Out-of-Sample Analysis

Financial Econometrics (M524) Lecture Notes

Ke-Li Xu

Indiana University

November 8, 2022

Contents

The framework

- Let y be the variable of forecasting interest.
- Question: which predictor(s) can predict y better.
- We will make P predictions of y: $\widehat{y}$
- The data index (letting $T = R + P$)

$$t = \underbrace{1, 2, ..., R}_{\text{Initial Est. sample}}, \underbrace{R+1, R+2, ..., R+P}_{\text{Prediction sample}}.$$

Data:	y_1	y_2	...	y_{R-1}	y_R	y_{R+1}	y_{R+2}	...	y_{R+P-1}	y_{R+P}
Prediction:						$\widehat{y}_{R+1}$	$\widehat{y}_{R+2}$	...	$\widehat{y}_{R+P-1}$	$\widehat{y}_{R+P}$

- 'True values' are in red.

- There's evidence that simple models can predict better than more complex models, as complex models contain more parameters thus induce more uncertainty.
- We will consider three models:
  - zero forecast (0 parameter)
  - historical mean (1 parameters)
  - predictive regression (at least two parameters)
- A very simple method of forming $\widehat{y}$ is based on the *historical mean (HM).*
- The forecast value of y_{t+1} is:

$$\widehat{y}_{t+1}^{HM} = t^{-1} \sum_{s=1}^{t} y_s .$$

Historical mean

- To obtain $\widehat{y}_{R+1}$:

Data:	$\boxed{y_1}$	$\boxed{y_2}$	...	$\boxed{y_{R-1}}$	$\boxed{y_R}$	y_{R+1}	y_{R+2}	$\cdots$	y_{R+P-1}	y_{R+P}
Prediction:						$\widehat{y}_{R+1}^{HM}$				

- To obtain $\widehat{y}_{R+2}$:

Data:	$\boxed{y_1}$	$\boxed{y_2}$	...	$\boxed{y_{R-1}}$	$\boxed{y_R}$	$\boxed{y_{R+1}}$	y_{R+2}	$\cdots$	y_{R+P-1}	y_{R+P}
Prediction:							$\widehat{y}_{R+2}^{HM}$			

- To obtain $\widehat{y}_{R+P}$:

Data:	$\boxed{y_1}$	$\boxed{y_2}$	...	$\boxed{y_{R-1}}$	$\boxed{y_R}$	$\boxed{y_{R+1}}$	$\boxed{y_{R+2}}$	$\cdots$	$\boxed{y_{R+P-1}}$	y_{R+P}
Prediction:										$\widehat{y}_{R+P}^{HM}$

Observations in $\boxed{\text{boxes}}$ are those used for a forecast.
(referred to as *expanding* or *recursive* scheme).

Application: Manager Sentiment example

Comparison of three methods

$$R^2_{OOS} = 4.37\%$$

Matlab command:

```
[MSE_0,MSE_hm,MSE,R2,yplot_true,yplot_hm,yplot_direct]
=oos_rsq_m524(er,MS,1,0.5);
```

When there are predictors...

The model is of the form

$$y_{t+1} = x_t' \beta + u_{t+1},$$

where y is the variable of forecasting interest, and x contains predictors (including intercept).

Data:	$\boxed{x_1}$	$\boxed{x_2}$	...	$\boxed{x_{R-1}}$	x_R	x_{R+1}	x_{R+2}	...	x_{R+P-1}	x_{R+P}
Data:	y_1	$\boxed{y_2}$	...	$\boxed{y_{R-1}}$	$\boxed{y_R}$	y_{R+1}	y_{R+2}	...	y_{R+P-1}	y_{R+P}
Prediction:						$\widehat{y}_{R+1}$	$\widehat{y}_{R+2}$	...	$\widehat{y}_{R+P-1}$	$\widehat{y}_{R+P}$

- For these P predictions $\widehat{y}$, we need to estimate the model P times, *one for each*:

$$\widehat{y}_{t+1} = x_t' \widehat{\beta}_t,$$

  where $\widehat{\beta}_t$ is obtained using observations $\{y_{i+1}, x_i : i = 1, ..., t-1\}$ (using the information up to t).

