Predictability at Long Horizons

Financial Econometrics (M524) Lecture Notes

Ke-Li Xu

Indiana University

September 28, 2022

Contents

- Multiple-horizon predictive regression
- Estimation
    - Direct estimator
    - Implied estimator
- Testing
    - Long-run variance
    - Horizon-h regression: mean only
    - Horizon-h regression: slope only
    - Testing in action: Three standard errors (Newey-West, Hansen-Hodrick, Hodrick)
- The dependent variable: When relevant?
- Why uses DP as predictor: A present value model
- Two ways of defining horizon-h returns

Multiple-horizon predictive regression

- Nobel Press Release (2013, Awarded to Fama, Hansen and Shiller):

"There is no way to predict the price of stocks and bonds over the next few days or weeks. But it is quite possible to foresee the broad course of these prices over longer periods, such as the next 3 to 5 years."

- One-period return: y_t
- Predictor: x_{t-1}
- horizon-h predictability: regress $\sum_{j=0}^{h-1} y_{t+j}$ on x_{t-1}
- Issues of accumulating and overlapping (e.g. $h = 3$ as below)

$$
\begin{array}{ll}
y_2 + y_3 + y_4 & x_1 \\
y_3 + y_4 + y_5 & x_2 \\
y_4 + y_5 + y_6 & x_3 \\
\vdots & \vdots \\
y_t + y_{t+1} + y_{t+2} & x_{t-1} \\
\vdots & \vdots \\
y_{T-2} + y_{T-1} + y_T & x_{T-3}
\end{array}
$$

- h can be large (e.g. $h = 60$)

Multiple-horizon predictive regression: Direct estimator

- Horizon-one predictive regression:

$$y_t = \mu_y + \beta x_{t-1} + u_t. \tag{1}$$

- Horizon-h predictive regression:

$$\sum_{j=0}^{h-1} y_{t+j} = \mu(h) + \theta(h) x_{t-1} + u_t(h). \tag{2}$$

- Direct estimator $\widehat{\theta}(h)$: OLS of $\sum_{j=0}^{h-1} y_{t+j}$ on x_{t-1}.

Multiple-horizon predictive regression: Implied estimator

- The expression for $\theta(h)$:

$$\theta(h) = (\Upsilon_0 + \Upsilon_1 + ... + \Upsilon_{h-1})\beta, \tag{3}$$

where $\Upsilon_j = Cor(x_{t+j}, x_t)$ is the j-th autocorrelation.

Why (3) holds?

$$
\begin{aligned}
y_t &= \mu_y + \beta x_{t-1} + u_t, \\
y_{t+j-1} &= \mu_y(j) + \beta(j) x_{t-1} + u_t^{(j)}, \qquad (4)
\end{aligned}
$$

where $j \geq 1$.

- Take the covariance between (4) and x_{t-1},

$$Cov(y_{t+j-1}, x_{t-1}) = \beta(j) Var(x_{t-1}).$$

- Then

$$
\begin{aligned}
\beta(j) &= [Var(x_{t-1})]^{-1} Cov(y_{t+j-1}, x_{t-1}) \\
&= [Var(x_{t-1})]^{-1} Cov(\mu_y + \beta x_{t+j-2} + u_{t+j-1}, x_{t-1}) \\
&= [Var(x_{t-1})]^{-1} Cov(\beta x_{t+j-2}, x_{t-1}) \\
&= [Var(x_{t-1})]^{-1} Cov(x_{t+j-2}, x_{t-1}) \beta \\
&= \Upsilon_{j-1} \beta.
\end{aligned}
$$

- Thus

$$\theta(h) = \sum_{j=1}^{h} \beta(j) = \sum_{j=0}^{h-1} \Upsilon_j \beta.$$

Identification of $\mu(h)$ (What is the true value of $\mu(h)$?)

$$\mu(h) = h\mu_y + \left(h - \sum_{j=0}^{h-1} \Upsilon_j\right) \beta Ex. \tag{5}$$

- By (1),

$$Ey_{t+j} = \mu_y + \beta Ex_{t+j-1}. \tag{6}$$

- Sum (6) over j,

$$\sum_{j=0}^{h-1} Ey_{t+j} = h\mu_y + \beta hEx. \tag{7}$$

- By (2),

$$\sum_{j=0}^{h-1} Ey_{t+j} = \mu(h) + \theta(h)Ex. \tag{8}$$

- Combining (7) and (8), and solving for $\mu(h)$, we have $\mu(h) = h\mu_y + [\beta h - \theta(h)]Ex \overset{(3)}{=} h\mu_y + [\beta h - \sum_{j=0}^{h-1} \Upsilon_j \beta]Ex$. Thus (5) holds.

Implied estimator of $\theta(h)$

- If x_t is AR(1): $x_t = \mu_x + \rho x_{t-1} + v_t$. Then

$$\theta(h) = \sum_{j=0}^{h-1} \rho^j \beta. \tag{9}$$

- If x_t is ARMA(1,1): $x_t = \mu_x + \rho x_{t-1} + v_t + \psi v_{t-1}$. Then for $h \geq 2$,

$$\theta(h) = \beta + \sum_{j=0}^{h-2} \rho^j \left(\rho + \frac{\psi(1-\rho^2)}{1+2\rho\psi+\psi^2} \right) \beta. \tag{10}$$

- Implied estimator $\widehat{\theta}_{IM}(h)$ based on AR(1): insert $\widehat{\rho}$ and $\widehat{\beta}$ into (9).
- Implied estimator $\widehat{\theta}_{IM}(h)$ based on ARMA(1,1): insert $\widehat{\rho}$, $\widehat{\psi}$ and $\widehat{\beta}$ into (10).

The autocorrelation structure Υ_j of ARMA(1,1):

Lag j	0	1	2	3	4	...
AR(1)	1	ρ	ρ^2	ρ^3	ρ^4	...
ARMA(1,1)	1	λ	$\lambda\rho$	$\lambda\rho^2$	$\lambda\rho^3$	...

