EMH, the RW & Predictability

Bonus Material

Review: Efficient Markets Hypothesis (EMH)

• Fama (1970)

"A market in which prices always *fully reflect* available information is '*efficient*'."

If we have new information (a new earnings announcement) prices will adjust immediately (or very fast). Prices (significantly) jump with relevant information. But, they have to jump enough to make profiting from new information impossible.

• Efficiency can only be defined with reference to a specific type of information set, I_t. Three versions of EMH, according to I_t:

(a) **Weak form:** $I_{WF,t}$:$\{P_t, P_{t-1}, \ldots, P_{t-q}, \ldots\}$

(b) **Semi-strong form:** $I_{SSF\ t}$: $\{I_{WF,t}$+ Public Information at $t\}$

(c) **Strong form**: $I_{SF\ t}$: $\{I_{SSF,t}$+ Private/Inside Information at $t\}$

Review EMH: Violations

- Violations:
  - Technical traders devising profitable strategies (weak EMH)
  - Reading a newspaper and devising a profitable trading strategy (semi-strong EMH)
  - Corporate insiders making profitable trades (strong EMH).

- Q: Can markets really be strong-form efficient?

- Perfectly rational factors may account for violations of EMH.

Review: The Random Walk Hypothesis

- We start testing the EMH by assuming log returns, r_t, follow a RW with a drift. We called this "Random Walk Model":

$$\Rightarrow r_t = \Delta p_t = \mu + \varepsilon_t = \Delta p_t$$

where $\varepsilon_t \sim D(0, \sigma^2)$.

- Different specifications for ε_t produce different testable hypothesis for the EMH-RW Model:

- **RW1**: ε_t is *independent and identically distributed* (*i.i.d.*) ~ D(0, σ^2). Not realistic. (Old tests: Cowles and Jones (1937)).

- **RW2**: ε_t is *independent* (allows for heteroskedasticity). Test using filter rules, technical analysis. (Alexander (1961, 1964), Fama (1965)).

- **RW3**: ε_t is *uncorrelated* (allows for dependence in higher moments). Test using autocorrelations, variance ratios, long horizon regressions.

Review: The RWH – Autocorrelations & ACF

- Assume r_t is covariance stationary and ergodic. Then

  $\gamma_k = \text{cov}(r_t, r_{t-k})$ - Auto-covariance between times t & $t-k$
  $\rho_k = \gamma_k / \gamma_0$. - Var$[r_t] = \gamma_0$

are not time dependent. We estimate both statistics with $\hat{\gamma}_k$ and $\hat{\rho}_k$.

- Under **RW1** Hypothesis (and some assumptions)

$$\sqrt{T}\,\hat{\rho}_k \xrightarrow{a} \text{N}(0, 1)$$

$$\Rightarrow \quad \text{SE}[\hat{\rho}_k] = 1/\sqrt{T} \quad \Rightarrow \text{CI}\,[\hat{\rho}_k] = \{\hat{\rho}_k \pm 2/\sqrt{T}\}$$

- To check autocorrelations up to order k, we use the ACF for r_t.

Conclusion from individual tests: Few significant small in absolute value autocorrelations for **monthly** VW & EW (more for EW). More significant results for **daily** VW & EW, but still small in absolute value.

Review: The RWH – Autocorrelation Joint Tests

- We already know two tests to check for joint zero autocorrelation in a time series: Box-Pierce Q and Ljung-Box tests.

-The Q & LB statistics test a joint hypothesis that the first p autocorrelations are zero: H_0: $\rho_1 = ... = \rho_p = 0$

Under **RW1** and using the asymptotic distribution of $\hat{\rho}_k$:

$$\text{Q} = T \sum_{k=1}^{p} \hat{\rho}_k^2 \xrightarrow{d} \chi_p^2.$$

$$\text{LB} = T * (T-2) * \sum_{k=1}^{p} \frac{\hat{\rho}_k^2}{T-k} \xrightarrow{d} \chi_p^2.$$

Conclusion from joint tests: Strong rejection of H_0 for **monthly** VW & EW (stronger for EW). Even stronger rejection of H_0 for **daily** VW & EW. Not a good result for the RWH.

Review: The RWH – Autocorrelation Joint Tests

- Q & LB tests are widely use, but they have two main limitations:

(1) The test was developed under the independence (**RW1**) assumption.

If y_t shows **heteroscedasticity**, the asymptotic distribution used for Q tests is not correct. The "robust" Q statistic uses $\tilde{\rho}_k$ instead of ρ_k:

$$\tilde{\rho}_k = \frac{\hat{\gamma}_k}{\tau_k} = \frac{\sum_{t=k+1}^{T}(y_t - \bar{y})(y_{t-k} - \bar{y})}{\sum_{t=k+1}^{T}(y_t - \bar{y})^2 \; (y_{t-k} - \bar{y})^2}$$

Thus, $Q^* = T \; \sum_{k=1}^{p} \tilde{\rho}_k^2 \xrightarrow{d} \chi_p^2$.

(2) The number of autocorrelations p is arbitrary. Optimality can be introduce through IC. We call this procedure as "*automatic selection*."

- It is common to reach different conclusion from Q and Q*.

7

Review: Q & Q* Tests with Monthly Data

Example: Q and LB tests with p = **3** & **12 lags** for the **monthly** EW & VW CRSP index returns from 1926:Jan – 2022:March (T = 1155):

- Q test for **monthly VW**

```
> Box.test(lr_vw, lag = 4, type="Box-Pierce")

        Box-Pierce test

data:  lr_vw
X-squared = 22.812, df = 4, p-value = 0.000138

> Box.test(lr_vw, lag = 12, type="Box-Pierce")

        Box-Pierce test

data:  lr_vw
X-squared = 34.696, df = 12, p-value = 0.0005234
```

Note: Ljung-Box tests show similar results.

8

Review: Q & Q* Tests with Monthly Data

Example (continuation): Q* tests with automatic lag selection. In R, the package *vrtest* has the Auto.Q function that computes this test. As always, you need to install *vrtest* first.

- Q* test for **monthly VW**

```
> Auto.Q(lr_vw, 12)
$Stat
[1] 3.059582

$Pvalue
[1] 0.08026232
```

Conclusion: Once we take into consideration potential heteroscedasticity in y_t, there is weak evidence for autocorrelation in monthly Value-weighted CRSP index returns from

9

Review: Q & Q* Tests with Monthly Data

Example (continuation):

- Q test for **monthly EW**

```
> Box.test(lr_ew, lag = 4, type="Box-Pierce")

        Box-Pierce test

data:  lr_ew
X-squared = 61.607, df = 4, p-value = 1.333e-12

> Box.test(lr_ew, lag = 12, type="Box-Pierce")

X-squared = 83.328, df = 12, p-value = 9.531e-13
```

- Q* test for **monthly EW**

```
library(vrtest)
> Auto.Q(lr_ew, 12)
$Stat
[1] 6.487553

$Pvalue
[1] 0.01086324
```

10

Review: Q & Q* Tests – Overall Evidence

- Overall, at the **monthly** level, there is strong evidence for joint autocorrelation in **EW returns**. That is, we reject the RW hypothesis (RWH) for **monthly EW** returns.

But, the evidence is much weaker –i.e., not significant at 5% level– for **monthly VW returns**, once we take into account potential heteroscedasticity for returns.

At the **daily** level, we have a strong rejection of the RWH for both series: **VW** & **EW returns.**

11

Review: The RWH – Variance Ratio (VR) Tests

- Intuition: For all 3 RW hypotheses, the variance of RW increments is linear in the time interval. If the interval is twice as long, the variance must be twice as big. That is, the variance of monthly data should be 4 times bigger than the variance of weekly data. (Recall the log approximation rules for *i.i.d.* returns.)