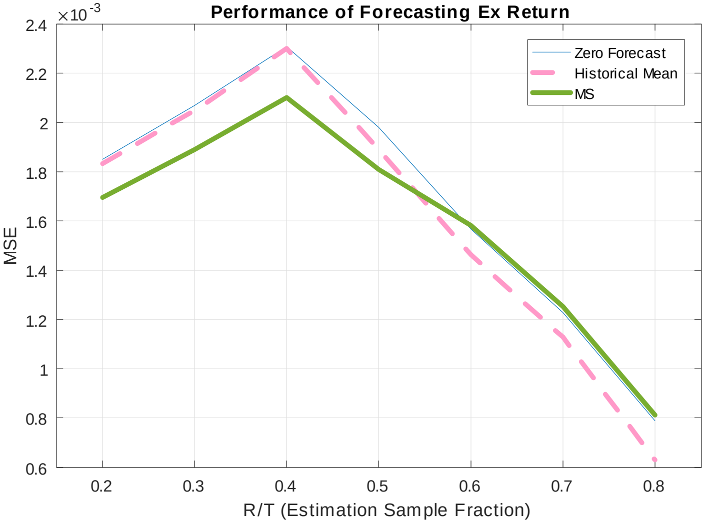
- E.g. to predict the first value, y_{R+1}, the observations in boxes are used for parameter estimation (for all three schemes which will be discussed below).

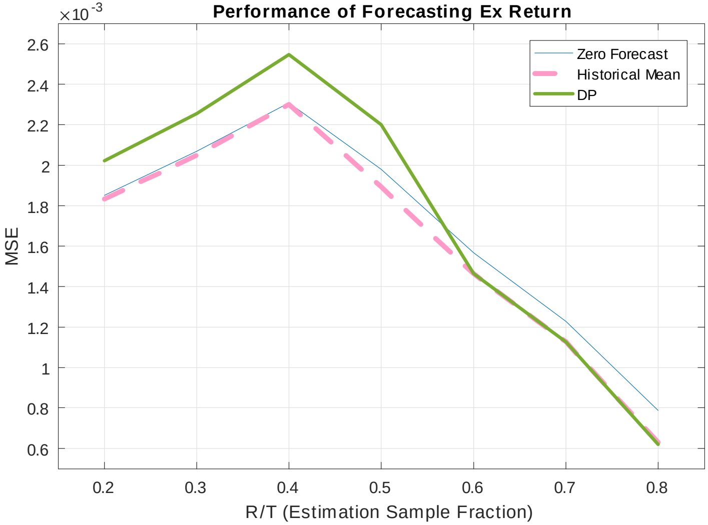
How to evaluate a forecast method as good?

- A commonly used measure is mean squared error (MSE):

$$MSE = \frac{1}{P} \sum_{t=R+1}^{T} (y_t - \widehat{y}_t)^2,$$

for a candidate method (which generates forecasts $\{\widehat{y}_t : t = R+1, ..., T\}$).


Performance of Forecasting Ex Return
×10^-3
2.4
2.2
2
1.8
1.6
1.4
1.2
1
0.8
0.6
MSE
0.2
0.3
0.4
0.5
0.6
0.7
0.8
R/T (Estimation Sample Fraction)
Zero Forecast
Historical Mean
MS


Performance of Forecasting Ex Return
×10^-3
2.6
2.4
2.2
2
1.8
1.6
1.4
1.2
1
0.8
0.6
MSE
0.2
0.3
0.4
0.5
0.6
0.7
0.8
R/T (Estimation Sample Fraction)
Zero Forecast
Historical Mean
DP

Out-of-sample (OOS) R-square

- $\widehat{y}_t^{HM}$ is very often used as a benchmark.
- For a candidate prediction method $\widehat{y}$ (e.g. predictive regression), the out-of-sample R^2 is a measure of prediction performance (relative to $\widehat{y}_t^{HM}$).
- Out-of-sample R^2 :

$$\begin{aligned} R^2_{OOS} &\triangleq 1 - \frac{\sum_{t=R+1}^{T}(y_t - \widehat{y}_t)^2}{\sum_{t=R+1}^{T}(y_t - \widehat{y}_t^{HM})^2} \\ &= 1 - \frac{MSE(\widehat{y}_t)}{MSE(\widehat{y}_t^{HM})}. \end{aligned}$$

- A catch for R^2_{OOS} is that R^2_{OOS} can be negative.
- If this is true

$$R^2_{OOS} < 0 \Longleftrightarrow \sum_{t=R+1}^{T} (y_t - \widehat{y}_t^{HM})^2 < \sum_{t=R+1}^{T} (y_t - \widehat{y}_t)^2,$$

  then it means $\widehat{y}_t^{HM}$ predicts better than the candidate method $\widehat{y}_t$.
- Otherwise (if $R^2_{OOS} > 0$), it means $\widehat{y}_t$ predicts better than $\widehat{y}_t^{HM}$.