(11)

Here

$$\lambda = \rho + \frac{\psi(1-\rho^2)}{1+2\rho\psi+\psi^2}.$$

Derivation of autocorrelations for ARMA(1,1)

- The model (assuming $\mu_x = 0$):

$$x_t - \rho x_{t-1} = v_t + \psi v_{t-1},$$

where v_t is iid, and $\sigma^2 = Ev_t^2$.

- The results in (11) follow from

$$\begin{aligned} \Upsilon_j &= \rho \Upsilon_{j-1}, \text{ for } j \geq 2. & (12) \\ \Upsilon_0 &= 1. & (13) \\ \Upsilon_1 &= \lambda, & (14) \end{aligned}$$

where

$$\lambda = \rho + \frac{\psi(1-\rho^2)}{1+2\rho\psi+\psi^2}.$$

Why (12) holds?

- Multiple $x_t - \rho x_{t-1} = v_t + \psi v_{t-1}$ by x_{t-j} $(j \geq 2)$:

$$\begin{aligned} E(x_t - \rho x_{t-1})x_{t-j} &= E(v_t + \psi v_{t-1})x_{t-j} \\ \Gamma_j - \rho\Gamma_{j-1} &= 0 \text{ (need } j \geq 2) \\ \Gamma_j &= \rho\Gamma_{j-1}, \end{aligned}$$

where $\Gamma_j = Cov(x_{t+j}, x_t)$ is the j-th autocovariance.

Why (14) holds?

- Multiple $x_t - \rho x_{t-1} = v_t + \psi v_{t-1}$ by x_{t-1} :

$$\begin{aligned} E(x_t - \rho x_{t-1})x_{t-1} &= E(v_t + \psi v_{t-1})x_{t-1} \\ \Gamma_1 - \rho\Gamma_0 &= \psi\sigma^2 \text{ (need } Ev_{t-1}x_{t-1} = \sigma^2\text{)}. \end{aligned} \tag{15}$$

- Square $x_t - \rho x_{t-1} = v_t + \psi v_{t-1}$:

$$\begin{aligned} E(x_t - \rho x_{t-1})^2 &= E(v_t + \psi v_{t-1})^2 \\ \Gamma_0 - 2\rho\Gamma_1 + \rho^2\Gamma_0 &= \sigma^2(1+\psi^2) \\ (1+\rho^2)\Gamma_0 - 2\rho\Gamma_1 &= \sigma^2(1+\psi^2). \end{aligned} \tag{16}$$

- Solving (15) and (16), you will get

$$\Gamma_0 = \sigma^2 \frac{1 + 2\rho\psi + \psi^2}{1-\rho^2}.$$

- Thus, from (15),

$$\begin{aligned} \Upsilon_1 &= \psi\sigma^2\Gamma_0^{-1} + \rho \\ &= \psi\frac{1-\rho^2}{1+2\rho\psi+\psi^2} + \rho \triangleq \lambda \end{aligned}$$

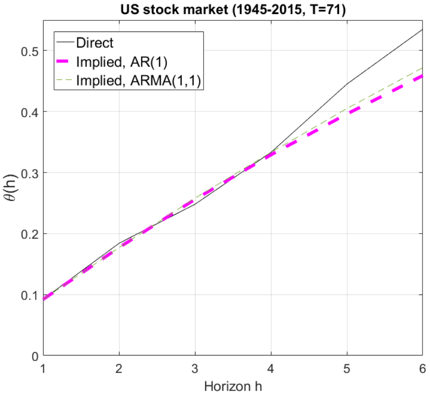
which yields (14). $\square$

US stock market (1945-2015), $T = 71$.

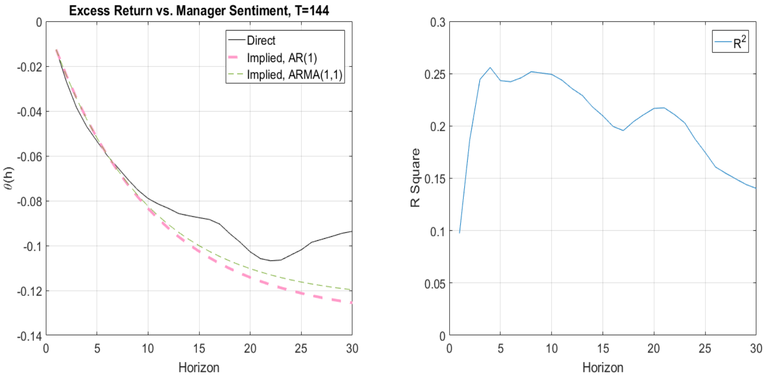
	$h=1$	$h=2$	$h=3$	$h=4$	$h=5$	$h=6$
$\theta(h)$, Direct	0.092 [2.03]	0.184 [2.91]	0.248 [3.43]	0.333 [4.13]	0.445 [4.97]	0.535 [5.57]
$\theta(h)$, Implied (AR)	0.092	0.177	0.256	0.329	0.396	0.459
$\theta(h)$, Implied (ARMA)	0.092	0.177	0.258	0.333	0.405	0.472
R^2	5.7%	11.2%	15.1%	2.08%	27.9%	33.0%

Standard t-stat is given in the bracket.

US stock market (1945-2015)



Manager sentiment, 2003/01 - 2014/12 ($T = 144$)



One way to interpret the figure on the left:

- Implied estimators (ARMA(1,1)) can be seen as 'true values' (although they are never known).
- Implied estimators (AR(1)) serve as a good approximation.
- Direct estimators are nonparametric estimators (can be very noisy for large horizons).