- If r_t is a covariance stationary process (constant first two moment, and covariance independent of time), then for the variance ratio of 2-period versus 1-period returns, VR(2):

$$\text{VR}(2) = \frac{\text{Var}[r_t(2)]}{2*\text{Var}[r_t]} = \frac{\text{Var}[r_t+r_{t+1}]}{2*\text{Var}[r_t]} =$$

$$= \frac{\text{Var}[r_t] + \text{Var}[r_{t+1}] + 2\,\text{Cov}[r_t,r_{t+1}]}{2*\text{Var}[r_t]} = \frac{2\,\sigma^2 + 2\,\gamma_1}{2\sigma^2} = 1 + \rho_1$$

where $r_t(2) = r_t + r_{t+1}$

Review: The RWH – Variance Ratio (VR) Tests

- $$\text{VR}(2) = \frac{\text{Va}[r_t(2)]}{2*\text{Var}[r_t]} = 1 + \rho_1.$$
- When $\rho_1 = 0 \Rightarrow \text{VR}(2) = 1$ (True under RW1, random walk)
- The intuition generalizes to longer horizons:

$$\text{VR}(q) = \frac{\text{Var}[r_t(q)]}{q*\text{Var}[r_t]} = 1 + 2 * \sum_{k=1}^{q-1}(1 - \frac{k}{q})\,\rho_k.$$

The VR(q) is a particular linear combination of the 1st (q - 1) autocorrelation coefficients (with linearly declining weights).

- Under RW1, we have H_0: VR(q) = 1.

  H_1: VR(q) ≠ 1.

Review: RWH – VR Tests & Distribution

- To do any testing we need the sampling distribution of the VRs (variance ratios) under H_0: VR(q) = 1. We use:

$$\frac{\sqrt{Tq}}{\sqrt{2*(q-1)}}(\widehat{\text{VR}}(q) - 1) \xrightarrow{a} N(\mathbf{0}, 1)$$

- For the special case of $q = 2$, we use

$$\sqrt{T}\,(\widehat{\text{VR}}(2) - 1) \xrightarrow{a} N(\mathbf{0}, 1)$$

- $\text{Var}[r_t(q)]$ is computed using the MLE formulation –i.e., dividing by T, not by (T - 1). For example, for r_t:

$$\text{Var}[r_t] = \frac{\sum_{t=1}^{T}(r_t - \bar{r})^2}{T}$$

- $\text{Var}[r_t(q)]$ is computed using **non-overlapping returns**.

Review: RWH – VR Tests & Computations

Example: We have monthly data from Jan 1973. Then, we compute monthly & bi-monthly variances using the **MLE formulation**:

$$\text{Var}[r_t] = \frac{\sum_{t=1}^{T}(r_t - \bar{r})^2}{T}$$
$$\text{Var}[r_t(2)] = \frac{\sum_{t=1}^{T}(r_t(2) - 2*\bar{r})^2}{T}$$

- We use **non-overlapping returns**, that is, for bi-monthly returns, we **add** two (contiguous, non-overlapping) monthly returns:

(1) monthly returns: r_t is computed as usual. For the first return:

$$r_{t=Jan\ 73} = \ln(P_{t=Jan\ 31,73}) - \ln(P_{t=Jan\ 1,73})$$

(2) bi-monthly returns. The first three $r_t(2)$ are computed as:

$$r_{t=Feb\ 73}(2) = r_{t=Feb\ 73} + r_{t=Jan\ 73}$$
$$r_{t=Apr\ 73}(2) = r_{t=Apr\ 73} + r_{t=Mar\ 73}$$
$$r_{t=June\ 73}(2) = r_{t=June\ 73} + r_{t=May\ 73}$$

The RWH: VR Test Monthly Evidence (VW)

Example: We check the RW Hypothesis, under RW3, for the monthly CRSP EW and VW Index returns. In R, the package *vrtest* has functions to compute the above mentioned VR tests.

- VR tests for **monthly VW**

```
library(vrtest)
kvec <- c(2,3,12)                              #Vector with different q
y <- lr_vw
> vr_1 <- VR.minus.1(y, kvec)                  # Stat should be close to 0 if RW
> vr_1
$VR.auto                                       # VR with Automatic ("optimal) q selection
[1] 0.1954746

$Holding.Periods
[1]  2  3 12

$VR.kvec                                       (VR – 1) stat for each q=kvec[i]
[1] 0.1007011 0.1187365 0.1212423

> sqrt(T*kvec)/sqrt(2*(kvec-1))*vr_1$VR.kvec   # VR test for each q=kvec[i] ~ N(0,1)
[1] 3.422358 3.494666 3.043158
```

The RWH: VR Test Monthly Evidence (EW)

Example (continuation):

• VR tests for **monthly EW**

```
> y <- lr_ew
> vr_1 <- VR.minus.1(y, kvec)                    # Stat should be close to 0 if RW
> vr_1
$VR.auto                                          # VR with Automatic ("optimal) q selection
[1] 0.1954746

$Holding.Periods
[1]  2  3 12

$VR.kvec                                          (VR – 1) stat for each q=kvec[i]
[1] 0.2043236 0.2789327 0.2180176

> sqrt(T*kvec)/sqrt(2*(kvec-1))*vr_1$VR.kvec      # VR test for each q=kvec[i] ~ N(0,1)
[1] 6.943998 8.209583 5.472199
```

Conclusion: Using the VR test (with $q = 2, 3, 12$), we reject the RW Hypothesis ⇒ tests are greater in absolute value than **1.96**.

The RWH: VR Tests – Issues

• Several issues has been raised regarding the VR's tests. The main issues are:

(1) Choice of q. In the previous examples, we have arbitrarily selected q. Similar to the situation with the Q test, there are suggestions to automatically (or "*optimally*," according to some loss function) select q. Choi (1999) is one example of this approach, (the *vrtest* R package uses this approach in the *Auto.VR* test).

(2) Poor asymptotic approximation. In simulations, it is found that the asymptotic Normal distribution is a poor approximation to the small-sample distribution of the VR statistic. The usual solution is to use a bootstrap (Kim's (2009) bootstrap gives the p-value of the automatic VR test in the *Auto.VR* function).

The RWH: VR Test Monthly Evidence (VW)

Example: We use VR tests with automatic selection and a bootstrap to check the RW Hypothesis for the monthly CRSP EW and VW Index returns. Again, we use AutoBoot.test function in R package *vrtest*.

- Automatic VR tests for **monthly VW**

```
y <- lr_vw
> AutoBoot.test(y, nboot=1000, wild="Normal", prob=c(0.025,0.975))        # Choi (1999)
$test.stat                    (Automatic variance ratio test statistic as in Choi (1999))
[1] 2.509324

$VRsum                        (1+ weighted sum of autocorrelation up to the optimal order)
[1] 1.195475

$pval
[1] 0.064

$CI.stat
    2.5%     97.5%
-2.836631  2.612363

$CI.VRsum
    2.5%     97.5%
0.8323731 1.1927214
```

The RWH: VR Test Monthly Evidence (EW)

Example (continuation):.

- Automatic VR tests for **monthly EW**

```
y <- lr_ew
> AutoBoot.test(y, nboot=1000, wild="Normal", prob=c(0.025,0.975))        # Choi (1999)
$test.stat                    (Automatic variance ratio test statistic as in Choi (1999))
[1] 4.173898
$VRsum                        (1+ weighted sum of autocorrelation up to the optimal order)
[1] 1.382554
$pval
[1] 0.021

$CI.stat
    2.5%     97.5%
-3.262026  3.359002

$CI.VRsum
    2.5%     97.5%
0.7687769 1.2610106
```

Conclusion: Using the Automatic VR test and a bootstrap, we have strong evidence against the RW Hypothesis for EW, but weak for VW.

The RWH: VR Tests – LM's Modifications

- Lo & MacKinlay (LM, 1988, 1989) propose modifications to the test:

- Allow for **overlapping returns**, and, thus, use more observations. But, overlapping returns will be autocorrelated, even if underlying process is not. We need to adjust for this feature.

- Use **unbiased estimators of variances** –i.e., divide by (*T* - *df*).

$$M_1(q) = \frac{\sqrt{3*T*q}}{\sqrt{2*(2q-1)*(q-1)}}(\overline{VR}(q) - 1) \xrightarrow{a} N(\mathbf{0}, 1),$$

where $\overline{VR}(q)$ is the VR statistic computed using overlapping returns.