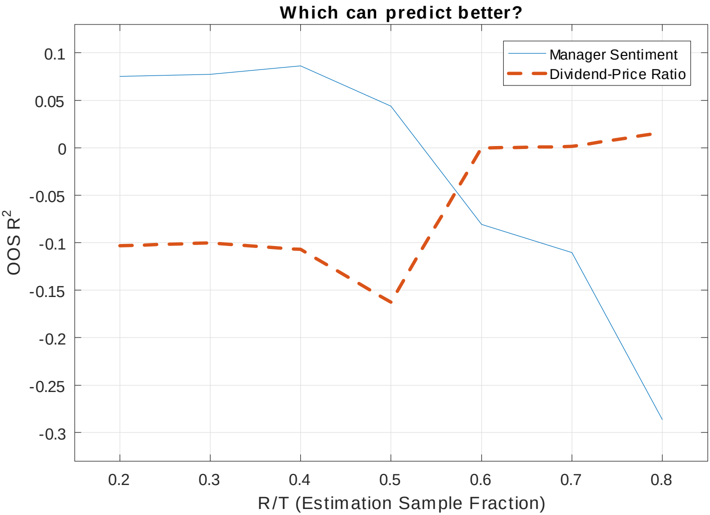
Comparison between in-sample R^2_{IS} and out-of-sample R^2_{OOS}

Suppose we use x_t as the predictor

$$R^2_{OOS} \triangleq 1 - \frac{\sum_{t=R+1}^{T}(y_t - x'_{t-1}\widehat{\beta}_{t-1})^2}{\sum_{t=R+1}^{T}(y_t - \widehat{y}_t^{HM})^2}$$

$$R^2_{IS} \triangleq 1 - \frac{\sum_{t=2}^{T}(y_t - x'_{t-1}\widehat{\beta})^2}{\sum_{t=2}^{T}(y_t - \overline{y})^2}$$

We must have $0 \leq R^2_{IS} \leq 1$. But this may not always be true for R^2_{OOS}.


Which can predict better?
Manager Sentiment
Dividend-Price Ratio
OOS R²
0.1
0.05
0
-0.05
-0.1
-0.15
-0.2
-0.25
-0.3
0.2
0.3
0.4
0.5
0.6
0.7
0.8
R/T (Estimation Sample Fraction)

Multiple horizons

Like before, defind $y_{t+1}(h) = y_{t+1} + ... + y_{t+h}$, for $t = 1, ..., T - h$.
Consider $h = 2$.

y_1	y_2	$\cdots$	y_{R-2}	y_{R-1}	y_R	y_{R+1}	$\cdots$	y_{T-1}	y_T
$y_1(2)$	$y_2(2)$	...	$y_{R-2}(2)$	$y_{R-1}(2)$	$y_R(2)$	$y_{R+1}(2)$	...	$y_{T-1}(2)$	
						$\widehat{y}_{R+1}(2)$	...	$\widehat{y}_{T-1}(2)$	

- We will make $P - 1$ predictions $(\widehat{y}(2))$; 'true values' are in red.
- In general, the number of predictions is $P - h + 1$.

Historical mean (multiple-horizon)

$$\widehat{y}_{t+1}^{HM}(h) = (t-h+1)^{-1} \sum_{s=1}^{t-h+1} y_s(h).$$

- To obtain $\widehat{y}_{R+1}(2)$:

y_1	$\cdots$	y_{R-2}	y_{R-1}	y_R	y_{R+1}	$\cdots$	y_{T-2}	y_{T-1}	y_T
$\boxed{y_1(2)}$	$\ldots$	$\boxed{y_{R-2}(2)}$	$\boxed{y_{R-1}(2)}$	$y_R(2)$	$y_{R+1}(2)$	$\ldots$	$y_{T-2}(2)$	$y_{T-1}(2)$	
					$\widehat{y}_{R+1}(2)$	$\ldots$	$\widehat{y}_{T-2}(2)$	$\widehat{y}_{T-1}(2)$	

Note that $y_R(2) = y_R + y_{R+1}$ is not available at time R.

- To obtain $\widehat{y}_{T-1}(2)$:

y_1	$\cdots$	y_{R-2}	y_{R-1}	y_R	y_{R+1}	$\cdots$	y_{T-2}	y_{T-1}
$\boxed{y_1(2)}$	$\ldots$	$\boxed{y_{R-2}(2)}$	$\boxed{y_{R-1}(2)}$	$\boxed{y_R(2)}$	$\boxed{y_{R+1}(2)}$	$\boxed{\cdots}$	$y_{T-2}(2)$	$y_{T-1}(2)$
					$\widehat{y}_{R+1}(2)$	$\ldots$	$\widehat{y}_{T-2}(2)$	$\widehat{y}_{T-1}(2)$

(Multiple horions) When there are predictors...