Manager sentiment

	$h = 1$	$h = 3$	$h = 6$	$h = 12$	$h = 24$
$\theta(h)$, Direct	-0.0126	-0.0385	-0.0598	-0.0834	-0.1041
	$[-3.90]$	$[-6.71]$	$[-6.59]$	$[-6.33]$	$[-5.22]$
$\theta(h)$, Implied (AR)	-0.0126	-0.0342	-0.0596	-0.0923	-0.1201
$\theta(h)$, Implied (ARMA)	-0.0126	-0.0344	-0.0595	-0.0906	-0.1153
R^2	9.7%	24.4%	24.2%	23.6%	18.8%

Standard t-stat is given in the bracket.

How to fit an ARMA model in Matlab?

```
>>EstMdl=estimate(arima(1,0,1),dp_4);
>>rho=EstMdl.AR{1};
>>fi=EstMdl.MA{1};
```

Note: the `estimate` command fits the demeaned series.

- Dividend-price ratio

$$dp_t = \underset{[-1.24]}{-0.21} + \underset{[19.9]}{0.94}\, dp_{t-1} + v_t - \underset{[-0.97]}{0.10}\, v_{t-1},$$

  where t-statistic is given in the bracket.

- Manager sentiment

$$MS_t = \underset{[0.41]}{0.02} + \underset{[29.8]}{0.90}\, MS_{t-1} + v_t + \underset{[1.04]}{0.13}\, v_{t-1}$$

Inference

- Horizon-h predictive regression:

$$\sum_{j=0}^{h-1} y_{t+j} = \mu(h) + \theta(h) x_{t-1} + u_t(h). \tag{17}$$

- Now consider testing hypotheses on coefficients of the regression above.
- The problem is non-trivial because the errors are overlapping.

Long-run variance (LRV)

- For a time series Z_t (denoting its mean and variance to be μ and σ^2), its *long run variance* (LRV) is defined as

$$\Omega \triangleq \lim_{T\to\infty} Var\left(\frac{1}{\sqrt{T}}\sum_{t=1}^{T} Z_t\right). \tag{18}$$

- Why might we be interested in Ω? It gives the asymptotic variance for the sample mean $\overline{Z} = T^{-1}\sum_{t=1}^{T} Z_t$ when observations Z_t are correlated.
- More precisely,

$$\overline{Z} \overset{A}{\sim} \mathcal{N}(\mu, \Omega/T), \tag{19}$$

where $\overset{A}{\sim}$ is read as "is approximately distributed as (when T is large)".
- If Z_t is iid (as if generated by random sampling in basic econometrics), then $\Omega = \sigma^2$, and (19) thus reduces to the elementary result

$$\overline{Z} \overset{A}{\sim} \mathcal{N}(\mu, \sigma^2/T). \tag{20}$$

- In a time series setting, (20) is generally incorrect (since it ignores autocovariances of Z_t), except if Z_t is a white noise.
- We will rely on (19) instead.
- An equivalent definition of LRV is

$$\Omega = \gamma_0 + 2\sum_{j=1}^{\infty} \gamma_j. \tag{21}$$

We will justify the equivalence below.

- Denote the autocovariances of Z_t as $\{\gamma_0, \gamma_1, ...\}$. An expansion shows

$$\begin{aligned} & T^{-1} Var\left(\sum_{t=1}^{T} Z_t\right) \\ = \ & \frac{1}{T}[T\gamma_0 + 2(T-1)\gamma_1 + 2(T-2)\gamma_2 + \cdots + 2\gamma_{T-1}] \\ = \ & \gamma_0 + 2\sum_{j=1}^{T-1}(1 - \frac{j}{T})\gamma_j. \end{aligned} \tag{22}$$

- (22) is an exact formula, meaning it is true for all T.
- For example, if $T = 3$,

$$\begin{aligned} & Var(Z_1 + Z_2 + Z_3) \\ = \ & 3Var(Z_1) + 2Cov(Z_1, Z_2) + 2Cov(Z_2, Z_3) + 2Cov(Z_1, Z_3) \\ = \ & 3\gamma_0 + 2\gamma_1 + 2\gamma_1 + 2\gamma_2 \\ = \ & 3\gamma_0 + 2(2\gamma_1 + \gamma_2). \end{aligned}$$

- When $T \to \infty$,

$$
\begin{aligned}
T^{-1} Var\left(\sum_{t=1}^{T} Z_t\right) &\overset{(22)}{=} \gamma_0 + 2\sum_{j=1}^{T-1}(1-\frac{j}{T})\gamma_j \\
&= \gamma_0 + 2\sum_{j=1}^{T-1}\gamma_j - \frac{2}{T}\sum_{j=1}^{T-1} j\gamma_j \\
&\to \gamma_0 + 2\sum_{j=1}^{\infty}\gamma_j,
\end{aligned}
$$

  if

$$
T^{-1}\sum_{j=1}^{T-1} j\gamma_j \to 0 \tag{23}
$$

  (γ_j shrinks sufficiently fast).

- Thus (21) holds.

- (23) is true quite broadly. ($T^{-1}\sum_{j=1}^{T-1} j\gamma_j \to 0$)
- For MA model, $\sum_{j=1}^{\infty} j\gamma_j$ is a sum of finite terms so (23) obviously holds.
- AR(1):

$$\begin{aligned}\sum_{j=1}^{T-1} j\gamma_j &= \sigma^2 \sum_{j=1}^{T-1} j\rho^j \\ &\to \sigma^2(\rho+\rho^2+\rho^3+\ldots)(1+\rho+\rho^2+\ldots) \\ &= \frac{\sigma^2\rho}{(1-\rho)^2}. \end{aligned} \tag{24}$$

$$\boxed{\begin{array}{r} \rho+\rho^2+\rho^3+\ldots+\rho^{T-1} \\ \rho^2+\rho^3+\ldots+\rho^{T-1} \\ \rho^3+\ldots+\rho^{T-1} \\ \vdots \\ \rho^{T-1} \end{array}}$$

- So $\sum_{j=1}^{\infty} j\gamma_j = \frac{\sigma^2\rho}{(1-\rho)^2}$. Thus (23) holds.