- Allow for possible **heteroscedasticity** of returns (more realistic)

$$M_2(q) = \frac{(\overline{VR}(q) - 1)}{\sqrt{\phi(q)}} \xrightarrow{a} N(\mathbf{0}, 1),$$

where

$$\phi(q) = \sum_{j=1}^{q}[\frac{2(q-j)}{q}]^2 * \{\frac{\sum_{t=j+1}^{T}(r_t - \bar{r})^2(r_{t-j} - \bar{r})^2}{[\sum_{t=1}^{T}(r_t - \bar{r})^2]^2}\}.$$

The RWH: LM Tests Monthly Evidence

Example (continuation): We check the RW Hypothesis, under RW3, for the monthly CRSP EW and VW Index returns using the LM's tests: M1 and M2. Again, we use the R package *vrtest.*

- Automatic VR tests for **monthly VW**

```
library(vrtest)
kvec <- c(2,3,12)                  #Vector with different q
y <- lr_vw
> Lo.Mac(y, kvec)                  # LM's tests M1 & M2 ~ asymptotic N(0,1)
$Stats
          M1        M2
k=2  3.422358 1.7485059
k=3  2.706957 1.4241521
k=12 1.099060 0.6373211
```

<u>Conclusion</u>: We reject H_0 (RW Model) using M1 for q = 2, 3; but, once we allow for heteroscedasticity (M2 tests), we cannot reject H_0.

The RWH: LM Tests Monthly Evidence

Example (continuation):

- Automatic VR tests for **monthly EW**

```
y <- lr_ew
> Lo.Mac(y, kvec)                    # LM's tests M1 & M2 ~ asymptotic N(0,1)
$Stats
        M1       M2
k=2  6.943998 2.5480302
k=3  6.359116 2.5009114
k=12 1.976326 0.9975538
```

Conclusion: Strong rejection of RW using M1, especially for $q = 2, 3$; but, using M2 test with $q = 12$, we cannot reject the RW Hypothesis.

Consistent with previous result, stronger evidence for EW returns than for VW returns.

The RWH: VR & LM Tests – Issues ($\xrightarrow{a}$ N(0,1))

- Several issues has been raised regarding the LM's tests:

(1) Poor asymptotic approximation. The asymptotic standard normal distribution provides a poor approximation to the small-sample distribution of the VR statistic. LM's tests tend to be biased and right-skewed, in finite samples.

- Proposed solutions:

- **Alternative asymptotic distributions**, as in Richardson and Stock (1989) or Chen and Deo (2006).

- **Bootstrapping**, as in Kim (2006) or Malliaropulos and Priestley (1999).

The RWH: VR & LM Tests – Issues (Joint Tests)

(2) Joint tests. The LM's tests are individual tests, where H_0 is tested for a specific value of q. But, under H_0, VR(q) = 1, for all q. LM's tests ignore the joint nature of testing for the RW Hypothesis.

- Proposed solutions:

- RS statistic, a Wald Test, as proposed by Richardson and Smith (1993):

$$RS(q) = T\,(\mathbf{VR} - \iota)'\,\Phi^{-1}\,T\,(\mathbf{VR} - \iota) \xrightarrow{d} \chi^2_q.$$

where **VR** is the (q×1) vector of q sample variance ratios, ι is the (q×1) unit vector, and Φ is the covariance matrix of **VR**.

- QP statistic, a Wald Test based on a "power transformed" VR statistic, as proposed by Chen and Deo (2006). QP asymptotically follows a χ^2_q distribution. This test is a one-sided test (H_1: VR(q) < 1 for all q.)

The RWH: VR & LM Tests – Issues (Joint Tests)

- Proposed solutions (continuation):

- CD statistic, a join test, as proposed by Chow and Denning (1993):

$$CD = \sqrt{T} \max_{1 \le i \le m} |\, M_2(q_i)|$$

which follows a complex distribution, the studentized maximum modulus [SMM] distribution with m and T degrees of freedom (m is the number of k values). This SMM distribution is tabulated in Hahn and Hendrickson (1971) and Stoline and Ury (1979).

In general, we use the simulated critical values obtained by simulations as done by Chow and Denning themselves or a bootstrap as in Kim (2006).

The RWH: VR & LM Tests – Monthly Evidence

Example: We check the monthly LM test results using a bootstrap instead of the asymptotic distribution. We use the *Boot.test* function in the R package *vrtest*, which provides two bootstrapped p-values: one for the LM statistic and the other one for the CD statistic.

- VR tests for **monthly VW**

```
> y <- lr_vw
> Lo.Mac(y, kvec)                # LM's tests M1 & M2
$Stats
        M1        M2
k=2  3.422358 1.7485059
k=3  2.706957 1.4241521
k=12 1.099060 0.6373211

> Boot.test(y, kvec, nboot=1000, wild="Normal", prob=c(0.025,0.975))       #Kim's Bootstrap
$Holding.Period
[1]  2  3 12

$LM.pval                         (Bootstrap p-values for the Lo-MacKinlay M2 tests)
[1] 0.067 0.157 0.503
```

The RWH: VR & LM Tests – Monthly Evidence

Example (continuation):

```
> Lo.Mac(y, kvec)                # LM's tests M1 & M2
$Stats
        M1        M2
k=2  3.422358 1.7485059
k=3  2.706957 1.4241521
k=12 1.099060 0.6373211

> Boot.test(y, kvec, nboot=1000, wild="Normal", prob=c(0.025,0.975))       #Kim's Bootstrap
$Holding.Period
[1]  2  3 12

$LM.pval                         (Bootstrap p-values for the Lo-MacKinlay M2 tests)
[1] 0.067 0.157 0.503

$CD.pval                         (Bootstrap p-value for the Chow-Denning test)
[1] 0.153

$CI                              (C.I. for Lo-Mackinlay M2 tests from Bootstrap distribution)
        2.5%     97.5%
k=2  -1.825961 1.827630
k=3  -1.847447 1.855263
k=12 -1.712367 2.152280
```

The RWH: VR & LM Tests – Monthly Evidence

Example (continuation):

```
> Wald(y, kvec)                         # RS Wald test
$Holding.Period
[1]  2  3 12

$Wald.stat
[1] 12.42735

$Critical.Values_10_5_1_percent
[1]  6.251389  7.814728 11.344867

> Chen.Deo(y, kvec)                     # QP Wald test
$Holding.Period
[1]  2  3 12

$VRsum
[1] 0.07335402

$QPn
[1,] 3.154226

$ChiSQ.Quantiles_1_2_5_10_20_percent
[1] 11.344867  9.837409  7.814728  6.251389  4.641628
```

The RWH: VR & LM Tests – Monthly Evidence

Example (continuation):

- VR tests for **monthly EW**

```
> y <- lr_ew
> Lo.Mac(y, kvec)              # LM's tests M1 & M2
$Stats
        M1        M2
k=2  6.943998 2.5480302
k=3  6.359116 2.5009114
k=12 1.976326 0.9975538

> Boot.test(y, kvec, nboot=1000, wild="Normal", prob=c(0.025,0.975))      #Kim's Bootstrap
$Holding.Period
[1] 5 20 60

$LM.pval                       (Bootstrap p-values for the Lo-MacKinlay M2 tests)
[1] 0.001 0.004 0.279

$CD.pval                       (Bootstrap p-value for the Chow-Denning test)
[1] 0.017
```

The RW Hypothesis: VR Test Monthly Evidence

Example (continuation):

$LM.pval (Bootstrap p-values for the Lo-MacKinlay M2 tests)
[1] 0.001 0.004 0.279

$CD.pval (Bootstrap p-value for the Chow-Denning test)
[1] 0.017

$CI (C.I. for Lo-Mackinlay M2 tests from Bootstrap distribution)

	2.5%	97.5%
k=2	-1.754012	1.708415
k=3	-1.710910	1.816157
k=12	1.563058	2.092434

> Wald(y, kvec) # RS Wald test
$Holding.Period
[1] 2 3 12

$Wald.stat
[1] 52.68679

$Critical.Values_10_5_1_percent
[1] 6.251389 7.814728 11.344867

The RW Hypothesis: VR Test Monthly Evidence

Example (continuation):

> Chen.Deo(y, kvec) # QP Wald test
$Holding.Period
[1] 2 3 12

$VRsum
[1] 0.1442001

$QPn
[1,] 6.524497

$ChiSQ.Quantiles_1_2_5_10_20_percent
[1] 11.344867 9.837409 7.814728 6.251389 4.641628

Conclusion: Consistent with previous result, solid evidence for the RW for EW returns, but weak evidence (only the Wald test rejects H_0) for VW returns.