- For $t = 1, ..., T - h$.
$$y_{t+1}(h) = x_t'\theta(h) + u_{t+1}.$$
- Idea: fit the model in the estimation sample, and construct the fitted value of $y_{t+1}(h)$ using the *most recent* x_t.
- In general, the number of predictions is $P - h + 1$, and the estimation sample size: $R - h$.

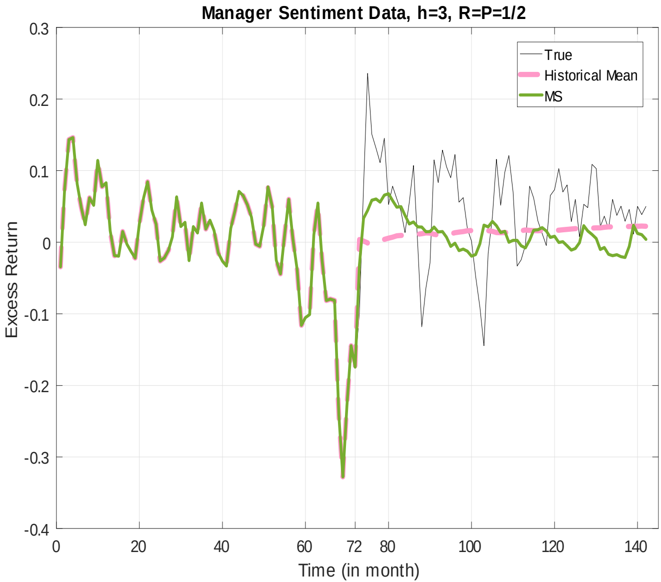
- E.g. Horizon-2 forecast ($h = 2$):

$\boxed{x_1}$	$\boxed{x_2}$	$\ldots$	$\boxed{x_{R-2}}$	x_{R-1}	$\underbrace{x_R}$	x_{R+1}	$\ldots$	x_{T-1}	x_T
y_1	y_2	$\ldots$	y_{R-2}	y_{R-1}	y_R	y_{R+1}	$\ldots$	y_{T-1}	y_T
$y_1(2)$	$\boxed{y_2(2)}$	$\ldots$	$\boxed{y_{R-2}(2)}$	$\boxed{y_{R-1}(2)}$	$y_R(2)$	$y_{R+1}(2)$	$\ldots$	$y_{T-1}(2)$	
						$\widehat{y}_{R+1}(2)$	$\ldots$	$\widehat{y}_{T-1}(2)$	

- Now we have $P - 1$ predictions. (true values in red)
- E.g.

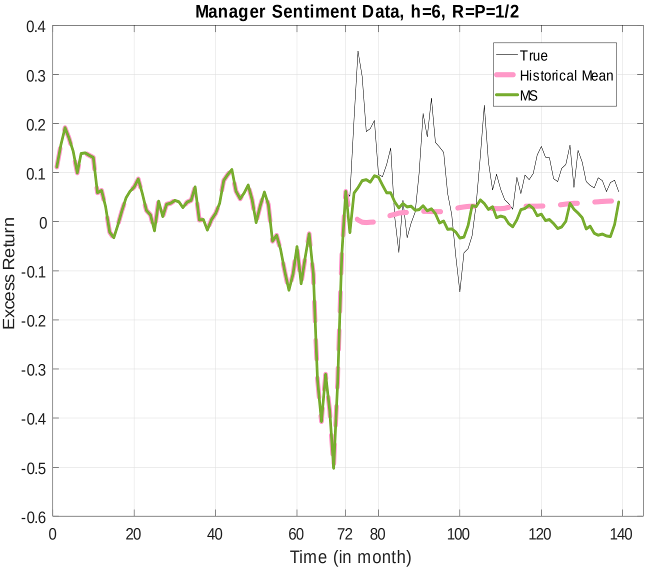
$$\widehat{y}_{R+1}(2) = x_R' \widehat{\theta}_R,$$

where $\widehat{\beta}_R$ uses the info up to R ($y_{R-1}(2)$ is the most observation used).
- Be careful that the predictor value used in the forecast is x_R.
- The observation x_{R-1} is not used (when you make your first forecast).
- Estimation sample size: $R - 2$. (estimation sample in boxes)



Manager Sentiment Data, h=3, R=P=1/2
0.3
0.2
0.1
0
-0.1
-0.2
-0.3
-0.4
Excess Return
0
20
40
60
72
80
100
120
140
Time (in month)
True
Historical Mean
MS