LRV of an MA(q) process

- MA(q):
$$Z_t = e_t + a_1 e_{t-1} + ... + a_q e_{t-q},$$
where $\{e_t\}$ is a serially uncorrelated sequence, and $\sigma^2 = Ee_t^2$.
- Then
$$LRV(Z_t) = \sigma^2 (1 + a_1 + ... + a_q)^2. \tag{25}$$
- We verify this important formula by considering MA(1) and MA(2) models.

- Consider MA(1):

$$Z_t = e_t + a_1 e_{t-1}.$$

- We have

$$\begin{aligned}\gamma_0 &= (1 + a_1^2)\sigma^2, \\ \gamma_1 &= a_1\sigma^2.\end{aligned}$$

- Thus $\gamma_0 + 2\gamma_1 = \sigma^2(1 + a_1)^2$.
- Consider MA(2):

$$Z_t = e_t + a_1 e_{t-1} + a_2 e_{t-2}.$$

- We have

$$\begin{aligned}\gamma_0 &= (1 + a_1^2 + a_2^2)\sigma^2 \\ \gamma_1 &= (a_1 + a_1 a_2)\sigma^2 \\ \gamma_2 &= a_2\sigma^2.\end{aligned}$$

LRV of an AR(1) process

- AR(1):

$$Z_t = \rho Z_{t-1} + e_t,$$

where $\{e_t\}$ is a serially uncorrelated sequence, and $\sigma^2 = Ee_t^2$.

- AR(1) can be written as MA(∞):

$$Z_t = e_t + \rho e_{t-1} + \rho^2 e_{t-2} + ...$$

- Extending (25), we have

$$\begin{aligned} LRV(Z_t) &= \sigma^2 (1 + \rho + \rho^2 + ...)^2 \\ &= \frac{\sigma^2}{(1-\rho)^2}. \end{aligned}$$

Horizon-h regression: Mean only

- Now come back to the horizon-h regression (17).
- Suppose for now there is no stochastic predictor

$$\sum_{j=0}^{h-1} y_{t+j} = \mu(h) + u_t(h). \tag{26}$$

- Write

$$Y_t = \mu(h) + u_t(h),$$

  where $Y_t = \sum_{j=0}^{h-1} y_{t+j}$.
- (26) should be understood as aggregation of horizon-1 models: $y_t = \mu + u_t$, where u_t is iid(0,σ_u^2).
- Then the true values are: $\mu(h) = h\mu$ and $u_t(h) = \sum_{j=0}^{h-1} u_{t+j}$.
- Question: How to test $H_0 : \mu(h) = 0$, based on the regression (26).

$$\text{'t-stat'} = \frac{T^{-1/2}\sum_{t=1}^{T-h+1} Y_t}{\sqrt{LRV(Y_t - \mu(h))}}.$$

- Note that $Y_t - \mu(h)$ is MA$(h-1)$. Thus

$$LRV(Y_t - \mu(h)) = h^2\sigma_u^2. \tag{27}$$

- This is in contrast to the standard t-stat

$$\text{'naive t-stat'} = \frac{T^{-1/2}\sum_{t=1}^{T-h+1} Y_t}{\sqrt{Var(Y_t - \mu(h))}},$$

where $Var(Y_t - \mu(h)) = Var(u_t(h)) = h\sigma_u^2$.

What do we learn from (27)?

- Standard t-test is not valid. (It's going to over-reject.)
- Standard error gets larger, as h increases.

Horizon-h regression: Slope only

- Now consider horizon-h regression (17) with mean zero:

$$\sum_{j=0}^{h-1} y_{t+j} = \theta(h)x_{t-1} + u_t(h). \tag{28}$$

- Let $Z_t = x_{t-1}u_t(h)$.
- Under the null $\beta = 0$,

$$Z_t = x_{t-1}u_t + x_{t-1}u_{t+1} + ... + x_{t-1}u_{t+h-1}.$$

- Z_t only has finite dependence: autocovariances look like $\{\gamma_0, \gamma_1, ..., \gamma_{h-1}, 0, 0, ...\}$. [called (h-1)-dependence; serial dependence exists up to order $h-1$]
- Even the summands are uncorrelated, Z_t is *not* an MA$(h-1)$ process, since $x_{t-1}u_t$ is not a lag of $x_{t-1}u_{t+1}$ (tricky).

[Cautionary note: If you thought Z_t is MA$(h-1)$, then you would have $LRV(Z_t) = h^2E(x_{t-1}^2u_t^2) \overset{Homo}{=} \frac{\sigma_v^2\sigma_u^2}{1-\rho^2}h^2$, which is incorrect, as we show below.]

- In fact, we will show that (if $\beta = 0$)

$$LRV(Z_t) = \frac{\sigma_v^2 \sigma_u^2}{1-\rho^2} \left(h + 2 \sum_{k=1}^{h-1} \rho^k (h-k) \right). \tag{29}$$

- Thus,

$$\text{'t-stat'} = \frac{T^{1/2} \widehat{\theta}(h)}{(Ex^2)^{-1} \sqrt{LRV(Z_t)}}.$$

- This is in contrast to the standard t-stat

$$\text{'naive t-stat'} = \frac{T^{1/2} \widehat{\theta}(h)}{(Ex^2)^{-1} \sqrt{Var(Z_t)}},$$

where

$$Var(Z_t) = \frac{\sigma_v^2 \sigma_u^2}{1-\rho^2} h. \tag{30}$$

How $LRV(Z_t)$ differs from $Var(Z_t)$?