The RW Hypothesis: VR Test Daily Evidence

Example: We check the RW Hypothesis, under RW3, for the daily CRSP EW and VW Index returns.

- VR tests for **daily VW**

```
kvec <- c(5, 20, 60)                      #Vector with different q
y <- lr_vw
vr_1 <- VR.minus.1(y, kvec)               # Stat should be close to 0 if RW
> vr_1
$VR.auto                    (value of VR-1 with automatic selection of holding vectors)
[1] 0.08049192

$Holding.Periods
[1]  5  20 60

$VR.kvec                    (the values of VR-1 for the chosen holding periods)
[1] 0.06015875  0.11155693  0.16958754

> sqrt(T*kvec)/sqrt(2*(kvec-1))*vr_1$VR.kvec  # VR test for each q=kvec[i] (~ N(0,1) dist)
[1] 1.616329 2.750494 4.109789
```

The RW Hypothesis: VR Test Daily Evidence

Example (continuation):

```
> AutoBoot.test(y, nboot=300, wild="Normal", prob=c(0.025,0.975))          # Choi (1999)
$test.stat
[1] 4.354851

$VRsum
[1] 1.080492

$pval
[1] 0.02333333

$CI.stat
    2.5%     97.5%
-3.423189  4.067023

$CI.VRsum
    2.5%     97.5%
0.9483973 1.0656480
```

The RW Hypothesis: VR Test Daily Evidence

Example (continuation):

```
> Lo.Mac(y, kvec)                  # LM's tests M1 & M2
$Stats
         M1       M2
k=5  4.372645 1.757401
k=20 3.574490 1.573525
k=60 3.057608 1.536068

> Boot.test(y, kvec, nboot=1000, wild="Normal", prob=c(0.025,0.975))        #Kim's Bootstrap
$Holding.Period
[1]  2  3 12

$LM.pval                         (Bootstrap p-values for the Lo-MacKinlay M2 tests)
[1] 0.06333333 0.08000000 0.07333333

$CD.pval                         (Bootstrap p-value for the Chow-Denning test)
[1] 0.11333

$CI                              (C.I. for Lo-Mackinlay M2 tests from Bootstrap distribution)
        2.5%    97.5%
k=5  -1.602225 2.333427
k=20 -1.594718 1.935643
k=60 -1.748524 1.782090
```

The RW Hypothesis: VR Test Daily Evidence

Example (continuation):

```
> Wald(y, kvec)                                        # RS Wald test
$Holding.Period
[1]  5 20 60

$Wald.stat
[1] 21.19834

$Critical.Values_10_5_1_percent
[1]  6.251389  7.814728 11.344867

> Chen.Deo(y, kvec)                                    # QP Wald test

$VRsum
[1] 0.05863072

$QPn
        [,1]
[1,] 3.639522

$ChiSQ.Quantiles_1_2_5_10_20_percent
[1] 11.344867  9.837409  7.814728  6.251389  4.641628
```

The RW Hypothesis: Evidence

• Tests based on CRSP value-weighted (VW) and equal weighted (EW) indices from **1925** & individual securities from **1962**.

• Daily, weekly and monthly returns from VW and EW indices show significant (positive) autocorrelation.

• VR(q) > 1 statistics reject RW3 for EW index but not VW index. Market capitalization or size may be playing a role. Rejection of RW stronger for smaller firms. Their returns more serially correlated.

• For individual securities, VR(q) < 1, suggesting small and negative correlations (and not significant).

• Rejection of the RWH does not imply that EMH is rejected. The main implication is for theory models: Need to incorporate autocorrelations.

Predictability

Bonus Material

Traditional Views of EMH (1960-1970)

- CAPM is a good measure of risk
- Usual findings:
  (a) Stock, bond and foreign exchange changes are not predictable
  (b) Constant equity premium
- Market volatility does not change much through time
- Professional managers do not reliably outperform simple indices and passive portfolios once one corrects for risk
- Summary of State of the Art, late 1970s (Jensen, 1978):

"I believe there is no other proposition in economics which has more solid evidence supporting it than the Efficient Markets Hypothesis."

Modern Empirical Research (1980 - present)

- Rejection of the RW Hypothesis.
- Stock returns are predictable.

1. Valuation ratios (D/P, E/P, B/M ratios)
2. Interest rates (term spread, short-long T-bill rates, etc.)
3. Decision of market participants (corporate financing, consumption).
4. Cross-sectional equity pricing.
5. Bond and foreign exchange returns are also predictable.

- Some funds seem to outperform simple indices, even after controlling for risk through market betas.
- New equilibrium (theory) models with time-varying equity premium.

Predictive Regressions

- Motivation:
  1. Mounting evidence that stock and bond returns are predictable.
  2. Q: Market inefficiency vs Rational variation in expected returns?

- Economic questions:
  1. Do the expected returns on bonds and stocks move together?
  2. Do the same variables forecast bond and stock returns?
  3. Is the variation in expected returns related to business cycles?

- Setup:

Regress future returns, $r_{t+\tau}$, on variables $\boldsymbol{x}_t$ known at time t.

$$r_{t+\tau} = \mu_t + \beta\, \boldsymbol{x}_t + \varepsilon_{t+\tau} \qquad (1)$$

where τ can be one month, one quarter, and one to four years.

Predictive Regressions: Fama & French (1989)

Example: Fama and French (JFE, 1989). (Tables in next slide.)

- $r_{t+\tau}$: value- & equal-weighted market portfolios of NYSE; value-weighted corporate bond portfolios.

- $\boldsymbol{x}_t$ variables:

– Dividend yields, D_t/P_t: Add monthly dividends for the year preceding time t divided by the value of the portfolio at time t

– Term Premium, $TERM_t$: Long term government bond yield minus treasuries –see, Keim and Stambaugh (1986).

– Default premium, DEF_t: AAA bond yields minus BAA bond yields –see, Keim and Stambaugh (1986).

<u>Sample</u>: Non-overlapping data for quarterly (T=244) & annual (T=61) data. For longer horizons (bi-annual+), overlapping observations.

Predictive Regressions: Fama & French (1989)

Table 2

Slopes, t-statistics, and R^2 from multiple regressions of excess returns on the term spread ($TERM$) and the value-weighted dividend yield (D/P) or the default spread (DEF); 1927–1987.[a]

T	Aaa	Aa	A	Baa	LG	VW	EW	Aaa	Aa	A	Baa	LG	VW	EW
	$r(t, t+T) = a + bD(t)/P(t) + cTERM(t) + e(t, t+T)$													
	Slopes for D/P							t-statistics for D/P slopes						
M	0.04	0.01	−0.05	−0.03	0.07	0.21	0.43	0.96	0.21	−0.55	−0.23	0.43	0.78	1.26
Q	0.20	0.14	0.03	0.14	0.48	1.09	2.05	1.07	0.69	0.09	0.31	0.79	1.09	1.58
1	0.30	−0.21	−0.64	−0.49	0.39	2.79	5.75	0.65	−0.45	−0.87	−0.68	0.30	1.31	2.31
2	1.09	0.25	0.99	0.91	3.92	8.89	15.66	1.38	0.31	0.95	0.62	1.62	3.13	4.98
3	1.83	1.17	2.91	2.87	7.83	12.20	20.21	2.24	1.66	6.49	2.62	3.21	3.77	4.91
4	2.72	2.16	4.14	4.47	10.54	15.37	23.29	3.91	2.76	5.20	6.22	5.94	5.08	4.74
	Slopes for $TERM$							t-statistics for $TERM$ slopes						
M	0.23	0.24	0.26	0.26	0.26	0.31	0.47	3.12	3.57	3.23	3.19	2.27	1.68	1.93
Q	0.57	0.52	0.59	0.55	0.60	0.65	0.96	1.75	1.63	1.83	1.74	1.37	1.03	1.10
1	3.36	3.30	3.92	3.36	3.61	1.56	2.69	5.44	5.25	5.81	4.88	3.78	0.82	1.01
2	4.23	4.22	4.46	4.15	4.41	−1.27	−0.38	3.98	3.83	2.83	2.80	1.85	−0.38	−0.07
3	5.01	4.62	4.86	4.63	4.66	−1.54	−0.17	3.52	3.27	2.56	2.47	1.70	−0.36	−0.03
4	5.09	4.61	5.50	5.22	6.06	0.32	4.07	2.39	2.15	2.36	2.66	2.95	0.09	0.73
	Regression R^2													
M	0.04	0.04	0.03	0.02	0.01	0.01	0.01							
Q	0.06	0.04	0.03	0.02	0.02	0.02	0.04							
1	0.39	0.33	0.30	0.20	0.11	0.02	0.07							
2	0.28	0.22	0.17	0.13	0.15	0.12	0.22							
3	0.26	0.19	0.21	0.18	0.24	0.18	0.27							
4	0.24	0.19	0.28	0.26	0.36	0.25	0.34							