$R^2_{OOS} = 10.4\%$

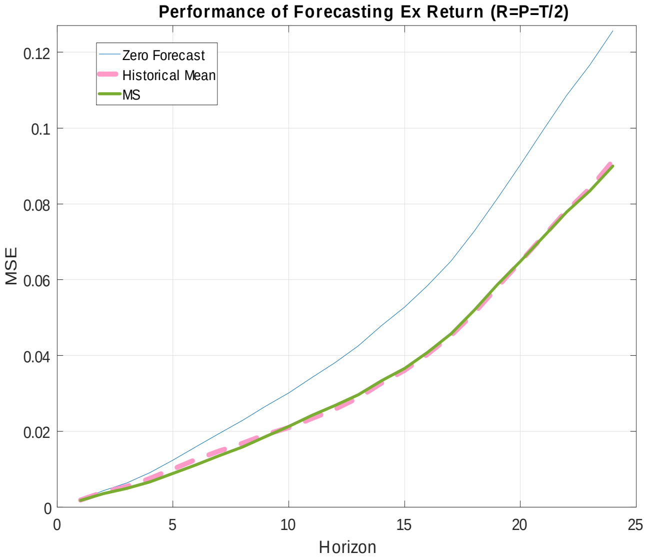


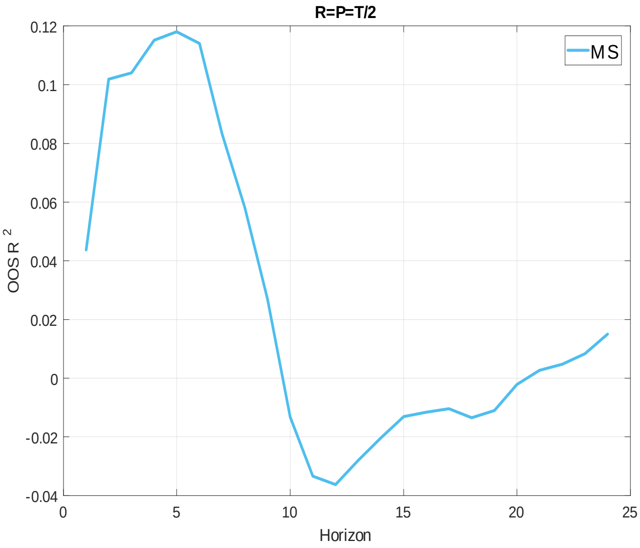
Manager Sentiment Data, h=6, R=P=1/2
0.4
0.3
0.2
0.1
0
-0.1
-0.2
-0.3
-0.4
-0.5
-0.6
Excess Return
0
20
40
60
72
80
100
120
140
Time (in month)
True
Historical Mean
MS

$R^2_{OOS} = 11.4\%$

Manager Sentiment Data, h=12, R=P=1/2
True
Historical Mean
MS
Excess Return
Time (in month)
0.6
0.4
0.2
0
-0.2
-0.4
-0.6
0
20
40
60
72
80
100
120
140

$R^2_{OOS} = -3.62\%$


Performance of Forecasting Ex Return (R=P=T/2)
Zero Forecast
Historical Mean
MS
MSE
0.12
0.1
0.08
0.06
0.04
0.02
0
0
5
10
15
20
25
Horizon


R=P=T/2
MS
OOS R 2
Horizon
0.12
0.1
0.08
0.06
0.04
0.02
0
-0.02
-0.04
0
5
10
15
20
25

Implied estimator-based forecast

- We first review some previous coverage.
- Consider the predictive regression with one predictor

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

- 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 estimators $\widehat{\mu}(h)$ and $\widehat{\theta}(h)$: OLS of $\sum_{j=0}^{h-1} y_{t+j}$ on an intercept and x_{t-1}.

- Implied estimators of $\mu(h)$ and $\theta(h)$ need a model for x_t. We consider the AR(1): $(x_t - Ex) = \rho(x_{t-1} - Ex) + v_t$.
- Implied estimators:

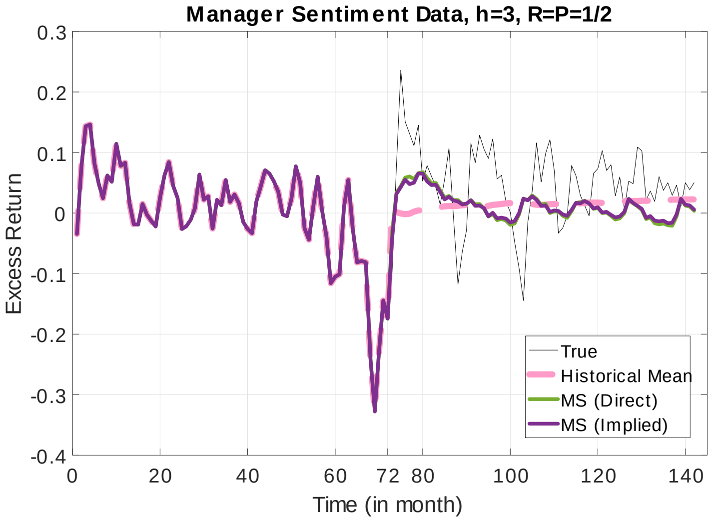
$$\begin{aligned} \widehat{\mu}^{IM}(h) &= h\widehat{\mu}_y + \left(h - \textstyle\sum_{j=0}^{h-1} \widehat{\rho}^j\right) \widehat{\beta}\widehat{Ex}, \qquad (3) \\ \widehat{\theta}^{IM}(h) &= \sum_{j=0}^{h-1} \widehat{\rho}^j \widehat{\beta}, \qquad (4) \end{aligned}$$