$$
\boxed{
\begin{array}{ll}
\text{h=1:} & LRV(Z_t) = \frac{\sigma_v^2\sigma_u^2}{1-\rho^2} \\
 & Var(Z_t) = \frac{\sigma_v^2\sigma_u^2}{1-\rho^2} \\[1em]
\text{h=2:} & LRV(Z_t) = \frac{\sigma_v^2\sigma_u^2}{1-\rho^2}(2+2\rho) \\
 & Var(Z_t) = 2\frac{\sigma_v^2\sigma_u^2}{1-\rho^2} \\[1em]
\text{h=3:} & LRV(Z_t) = \frac{\sigma_v^2\sigma_u^2}{1-\rho^2}(3+4\rho+2\rho^2) \\
 & Var(Z_t) = 3\frac{\sigma_v^2\sigma_u^2}{1-\rho^2}
\end{array}
}
$$

Why does (29) hold?
Noting that Z_t is (h-1)-dependent,

$$
\begin{aligned}
& LRV(Z_t) \\
& \overset{(21)}{=} Ex_{t-1}^2 u_t^2(h) + 2\sum_{k=1}^{h-1} Ex_{t-1}x_{t-1+k}u_t(h)u_{t+k}(h) \\
= \quad & Ex_{t-1}^2 \left(\textstyle\sum_{j=0}^{h-1} u_{t+j}\right)^2 + 2\sum_{k=1}^{h-1} Ex_{t-1}x_{t-1+k} \underbrace{(\sum_{j=0}^{h-1} u_{t+j})(\sum_{j=0}^{h-1} u_{t+k+j})} \\
= \quad & Ex_{t-1}^2 \sigma_u^2 h + 2\sum_{k=1}^{h-1} Ex_{t-1}x_{t-1+k} \underbrace{\sigma_u^2(h-k)} \\
= \quad & \frac{\sigma_v^2}{1-\rho^2}[\sigma_u^2 h + 2\sum_{k=1}^{h-1} \rho^k \sigma_u^2 (h-k)].
\end{aligned}
$$

Thus (29) holds.
An intermediate step: (suppose $h = 3$ and $k = 1$),
$E(u_t + u_{t+1} + u_{t+2})(u_{t+1} + u_{t+2} + u_{t+3}) = 2\sigma_u^2$.

- Now we introduce a new expression for $LRV(Z_t)$, which exploits the special structure of long-horizon regression.
-
$$LRV(Z_t) = E\left[\left(\sum_{j=0}^{h-1} x_{t-1-j}\right)^2 u_t^2\right]. \qquad (31)$$
- To show it, consider $h = 2$,

$$\begin{aligned} & LRV(Z_t) \overset{(21)}{=} Var(x_{t-1}u_t(h)) + 2Cov(x_{t-1}u_t(h), x_{t-2}u_{t-1}(h)) \\ = \quad & Ex_{t-1}^2(u_t + u_{t+1})^2 + 2Ex_{t-1}x_{t-2}(u_t + u_{t+1})(u_{t-1} + u_t) \\ = \quad & Ex_{t-1}^2(u_t^2 + u_{t+1}^2) + 2Ex_{t-1}x_{t-2}u_t^2 \\ & \overset{*}{=} E(x_{t-1}^2 + x_{t-2}^2)u_t^2 + 2Ex_{t-1}x_{t-2}u_t^2 \\ = \quad & E(x_{t-1} + x_{t-2})^2 u_t^2, \end{aligned}$$

where the step $\overset{*}{=}$ uses stationarity.

- (31) holds only when $\beta = 0$. (which is fine for a test for predictability)
- The well-known Hodrick standard error is based on the expression (31).

Three standard errors

- We now have three formulas for $LRV(Z_t)$: (22), (21) and (31).

$$
\begin{aligned}
& LRV(Z_t) \\
& \overset{(22)}{=} \lim_{T\to\infty} \left[\gamma_0 + 2\textstyle\sum_{j=1}^{T-1}(1-\frac{j}{T})\gamma_j\right] \\
& \overset{(21)}{=} \gamma_0 + 2\sum_{j=1}^{\infty}\gamma_j \\
& \overset{(31)}{=} E\left[\left(\textstyle\sum_{j=0}^{h-1} x_{t-1-j}\right)^2 u_t^2\right].
\end{aligned}
$$

- These three formulas define three popular approaches to constructing standard error for $\theta(h)$:
    - Newey-West (NW),
    - Hansen-Hodrick (HH),
    - Hodrick (H).

$$\begin{aligned}
\widehat{LRV}_{NW}(Z_t) &= \widehat{\gamma}_0 + 2\sum_{j=1}^{M}(1-\frac{j}{M+1})\widehat{\gamma}_j, \\
\widehat{LRV}_{HH}(Z_t) &= \widehat{\gamma}_0 + 2\sum_{j=1}^{h-1}\widehat{\gamma}_j, \\
\widehat{LRV}_{H}(Z_t) &= T^{-1}\sum_{t=h+1}^{T}\left[\left(\sum_{j=0}^{h-1}x_{t-1-j}\right)^2\widehat{u}_t^2\right],
\end{aligned}$$

  where M is a truncation parameter (needed to be determined in applications), $\widehat{\gamma}_j$'s are sample autocovariances of $x_{t-1}\widehat{u}_t(h)$, $\widehat{u}_t$ is the horizon-one regression residual, and $\widehat{u}_t(h)$ is the horizon-h regression residual.

- These are for the model (28), with no intercept.

- We have just discussed two simplified models (26) and (28):

$$\sum_{j=0}^{h-1} y_{t+j} = \mu(h) + u_t(h),$$

$$\sum_{j=0}^{h-1} y_{t+j} = \theta(h)x_{t-1} + u_t(h).$$

- Neither of them is actually used in practice. We use them here for easy calculations, and more importantly, for the purpose of understanding how overlapping observations could cause trouble in inference.
- *Main lesson*: Standard test statistic (naive t-stat or White t-stat) should not be used, since it ignores serial dependence. Instead, LRV-based test statistic should be used.

Testing in action: Three standard errors

- We now finally go back the model (17), with the intercept:

$$\sum_{j=0}^{h-1} y_{t+j} = \mu(h) + \theta(h) x_{t-1} + u_t(h).$$

- The formulas given above (for the model with no intercept) needs to be changed; just change x_{t-1} to $\widetilde{x}_{t-1}$, where $\widetilde{x}_{t-1} = x_{t-1} - \overline{x}$ with $\overline{x} = (T-1)^{-1} \sum_{t=2}^{T} x_t$.