Predictive Regressions: Fama & French (1989)

T	Aaa	Aa	A	Baa	LG	VW	EW	Aaa	Aa	A	Baa	LG	VW	EW
	$r(t, t+T) = a + bDEF(t) + cTERM(t) + e(t, t+T)$													
	Slopes for DEF							t-statistics for DEF slopes						
M	0.07	0.07	0.07	0.05	0.27	0.04	0.41	0.78	0.74	0.50	0.30	0.99	0.09	0.67
Q	0.31	0.34	0.47	0.54	1.30	0.99	2.78	0.85	0.91	0.84	0.85	1.30	0.56	1.12
1	0.76	0.41	0.76	1.49	4.12	4.38	11.59	0.79	0.46	0.57	1.12	1.62	1.15	2.38
2	4.18	3.51	6.41	6.70	13.49	14.62	29.22	1.96	1.75	2.86	2.47	3.04	2.18	2.73
3	7.21	7.14	11.12	11.83	22.25	19.96	37.61	2.01	2.14	3.21	5.51	4.61	2.25	2.52
4	10.11	9.66	13.62	15.29	27.07	24.56	41.41	2.11	2.07	3.20	4.46	4.64	2.42	2.80
	Slopes for $TERM$							t-statistics for $TERM$ slopes						
M	0.22	0.23	0.24	0.24	0.22	0.34	0.47	2.81	3.21	3.02	3.04	1.99	1.83	2.00
Q	0.55	0.48	0.49	0.47	0.42	0.70	0.83	1.51	1.36	1.48	1.41	0.97	1.01	0.92
1	3.25	3.15	3.58	2.89	2.73	1.20	1.35	4.41	4.49	4.89	3.96	3.08	0.60	0.52
2	3.48	3.41	3.13	2.73	2.10	−2.53	−3.47	2.89	3.07	2.04	1.97	1.19	−0.74	−0.81
3	3.63	3.06	2.75	2.32	1.02	−3.30	−4.29	1.80	1.74	1.26	1.24	0.50	−0.84	−0.80
4	3.09	2.56	2.94	2.29	1.68	−1.89	−0.45	0.98	0.88	1.03	0.99	0.76	−0.51	−0.12
	Regression R^2													
M	0.04	0.04	0.03	0.02	0.01	0.00	0.01							
Q	0.05	0.04	0.03	0.03	0.03	0.01	0.02							
1	0.39	0.33	0.29	0.20	0.15	0.00	0.07							
2	0.32	0.26	0.25	0.22	0.25	0.05	0.16							
3	0.34	0.29	0.33	0.31	0.36	0.09	0.20							
4	0.34	0.31	0.40	0.41	0.45	0.13	0.23							

[a] The regressions for T = one month (M), one quarter (Q), and one year use nonoverlapping returns. The regressions for two- to four-year returns use overlapping annual observations. The numbers of observations in the regressions are (M) 732, (Q) 244, (1 yr) 61, (2 yr) 60, (3 yr) 59, and (4 yr) 58. The standard errors in the t-statistics for the slopes are adjusted for heteroscedasticity and (for two- to four-year returns) the sample autocorrelation of overlapping residuals with the method of Hansen (1982) and White (1980). See note to table 1 for definition of portfolios.

Predictive Regressions: Fama-French Findings

Example (continuation):

• Findings:

– $\boldsymbol{x}_t$ variables work, especially D_t/P_t with high t-stats & high R^2 for forecast horizons beyond 1 year.

– (Conditional) Expected returns move with the predictors, $\boldsymbol{x}_t$:

$$E_t[r_{t+\tau}] = \hat{\mu}_t + \hat{\beta}\, \boldsymbol{x}_t$$

That is, even with $\mu_t = \mu$, expected future returns are **time-varying**!

– Regression coefficients and R^2 **increase** with the forecast horizon.

• Another well-cited paper is Lamont (JF, 1998), who finds that other financial ratios also work as predictors: **dividends yield** & **earnings yield**. Lamont also find that the dividend payout ratio has cross-sectional predictive power.

Predictive Regressions: Fama-French Findings

Example (continuation):

• Interpretation of slope estimate for D_t/P_t (similar of other financial ratios with P_t in the denominator):

- There is a positive relation between D_t/P_t and $r_{t+\tau}$. A high (low) D_t/P_t forecasts high (low) subsequent returns (higher $P_{t+\tau}$!). Since we tend to observe high D_t/P_t when P_t is low, we have evidence for **mean reversion** in stock prices.

- Let's look at the one-year D_t/P_t EW slope coefficient: **5.75**. Then, a 1% increase in dividend increase expected (total) returns by 5.75% (an investor gets 1% dividend plus 4.75% extra return). **Big number**!

- Using the above **5.75** slope, we can derive an informal range for the expected 1-year return: In the past 40 years D_t/P_t ranged from 1% to 6%, ignoring $\hat{\mu}_t$, the range for $E_t[r_{t+1y}]$ is {5.75% - 34%}. **Very big!**

Predictive Regressions: Fama-French Findings

Example (continuation):

• Interpretation of R^2 for D_t/P_t (again, similar interpretation for other ratios with P_t in the denominator):

- R^2 are small, but they start to be worth paying attention to for horizons of 1-year ahead or longer. "Small" and "big" are relative term, remember that according to the RW the R^2 should be 0! Then, any $R^2 >$ 0 is "interesting."

- For the EW returns, D_t/P_t predicts **7%** of the variability of one-year ahead returns and **23%** of the variability of 4-year ahead returns. These are results that, on average, can produce profitable investment strategies.

Predictive Regressions: Fama-French Findings

Example (continuation):

• Rational explanations for time-variation of expected return:

- Time-varying risk aversion
- Time-varying amount of risk
- Parallel behavior explanation (investor sentiment).

• Remark: We expect low prices –relative to D_t, E_t, $Book_t$– to be followed by high returns (high prices). Going back to the EMH, can we profit from this predictability?

Predictive Regressions: Updating Fama-French Results

Example: We use Shiller's data (**1871:Jan - 2021:Dec**) to redo the **monthly** predictive regressions of Fama-French (see FEc_prog_Pred for code and links to data).

Findings: With the exception of **lagged excess returns** and the **default yield spread** (AAA yield – BBB yield) nothing is significant.

	Independent Variable: Excess Returns at t+1 (1871-2021)						
	r_t	D_t/P_t	DY_t	E_t/P_t	D_t/E_t	DFY_t	DFR_t
μ	0.00398 (0.0011)	0.00992 (0.0141)	0.01694 (0.0142)	0.02570 (0.0150)	-0.0183 (0.0195)	0.00979 (0.0025)	0.00435 (0.0030)
β	0.11256 (0.0234)	0.00095 (0.0025)	0.00218 (0.0025)	0.00410 (.0029)	-0.0041 (0.0035)	-0.0939 (0.0435)	0.11446 (0.2169)
R^2	0.00731	0.0004	0.00001	0.0011	0.0008	0.0041	0.0002

Predictive Regressions: Updating Fama-French Results

Example: We use the expanded Goyal and Welch data (**1925:Dec - 2021:Dec**) to redo the **monthly** predictive regressions of Fama-French, using CRSP indexes (see FEc_prog_Pred for code and links to data).