- In (3), we can use a nonparametric estimator for $\widehat{Ex}$, or a parametric one:

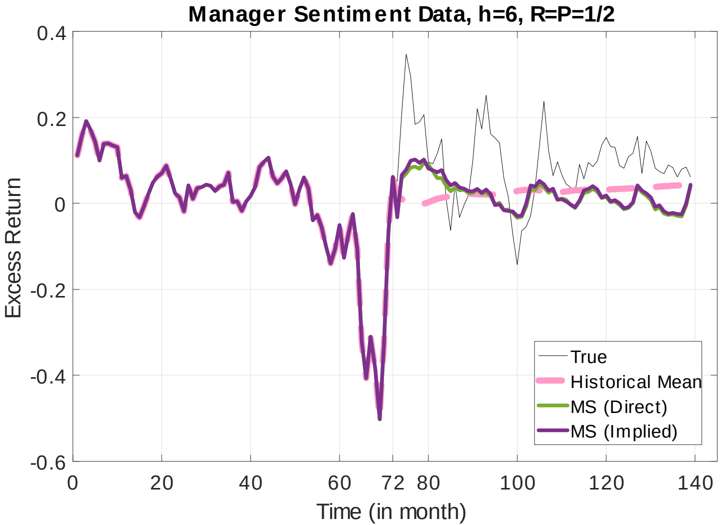
$$\widehat{Ex} = \frac{\widehat{\mu}_x}{1 - \widehat{\rho}}.$$

(based on AR(1) implication)

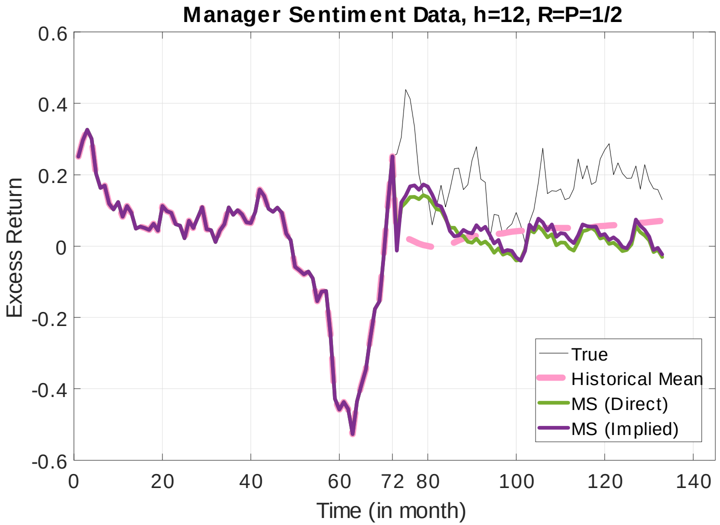
- Be careful that even the intercept is estimated differently: $\widehat{\mu}(h)$ is different from $\widehat{\mu}^{IM}(h)$.



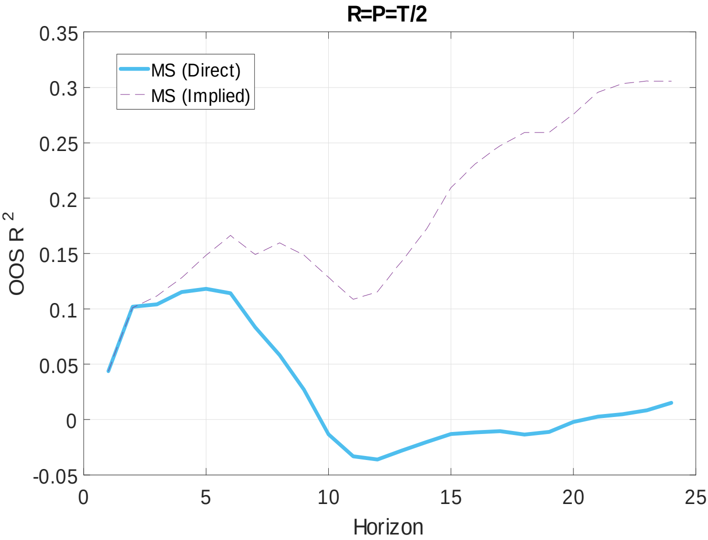
$$\left(R^2_{OOS,Direct}, R^2_{OOS,Implied}\right) = (10.4\%, 11.2\%)$$