- Thus $Z_t = \widetilde{x}_{t-1} u_t$, and

$$\widehat{LRV}_{NW}(Z_t) = \widehat{\gamma}_0 + 2\sum_{j=1}^{M}(1 - \frac{j}{M+1})\widehat{\gamma}_j, \quad (32)$$

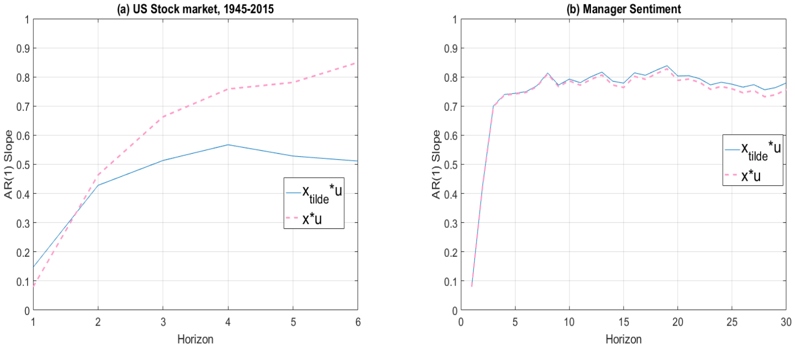
$$\widehat{LRV}_{HH}(Z_t) = \widehat{\gamma}_0 + 2\sum_{j=1}^{h-1}\widehat{\gamma}_j, \quad (33)$$

$$\widehat{LRV}_{H}(Z_t) = T^{-1}\sum_{t=h+1}^{T}\left[\left(\sum_{j=0}^{h-1}\widetilde{x}_{t-1-j}\right)^2 \widehat{u}_t^2\right], \quad (34)$$

where $\widehat{\gamma}_j$'s are sample autocovariances of $\widetilde{x}_{t-1}\widehat{u}_t(h)$.

- $$\text{t-stat} = \frac{T^{1/2}\widehat{\theta}(h)}{(T^{-1}\sum_{t=2}^{T}\widetilde{x}_{t-1}^2)^{-1}\sqrt{\widehat{LRV}(Z_t)}},$$

where $\widehat{LRV}(Z_t)$ is any one of (32)-(34).



- Blue line shows how the serial correlation increases with the horizon.
- On panel (b), blue and red lines are very close (because MS is already standardized as mean zero).

US stock market (1945-2015), $T = 71$.

	$h=1$	$h=2$	$h=3$	$h=4$	$h=5$	$h=6$
$\theta(h)$, Direct	0.092 [2.03]	0.184 [2.91]	0.248 [3.43]	0.333 [4.13]	0.445 [4.97]	0.535 [5.57]
White t-stat	2.18	2.98	3.51	4.85	7.08	9.36
Hodrick t-stat	2.18	2.16	1.88	1.85	1.96	1.96

Standard t-stat is given in the bracket.

Manager sentiment

	$h=1$	$h=3$	$h=6$	$h=12$	$h=24$
$\theta(h)$, Direct	-0.0126	-0.0385	-0.0598	-0.0834	-0.1041
	$[-3.90]$	$[-6.71]$	$[-6.59]$	$[-6.33]$	$[-5.22]$
White t-stat	-3.57	-5.97	-5.57	-5.47	-5.50
Hodrick t-stat	-3.57	-4.29	-3.61	-2.63	-1.99

Standard t-stat is given in the bracket.

The dependent variable: When relevant?

- We have studied the horizon-h predictive regression:

$$\sum_{j=0}^{h-1} y_{t+j} = \mu(h) + \theta(h) x_{t-1} + u_t(h).$$

- The most common dependent variable $Y_t \triangleq \sum_{j=0}^{h-1} y_{t+j}$ is the change of an economic variable s_t over h periods:

$$Y_t = s_{t+h-1} - s_{t-1} = \Delta s_{t+h-1} + \Delta s_{t+h-2} + ... + \Delta s_t,$$

where $\Delta s_{t+1} = s_{t+1} - s_t$.

- Thus

$$y_t = \Delta s_t.$$

- s_t can be
  - log asset price (so Δs_t is the log return *without dividends*)
  - log dividends (e.g. (42))
  - nominal exchange rate
  - log DGP
- However, we need to be careful if y_t is the stock return *with dividends*. (Then $\sum_{j=0}^{h-1} y_{t+j}$ is not the return over h periods.)

Why uses DP as predictor: A present value model

- We now describe a present value model (or discounted cash flow model) for the asset price.
- This model is also useful to answer the question why the dividend-price ratio is most often used as the predictor for returns.
- Note that r_{t+1} is defined by price and dividend variables in levels, $r_{t+1} = \log(P_{t+1} + D_{t+1}) - \log P_t$.

- Their relation in logs needs a log-linear approximation, given below

$$r_{t+1} \approx k - \alpha dp_{t+1} + \Delta d_{t+1} + dp_t, \tag{35}$$

where $dp_t = d_t - p_t$ is the log dividend-price ratio, and k and α are two constants (functions of the longrun average log dividend-price ratio $\overline{dp}$):

$$\begin{aligned} \alpha &= 1/[1 + \exp(\overline{dp})], \\ k &= \log[1 + \exp(-\overline{dp})] + \rho\overline{dp}. \end{aligned}$$

- α is the discount factor. The empirical value of $\overline{dp}$ is close to -3.2 [i.e. log(0.04)] for US data. Thus α is close to 0.96.