Findings: Now, only **lagged excess returns** are significant. We just see "momentum" at work at the monthly level.

	Independent Variable: Excess Returns at t+1 (125-2021)						
	r_t	D_t/P_t	DY_t	E_t/P_t	D_t/E_t	DFY_t	DFR_t
μ	0.00354 (0.0016)	0.02127 (0.0199)	0.02709 (0.0201)	0.02605 (0.0199)	-0.00118 (0.0277)	0.00724 (0.0030)	0.01038 (0.0025)
β	0.08548 (0.0294)	0.00296 (0.0034)	0.00396 (0.0034)	0.00424 (.0038)	-0.0009 (0.0049)	0.10056 (0.2165)	-0.0475 (0.0438)
R^2	0.01267	0.0007	0.00117	0.00216	0.00002	0.00017	0.00104

Predictive Regressions: Methodological Issues

• **Data snooping**. Are D_t/P_t, $TERM_t$, Payout Ratios the only variables used in those regressions? The standard finance and economic databases used in academic and industry research (CRSP, Compustat, Refinivit) have thousands of potential predictors.

Recall Type I error: If we use 100 regressors, 5 will be significant at the 5% level!

• **Peso problem.** In the sample, we do not observe a "crash," which are very low probability events, but agents do compute that probability in the expectation. Then, on average, the sample average is biased!

• **Regime Change**. Always a potential problem. Maybe coefficients change with the business cycle, Fed policy, bull/bear markets, etc.

Predictive Regressions: Methodological Issues

• **Endogeneity.** Regressors are only predetermined, but not exogenous. OLS slopes have a small bias (Stambaugh, 1986). Traditional OLS S.E. are likely not appropriate (Hodrick, 1992).

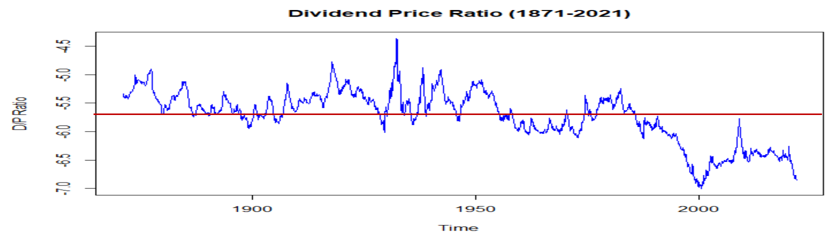
• **Persistence of Financial Ratios.** Valuation ratios are persistent and their innovations are correlated with returns, causing

– biased predictive coefficients: Stambaugh (1999)

– over-rejection by standard *t*-test: Cavanagh-Elliott-Stock (1995)

Note: These issues are less relevant for interest rates & recently proposed predictor variables (persistent, but less correlated with r_t).

Predictive Regressions: Valuation Ratios – Persistence



• D_t/P_t is persistent, D_t/P_t stays "high" or "low" for a long time. It moves around a constant mean (in red) & has no trend (stationary?).

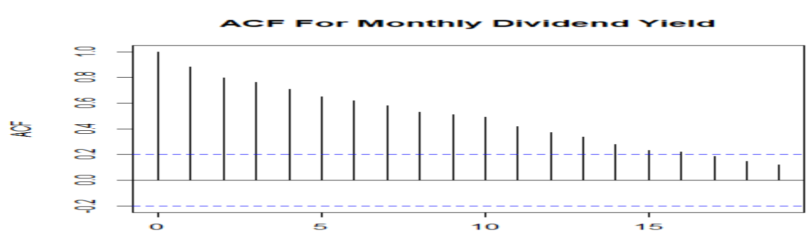
• There is some evidence for **mean reversion**, but it can take many years (decades?) to get back to the mean.

• Given the persistence in D_t/P_t, the Fama-French results imply that we should also have persistence in the forecast of expected returns. That is, we have high (low) expected returns for a long time (decades?)!

Predictive Regressions: Valuation Ratios – Persistence

• Issue: How persistent is D_t/P_t?
- D_t/P_t is likely to be persistent: it reflects long-run expectations.
- But, is D_t/P_t stationary? unit root? explosive?

• To answer the above question, we compute the ACF for D_t/P_t. (Recall that a persistent series will show a slow decay in the ACF.)



• The first order autocorrelation is **0.882**. Very persistent series! That is, next period dividend yield is very likely to be similar to this period.

Predictive Regressions: Valuation Ratios – Recessions

• There seems to be a relation (non-linear?) between D_t/P_t & the business cycle. We see big spikes in D_t/P_t when there is a recession (clear spike in the 1930s and in 2008-2009). Though these spikes are relatively short-lived (years, not decades).

Thus, expected returns vary with the business cycle (not a surprise): A big increase when there is a recession (risk is higher).

• Potential Problem with D_t: "**too smooth**" (measurement error?). The observed data may not be the "true" series of interest.

Subtle point: Since D_t is too smooth, all the predictability comes from P_t. What news affect more future stock prices (& returns): "Cash Flows news or Discount Rates news"? Discount rates news.

Predictive Regressions: Stambaugh Bias

• One econometric issue in Fama and French (1989): Regressors are only predetermined, but not exogenous.

• Start with predictive regression for returns, r_{t+1}:

$$r_{t+1} = \alpha + \beta\, \boldsymbol{x}_t + \varepsilon_{t+1}$$

$\boldsymbol{x}_t$: D_t/P_t –i.e., the dividend price ratio

Note: $\boldsymbol{x}_t$ depends on the price at the beginning of t, the change of $\boldsymbol{x}$ at the end of t+1 reflects changes in price from t to t+1 , as does r_{t+1}; $E[\varepsilon_{t+1} | \boldsymbol{x}_{t+1}, \boldsymbol{x}_t] \neq 0$, more generally, $E[\varepsilon_t | \boldsymbol{x}_s, \boldsymbol{x}_w] \neq 0$, $s < t < w$.

• Assumption (**A2**) is violated!

• In addition, $\boldsymbol{x}_t$ is persistent. It can be modeled with an ARMA.

Predictive Regressions: Stambaugh Bias

• Stambaugh (1999) assumes that $\boldsymbol{x}_t$ follows an AR(1)

$$\boldsymbol{x}_t = \mu + \phi\, \boldsymbol{x}_{t-1} + \nu_t \qquad (2)$$

where ν_t & ε_t follow a multivariate N(0, $\boldsymbol{\Sigma}$), independent across t.

Results: b (OLS estimate) is biased upward, positively skewed, and has higher variance and kurtosis than the normal sampling distribution of the OLS estimator.

• Stambaugh bias:

$$E[b - \beta] = (\sigma_{\varepsilon\nu} / \sigma_\nu^2)\, E[\hat{\phi} - \phi]$$

It turns out $\hat{\phi}$ has a downward bias and $\sigma_{\varepsilon\nu}$ is negative

⇒ b shows an upward bias. Conventional t-tests are misleading.

Finding: Correcting the bias weakens the predictability evidence.

Predictive Regressions: Dealing with Stambaugh Bias

• Since conventional t-tests are misleading, there are many suggestions to check if the predictability of the very persistent valuation ratios remains after correcting for the bias.

• One approach is Lewellen (2004): Adjust the OLS estimator under worst case scenario for persistence (ϕ = 1):

$$b_{adj} = b - (\sigma_{\varepsilon\nu} / \sigma_\nu^2)\, E[\hat{\phi} - 1]$$

• In practice, the estimated persistence is very close to one. The bias correction is small. Predictability survives:

- D_t/P_t predicts market returns from 1946–2000 and sub-samples.
- B/M and E_t/P_t predict returns during the shorter sample 1963–2000.

• Interesting Result: In a (1)-(2) framework NW SE are not reliable in small samples. Result from Hodrick (1992) & Kim and Nelson (1993).