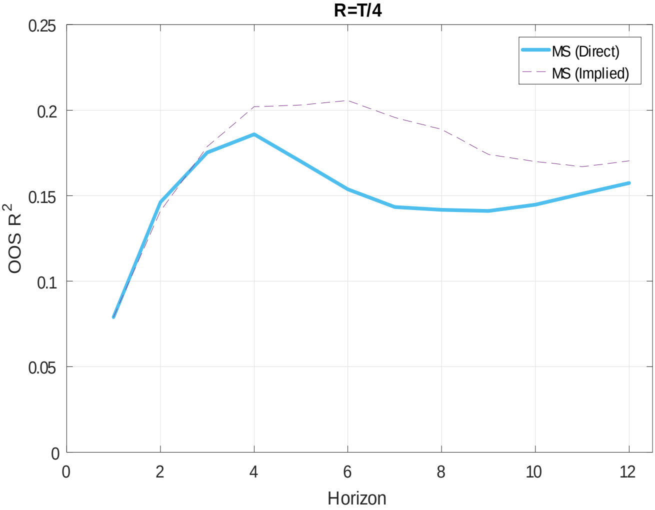


$$\left(R^2_{OOS,Direct}, R^2_{OOS,Implied}\right) = (11.4\%, 16.6\%)$$

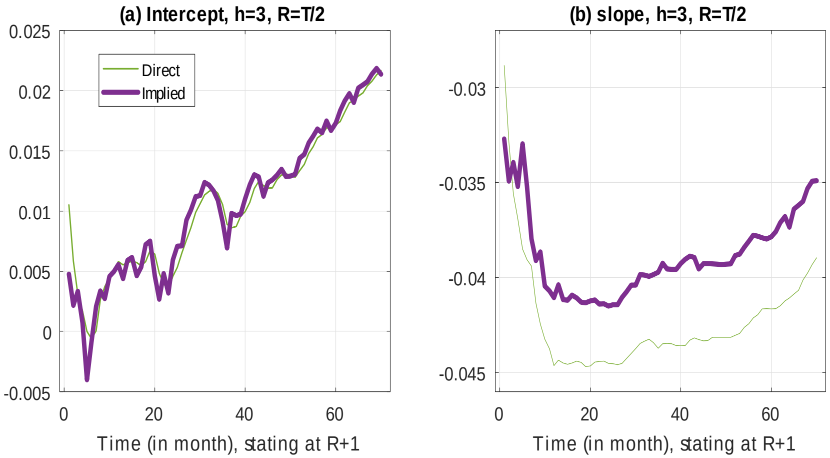


$$\left(R^2_{OOS,Direct}, R^2_{OOS,Implied}\right) = (-3.62\%, 11.5\%)$$

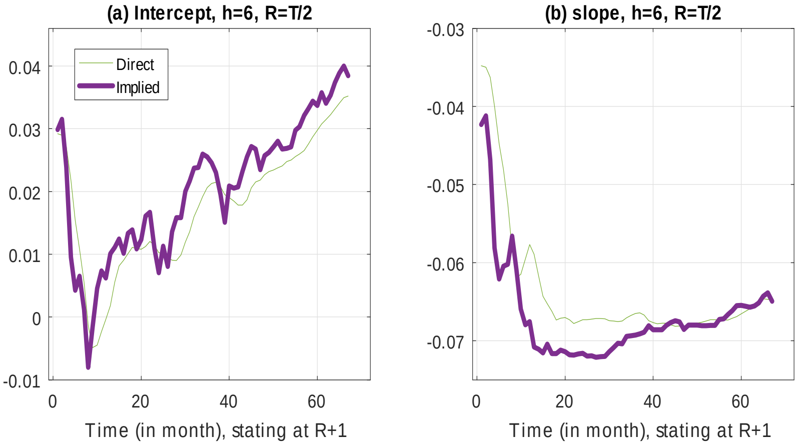

R=P=T/2
MS (Direct)
MS (Implied)
OOS R 2
0.35
0.3
0.25
0.2
0.15
0.1
0.05
0
-0.05
0
5
10
15
20
25
Horizon


R=T/4
MS (Direct)
MS (Implied)
0.25
0.2
0.15
0.1
0.05
0
OOS R 2
0
2
4
6
8
10
12
Horizon