- Iterating forward on Equation (35), letting F be the forward operator $(Fx_t = x_{t+1})$

$$(1 - \alpha F) dp_t = -k + r_{t+1} - \Delta d_{t+1}.$$

Thus

$$dp_t = -k(1 + \alpha + \alpha^2 + ...) + \sum_{i=0}^{\infty} \alpha^i (r_{t+1+i} - \Delta d_{t+1+i}). \tag{36}$$

- Equation (36) is one of the central tenets of the return predictability literature, the so-called present-value equation.
- Present-value equation shows that "dp_t ought to forecast either future returns r_{t+1} or future dividend growth rates Δd_{t+1}, or both".

Why (35) holds?

- The key step is the approximation of $\log(1 + D_{t+1}/P_{t+1})$.
- Let $f(\theta) = \log(1 + \exp(\theta))$,

$$\begin{aligned} & \log(1 + D_{t+1}/P_{t+1}) \\ = & \log(1 + \exp(dp_{t+1})) \\ = & f(dp_{t+1}) \\ \approx & -\log\alpha + (dp_{t+1} - \theta)(1 - \alpha), \end{aligned} \tag{37}$$

  where $\theta = \overline{dp} = mean(dp_t)$, $\alpha = 1/[1 + \exp(\theta)]$.
- In the step of log linear approximation, we use

$$f(dp_{t+1}) \approx f(\theta) + f'(\theta)(dp_{t+1} - \theta),$$

  where $f'(\theta) = \exp(\theta)/[1 + \exp(\theta)]$.

Then

$$
\begin{aligned}
r_{t+1} &= \log(P_{t+1} + D_{t+1}) - \log P_t \\
&= \log P_{t+1}(1 + D_{t+1}/P_{t+1}) - \log P_t \\
&= p_{t+1} - p_t + \log(1 + D_{t+1}/P_{t+1}) \\
&\overset{(37)}{\approx} p_{t+1} - p_t - \log\alpha + (dp_{t+1} - \theta)(1-\alpha) \\
&= \underbrace{-\theta(1-\alpha) - \log\alpha}_{=k} + p_{t+1} - p_t + (1-\alpha)dp_{t+1} \\
&= k + p_{t+1} - p_t - \Delta d_{t+1} + \Delta d_{t+1} + (1-\alpha)dp_{t+1} \\
&= k - dp_{t+1} + dp_t + \Delta d_{t+1} + (1-\alpha)dp_{t+1} \\
&= k - \alpha dp_{t+1} + \Delta d_{t+1} + dp_t, \qquad (38)
\end{aligned}
$$

as desired. Thus (35) holds.

Todd suggested the coefficient should be one?

- Log-linear relationship (35):

$$r_{t+1} = k \underbrace{-\alpha dp_{t+1} + \Delta d_{t+1}}_{\text{"error"}} + dp_t. \tag{39}$$

- It is true that the coefficient of dp_t is one in (39). However, a regression needs to make sure the regression error is orthogonal to the predictor dp_t, which is not true in (39).
- Using an AR(1) model for dp_t :

$$dp_{t+1} = \mu_{dp} + \rho dp_t + v_{t+1},$$

and a regression for Δd_{t+1} (dividend growth):

$$\Delta d_{t+1} = \mu_{dg} + \beta_{dg} dp_t + w_{t+1},$$

we have

$$r_{t+1} = k - \alpha\mu_{dp} + \mu_{dg} + \underbrace{(1 - \alpha\rho + \beta_{dg})}_{\text{slope}} dp_t \underbrace{-\alpha v_{t+1} + w_{t+1}}_{\text{error}}. \tag{40}$$

- Then the popular return/log-dividend-price-ratio regression is essentially estimating $1 - \alpha\rho + \beta_{dg}$.

In our dataset of US stock market, 1945-2015, annual observations,

$$\begin{aligned}
mean(DP_t) &= 0.0339 \\
\overline{dp} &= mean(dp_t) = -3.48 \\
\alpha &= 1/[1+\exp(\overline{dp})] = 0.97 \\
\rho &= 0.925 \\
\beta_{dg} &= \underset{[\text{tstat: } -0.54]}{-0.0104} \\
1 - \alpha\rho + \beta_{dg} &= 0.092 \;\; \text{(true coefficient).}
\end{aligned}$$

Derivation of (40).

$$
\begin{aligned}
r_{t+1} &= k + dp_t - \alpha dp_{t+1} + \Delta d_{t+1} \\
&= k + dp_t - \alpha(\mu_{dp} + \rho dp_t + v_{t+1}) + \mu_{dg} + \beta_{dg} dp_t + w_{t+1} \\
&= k - \alpha\mu_{dp} + \mu_{dg} + \underbrace{(1 - \alpha\rho + \beta_{dg})} dp_t - \alpha v_{t+1} + w_{t+1}.
\end{aligned}
$$

Extending (35) to multiple-period returns

- Multi-period returns:

$$r_{t:t+h} = \log(P_{t+h} + D_{t+h}) - \log P_t.$$

- (37) gives the approximation of $\log(1 + D_{t+1}/P_{t+1})$, as a linear function of dp_{t+1}.
- Extending (37) to a multiple horizon:

$$\log(1 + D_{t+h}/P_{t+h}) \approx -\log\alpha + (dp_{t+h} - \theta)(1 - \alpha), \tag{41}$$

where $\theta = \overline{dp} = mean(dp_t)$, $\alpha = 1/[1 + \exp(\theta)]$. Assuming dp_t is stationary, the parameters still have the same meanings.

- Similar to (38), we have

$$
\begin{aligned}
r_{t:t+h} &= \log(P_{t+h} + D_{t+h}) - \log P_t \\
&= \log P_{t+h}(1 + D_{t+h}/P_{t+h}) - \log P_t \\
&= p_{t+h} - p_t + \log(1 + D_{t+h}/P_{t+h}) \\
&\overset{(41)}{\approx} p_{t+h} - p_t - \log\alpha + (dp_{t+h} - \theta)(1-\alpha) \\
&= \underbrace{-\theta(1-\alpha) - \log\alpha}_{=k} + p_{t+h} - p_t + (1-\alpha)dp_{t+h} \\
&= k + p_{t+h} - p_t - (d_{t+h} - d_t) + (d_{t+h} - d_t) + (1-\alpha)dp_{t+h} \\
&= k - dp_{t+h} + dp_t + (d_{t+h} - d_t) + (1-\alpha)dp_{t+h} \\
&= k - \alpha dp_{t+h} + \sum_{i=0}^{h} \Delta d_{t+i} + dp_t, \qquad (42)
\end{aligned}
$$

where we use $d_{t+h} - d_t = \sum_{i=0}^{h} \Delta d_{t+i}$.