Predictive Regressions: Long Horizon Returns (Aside)

- D_t/P_t and other ratios forecast excess returns on stocks. Regression coefficients and R^2 rise with the forecast horizon.
- This is a result of the fact that the forecasting variable is **persistent**.
- Model (1)-(2), assuming $\alpha = \mu = 0$.

$$r_{t+1} = \alpha + \beta\, \boldsymbol{x}_t + \varepsilon_{t+1} \qquad (1)$$
$$\boldsymbol{x}_t = \mu + \phi\, \boldsymbol{x}_{t-1} + \nu_t \qquad (2)$$

Now, we compound 2-period returns (with log returns, we add them):

$$\begin{aligned} r_{t+2}(2) &= \beta\, \boldsymbol{x}_{t+1} + \varepsilon_{t+2} + \beta\, \boldsymbol{x}_t + \varepsilon_{t+1} \\ &= \beta\, (\boldsymbol{x}_{t+1} + \boldsymbol{x}_t) + \varepsilon_{t+2} + \varepsilon_{t+1} \\ &= \beta\, (\phi\, \boldsymbol{x}_t + \nu_{t+1} + \boldsymbol{x}_t) + \varepsilon_{t+2} + \varepsilon_{t+1} \\ &= \beta\, (1+\phi)\, \boldsymbol{x}_t + \beta\, \nu_{t+1} + \varepsilon_{t+2} + \varepsilon_{t+1} \\ &= \beta_2\, \boldsymbol{x}_t + \omega_{2t} \qquad \Rightarrow \beta_2 > \beta. \end{aligned}$$

Predictive Regressions: Long Horizon Returns (Aside)

$$r_{t+2}(2) = \beta_2\, \boldsymbol{x}_t + \omega_{2t}$$

where $\beta_2 = (1+\phi) > \beta$.

- The previous result generalizes:

$$\begin{aligned} r_{t+k}(k) &= \beta\, \boldsymbol{x}_{t+k} + \varepsilon_{t+k} + \beta\, \boldsymbol{x}_{t+k-1} + \varepsilon_{t+k-1} + \cdots + \beta\, \boldsymbol{x}_t + \varepsilon_{t+1} \\ &= \beta\, (\boldsymbol{x}_{t+k} + \boldsymbol{x}_{t+k-1} + \cdots + \boldsymbol{x}_t) + \varepsilon_{t+k} + \cdots + \varepsilon_{t+1} \\ &= \beta\, (1 + \phi + \phi^2 + \ldots + \phi^k)\, \boldsymbol{x}_t + \omega_{kt} \\ &= \beta_k\, \boldsymbol{x}_t + \omega_{kt} \qquad \Rightarrow \beta_k > \beta_{k-1}. \end{aligned}$$

The coefficient of the persistent ratio is increasing with the horizon of compounding returns.

Note: A more complicated derivation is needed for the increase in R^2.

Predictive Regressions: More Predictors

- Lots of variables have been proposed as predictors. A short list:

\- **Book-to-market** (**b/m$_{t-1}$**), **equity share in new issues** (S, **equis$_{t-1}$**), and **lagged returns**, as in Baker and Wurgler (2000) (**B-W**, next slide).

\- **Cross-sectional premium** (**csp**): The relative valuations of high- and low-beta stocks, as in Polk, Thompson, and Vuolteenaho (2006).

\- ***Net Equity Expansion*** (**ntis**): The ratio of 12-month moving sums of net issues by NYSE listed stocks divided by the total end-of-year market capitalization, as in Boudoukh, et al. (2007).

\- **Long Term Yield** (**lty**): Long-term government bond yields.

\- **Investment to Capital Ratio** (**i/k**): The ratio of aggregate (private nonresidential fixed) investment to aggregate capital for the whole economy, as in Cochrane (1991).

\- **Consumption, wealth, income ratio** (**cay**): Estimated from an equation from a model proposed by Lettau and Ludvigson (2001).

Predictive Regressions: More Predictors – B-W Results

Table V
Multivariate OLS Regressions for Predicting One-Year-Ahead Market Returns

OLS regressions of annual real equity market returns on multiple predictors:

$$R_{Et} = a + b_1 R_{Et-1} + b_2 BILL_{t-1} + b_3 TERM_{t-1} + b_4 D/P_{t-1} + b_5 B/M_{t-1} + b_6 S_{t-1} + u_t$$

where R_E denotes real percentage returns on the CRSP value-weighted (VW) or equal-weighted (EW) portfolio, $BILL$ denotes the return on Treasury bills, $TERM$ denotes the yield premium of long-term government bonds over treasuries, D/P denotes the dividend-to-price ratio, B/M denotes the book-to-market ratio, and S denotes the equity share in new issues. The dividend-to-price ratio, the book-to-market ratio, and the equity share are standardized to have zero mean and unit variance. t-statistics are shown in brackets using heteroskedasticity robust standard errors.

	1928–1997 Returns		1928–1962 Returns		1963–1997 Returns	
	VW CRSP	EW CRSP	VW CRSP	EW CRSP	VW CRSP	EW CRSP
Intercept	6.95	21.72	14.33	21.71	11.50	19.23
	[1.13]	[1.68]	[0.53]	[0.76]	[0.78]	[1.16]
R_E	0.05	0.08	0.27	0.20	−0.20	−0.09
	[0.39]	[0.82]	[1.12]	[1.09]	[−1.01]	[−0.68]
$BILL$	0.71	−0.85	4.96	9.60	0.66	4.20
	[0.89]	[−0.47]	[0.75]	[1.28]	[0.40]	[1.46]
$TERM$	−0.86	−3.66	−7.98	−10.86	0.15	6.09
	[−0.41]	[−0.96]	[−0.70]	[−0.84]	[0.08]	[1.45]
D/P	4.26	−1.58	−4.37	−9.17	14.51	63.21
	[1.13]	[−0.27]	[−0.51]	[−1.55]	[1.43]	[2.41]
B/M	1.51	13.50	19.59	34.10	−7.30	−14.30
	[0.38]	[2.38]	[1.99]	[6.34]	[−1.29]	[−1.47]
S	−7.88	−13.17	−8.84	−14.34	−8.27	−13.63
1	[−3.97]	[−3.77]	[−1.94]	[−2.21]	[−2.13]	[−2.48]
$\bar{R}^2$	0.12	0.28	0.27	0.51	0.12	0.29
N	70	70	35	35	35	35

Table from **Baker and Wurgler (2000)** – **Annual Data**.

Predictive Regressions: Updating B-W Results

Example: We use the expanded Goyal and Welch data (**1927 - 2021**) to redo the **annual** predictive regressions of Baker-Wurgler, using S&P **excess returns** (see FEc_prog_Pred for code & data). Script for ik:

```
Pred_da_a <- read.csv("http://www.bauer.uh.edu/rsusmel/4397/goyal-welch-a_27.csv",
head=TRUE, sep=",")
lr_sp <- Pred_da_a$sp_ret                # Value weighted S&P returns (with distributions)
ik <- Pred_da_a$ik                       # Investment-to-capital
TA <- length(lr_sp)
TI <- 21
y_a_ik <- lr_sp[(TI+1):TA] - Rf_a[(TI+1):TA]/100
ik_a <- ik[TI:(TA-1)]
fit_lag_y_ik <- lm(y_a_ik ~ ik_a)
> summary(fit_lag_y_ik)
Coefficients:
              Estimate Std. Error t value Pr(>|t|)
(Intercept)  0.23805    0.06206   3.836 0.000268 ***
ik_a        -0.07640    0.01747  -4.372 4.13e-05 ***
```

Predictive Regressions: Updating B-W Results

Example (continuation):

Findings: Consistent with the previous table for VW returns, **equity share in new equity** is significant. We also run predictive regressions for the other variables mentioned above. **Investment-to-capital** (ik, starting in **1947**) was very significant, with very high R^2. (Note: **cay** (starting in **1944**) & **csp** (starting in **1937**) were not significant).