- The important relationship (42) shows that "over a multiple horizon, dp_t ought to forecast either future returns $r_{t:t+h}$ or future dividend growth rates $d_{t+h} - d_t$, or both".
- So it's common to run two regressions:
  - Regression of $r_{t:t+h}$ on dp_t
  - Regression of $\sum_{i=0}^{h} \Delta d_{t+i}$ on dp_t

Two ways of defining horizon-h returns

- The return from $t-1$ to $t+h-1$:

$$r_t(h) = r_{t-1:t+h-1} = \log(P_{t+h-1} + D_{t+h-1}) - \log P_{t-1}. \tag{43}$$

- It is customary to use the notation $r_t(h)$, which reads as "horizon-h return from $t-1$ to $t+h-1$". (The index t in $r_t(h)$ is a little strange, but it reduces the standard return definition when $h = 1$.)
- Method 1: use the definition (43).
- Method 2: the summation method:

$$r_t^S(h) = \sum_{i=0}^{h-1} r_{t+i},$$

where S denotes 'sum'.
- These two methods are equivalent if dividends are *not* included.

- However, these two methods give different results when dividends are included.
- We can show that (for $h \geq 2$)

$$r_t^S(h) - r_t(h) = \sum_{i=0}^{h-2} M_{t+i}, \tag{44}$$

  where $M_t = \log(1 + D_t / P_t)$.
- i.e. $r_t^S(h)$ is always larger than $r_t(h)$.

Now we show (44).

- $h = 1:$

$$r_t(1) = r_t^S(1) = r_t = \log(P_t + D_t) - \log P_{t-1}.$$

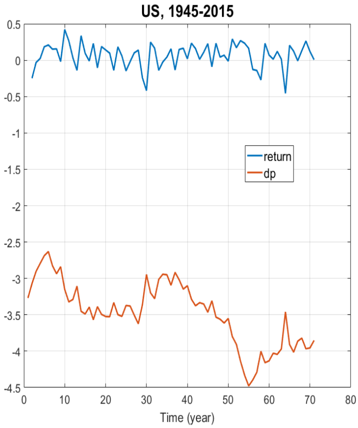
- $h = 2:$

$$\begin{aligned} r_t^S(2) &= r_{t+1} + r_t \\ &= [\log(P_{t+1} + D_{t+1}) \underbrace{- \log P_t] + [\log(P_t + D_t)}_{=M_t} - \log P_{t-1}] \\ &= r_t(2) + M_t. \end{aligned}$$

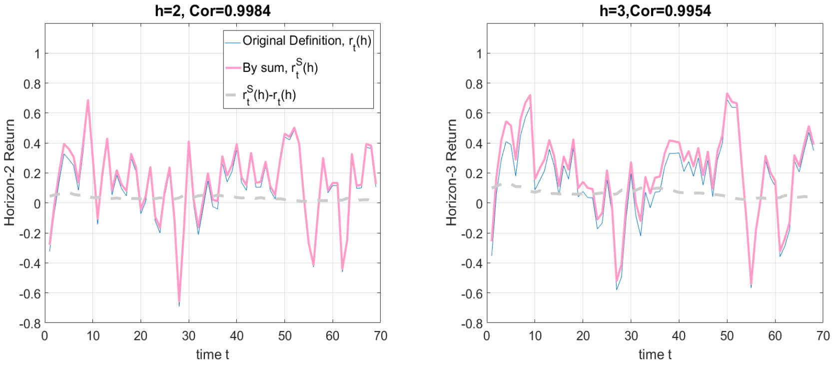
- $h = 3:$

$$\begin{aligned} & r_t^S(3) \\ = \; & r_{t+2} + r_{t+1} + r_t \\ = \; & [\log(P_{t+2} + D_{t+2}) \\ & \underbrace{- \log P_{t+1}] + [\log(P_{t+1} + D_{t+1})}_{=M_{t+1}} \underbrace{- \log P_t] + [\log(P_t + D_t)}_{=M_t} - \log P_{t-1}]. \end{aligned}$$

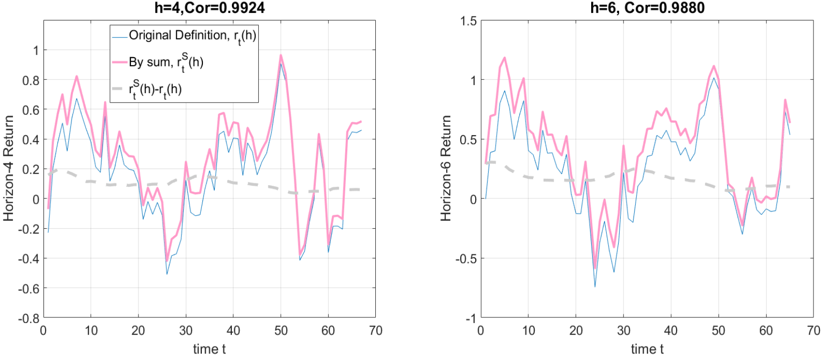
Regression data plots



Two ways of defining horizon-h returns



The gray line (graph on the left) shows $M_t = \log(1 + D_t / P_t)$.
The gray line (graph on the right) shows $M_t + M_{t+1}$.



Thus $r_t^S(h)$ is not exactly the return over h periods (which is $r_t(h)$ by definition), but is about 99% correlated with it.