	Independent Variable: Excess Returns at t+1 (1927 - 2021)						
	r_t	D_t/P_t	E_t/P_t	DFY_t	BM_t	$eqis_t$	ik_t
μ	-0.0087 (0.0096)	0.02039 (0.0668)	0.00511 (0.0626)	-0.0152 (0.0169)	-0.0212 (0.0214)	0.04094 (0.0174)	0.23805 (0.0621)
β	0.00234 (0.1093)	0.00856 (0.0195)	0.00501 (0.0225)	0.00555 (0.0119)	0.00023 (0.0003)	-0.0027 (0.0008)	-0.07640 (0.0175)
R^2	0.00001	0.0021	0.0005	0.0331	0.0046	0.1066	0.2121

Predictive Regressions: Way More Predictors

• With the advances in computer power, the success of finding predictors of future returns has continued almost exponentially. For example, using Machine Learning models (Neural Networks) we have:

- Gu, Kelly and Xiu (2020): **176 predictors**, grouped in **94 stock-level predictive characteristics** (Green et al. (2017)); **8 macroeconomic & financial variables predictors** (Welch and Goyal (2008)); and **74 industry dummies** (& even 94 * 8 interaction terms!).
- Bianchi, Buchner and Tamoni (2021): **128 monthly macroeconomic and financial variables** (McCracken and Ng (2015)).

• Always keep in mind that the standard finance databases for research (CRSP & Compustat) have **over 1,000 potential predictors** (without counting interactions). It is always possible to find more predictors!

Q: Why not use them all?

Predictive Regressions: In-sample vs Out-of-sample

• In a very well know paper, Goyal and Welch (2008) argue that the in-sample (IS) predictability seen in predictive regression, once evaluated out-of-sample (OOS), becomes very weak or just disappears.

• Setup of OOS Evaluation

(1) Perform Q τ-step-ahead forecasts using:

- Rolling predictive regressions, adding one observation at a time. That is, we obtain Q forecasts, $\hat{r}_{t+\tau}$.
- Use the mean of the rolling period at time t as the forecast. That is, we obtain Q forecasts, $\bar{r}_t$.

(2) Get Q rolling forecast errors, e_A, & Q mean forecasts, e_N.

(3) Compute MSE_A & MSE_N.

(4) Evaluate MSEs using the Diebold-Mariano test.

Predictive Regressions: In-sample vs Out-of-sample

- An OOS R^2 can be computed as:

$$R^2_{OOS} = 1 - \frac{MSE_A}{MSE_N}$$

with $MSE_A = \sum_{t=1}^{Q}(r_{t+\tau} - \hat{r}_{t+\tau})^2$

$$MSE_N = \sum_{t=1}^{Q}(r_{t+\tau} - \bar{r}_t)^2$$

Note: Goyal and Welch (2008) evaluate the MSEs using other tests, proposed by Clark and McCracken (2001) and McCracken's (2004) variation of the Diebold-Mariano test.

- Findings: Very difficult to identify any robust predictor of excess stock returns. There are short time intervals of significant OOS predictability, but these "*pockets of predictability*" are surrounded by long periods of little or no predictability, see Lansing, LeRoy & Ma (2022).

Predictive Regressions: Goyal & Welch (2008) Results

Table 1
Continued

			Full Sample								
				Forecasts begin 20 years after sample				Forecasts begin 1965			
			IS	IS for	OOS			IS for	OOS		
	Variable	Data	$\overline{R}^2$	OOS $\overline{R}^2$	$\overline{R}^2$	ΔRMSE	Power	OOS $\overline{R}^2$	$\overline{R}^2$	ΔRMSE	Power
Full Sample, Significant IS											
b/m	Book to market	1921–2005	**3.20***	1.13	−1.72	−0.01	42 (67)	−7.29	−12.71	−0.77	40 (61)
i/k	Invstmnt capital ratio	1947–2005	**6.63****	−0.25	−1.77	0.07	47 (77)		Same		
ntis	Net equity expansion	1927–2005	**8.15*****	−4.21	−5.07	−0.26	57 (78)	0.96	−6.79	−0.32	53 (72)
eqis	Pct equity issuing	1927–2005	**9.15*****	2.81	**2.04****	0.30	72 (85)	3.64	−1.00	0.12	66 (77)
all	Kitchen sink	1927–2005	**13.81****	2.62	−139.03	−5.97	- (-)	−20.91	−176.18	−6.19	- (-)
Full sample, no IS equivalent (caya, ms) or Ex-Post Information (cayp)											
cayp	Cnsmptn, wlth, incme	1945–2005	**15.72*****	20.70	**16.78*****	1.61	- (-)		Same		
caya	Cnsmptn, wlth, incme	1945–2005	-	-	−4.33	−0.14	- (-)		Same		
ms	Model selection	1927–2005	-	-	−22.50	−1.69	- (-)	-	−23.71	−1.79	- (-)
1927-2005 Sample, Significant IS											
d/y	Dividend yield	1927–2005	**2.71***					−0.35	−6.44	−0.30	30 (71)
e/p	Earning price ratio	1927–2005	**3.20***					−0.94	−3.15	−0.05	39 (64)
b/m	Book to market	1927–2005	**4.14***					−8.65	−19.46	−1.26	45 (64)

Table from **Goyal & Welch (2008)** – **Annual Data**.

Predictive Regressions: Updating Goyal-Welch Results

Example: We use the expanded Goyal and Welch data (**1927 - 2021**) to compute their **annual** OOS R^2, using rolling regressions starting in **1967**, and perform Diebold-Mariano (DM) tests for significant differences of the forecasts (R script on next page for ik).

Findings: Consistent with the results of Goyal and Welch (2008), we do not find a lot of consistent predictability out of sample. In general, DM tests fail to reject H_0 that the predictors do better than the unconditional mean in forecasting next year excess returns.

Predictive Regressions: Updating Goyal-Welch Results

Example (continuation): R Code for ik (OOS rolling regressions)

```
yy <- y_a_ik                              # Dependent variable (y_t+1) of rolling regression
xx <- ik_a                                # Independent  variable x_t
Alles = NULL                              # Initialize empty (a space to put forecasts errors)
k_for <- 40                               # Start of Rolling Sample
i <- k_for                                # Counter for while loop
TF <- length(yy)
while (i <= TF-1) {
  y_tp1 <- yy[1:i]
  x_t <- xx[1:i]
  pred_reg <- lm (y_tp1 ~ x_t)            # OLS predictive regression
  b_hat <- pred_reg$coefficients          # Extract coefficient
  y_hat <- b_hat[1]+b_hat[2]*xx[i+1]      # t+1 forecast
  f_e_a <- y_hat - yy[i+1]                # t+1 forecast error for model
  f_e_n <- mean(y_tp1) - yy[i+1]          # t+1 forecast error for mean
  f_2e <- c(f_e_a, f_e_n)                 # Combine both forecast errors in a vector
  Alles = rbind(Alles,f_2e)               # accumulate forecast errors in rows (two columns)
  i <- i+1
}
```

Predictive Regressions: Updating Goyal-Welch Results

Example (continuation): R Code for ik

```
# Checking accuracy of forecasts with OOS R^2
mse <- colSums(Alles^2)/(TF-k_for)
r2_oos <- 1 - mse[1]/mse[2]
> r2_oos
[1] 0.02177127                    ⇒ Reduction in R^2.

# Testing accuracy of forecasts with Diebold-Mariano
> dm.test(Alles[,1], Alles[,2], power=2)

        Diebold-Mariano Test

data:  Alles[, 1]Alles[, 2]
DM = -0.12985, Forecast horizon = 1, Loss function power = 2, p-value = 0.8975
alternative hypothesis: two.sided

>dm.test(Alles[,1], Alles[,2], power=1)
DM = -0.23874, Forecast horizon = 1, Loss function power = 1, p-value = 0.8128
alternative hypothesis: two.sided
```

Predictive Regressions: Final Remarks

- Big and active literature, lately using ML/AI models. It has found lots of potential predictors of excess stock returns, for example, Gu, Kelly and Xiu (2020) use Neural Networks to discover **176 predictors** (with interaction terms, they almost use almost 1,000 predictors!)

- Given the usual data mining results in large datasets, many of the discovered predictors are not "true predictors," but "false positive (FP) predictors." A lot of FP predictors will increase C.I. for forecasts.

- We have a typical model selection problem. If we use the General-to-specific approach, the question is: How to reduce the GUM? Several proposals: optimize R^2_{OOS}, OOS SR, minimize FP predictors, etc.

- Old question: Can we make money from these predictors? Not clear.