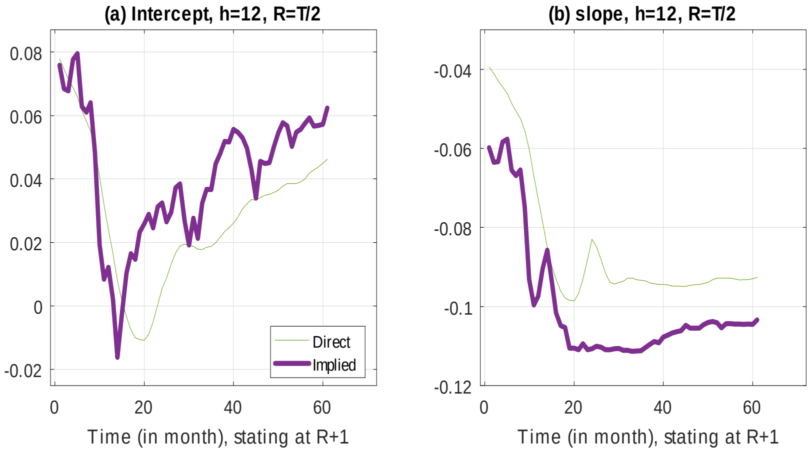
- We have seen that for this example, the forecast based on the implied estimator does improve on that based on the direct estimator, over almost all reasonable horizons.
- In the next few graphs, we will plot the point estimates of direct and implied estimator of intercept $\mu(h)$ and slope $\theta(h)$.
- They are nontrivially different.



(a) Intercept, h=3, R=T/2
Direct
Implied
0.025
0.02
0.015
0.01
0.005
0
-0.005
0
20
40
60
Time (in month), stating at R+1
(b) slope, h=3, R=T/2
-0.03
-0.035
-0.04
-0.045
0
20
40
60
Time (in month), stating at R+1



(a) Intercept, h=6, R=T/2
Direct
Implied
0.04
0.03
0.02
0.01
0
-0.01
0
20
40
60
Time (in month), stating at R+1
(b) slope, h=6, R=T/2
-0.03
-0.04
-0.05
-0.06
-0.07
0
20
40
60
Time (in month), stating at R+1


(a) Intercept, h=12, R=T/2
0.08
0.06
0.04
0.02
0
-0.02
0
20
40
60
Time (in month), stating at R+1
Direct
Implied
(b) slope, h=12, R=T/2
-0.04
-0.06
-0.08
-0.1
-0.12
0
20
40
60
Time (in month), stating at R+1

Multiple-horizon forecast vs. multiple-step forecast

- *Multiple-step* forecast: the target is y_{t+j} (for $j \geq 1$)
- *Multiple-horizon* forecast: the target is $\sum_{j=1}^{h} y_{t+j}$ (our interest here)
- Macroeconomists are often interested in forecasting y_{t+j}
- There are two methods for multiple-step forecast: *direct* and *iterated*.
- These two methods correspond to the direct method and implied method in multiple-horizon forecast we have already discussed.

- Suppose the horizon $h = 2$. We are forecasting $y_{t+1} + y_{t+2}$.
- The method now is based on *summing multiple-step forecasts*: we forecast y_{t+1} and y_{t+2} separately, and add two forecasts together to form the forecast for $y_{t+1} + y_{t+2}$.
- First consider the direct method:

$$\widehat{y}_{t+1} = \widehat{\mu}_y + \widehat{\beta} x_t, \tag{5}$$

$$\widehat{y}_{t+2} = \widehat{\mu}_y(2) + \widehat{\beta}(2) x_t, \tag{6}$$

  where both regressions are simple OLS.
- Then the forecast of $y_{t+1} + y_{t+2}$ is $\widehat{y}_{t+1} + \widehat{y}_{t+2}$.
- This method is identical to the forecast we introduced earlier, based on a direct horizon-two predictive regression.

- Now we discuss the iterated method.
- To forecast y_{t+1}, we use the same forecast as in (5).
- To forecast y_{t+2}, we use a different forecast from (6):

$$\widehat{y}_{t+2}^{IM} = [\widehat{\mu}_y + \widehat{\beta}(1 - \widehat{\rho})\widehat{Ex}] + \widehat{\beta}\widehat{\rho}x_t. \tag{7}$$

- To obtain (7), the iterated method needs a model for x_t. We use AR(1):

$$(x_t - Ex) = \rho(x_{t-1} - Ex) + v_t. \tag{8}$$

- Then

$$\begin{aligned} \widehat{y}_{t+2}^{IM} &\overset{\text{Step 1}}{=} \widehat{\mu}_y + \widehat{\beta}\widehat{x}_{t+1} \\ &\overset{\text{Step 2}}{=} \widehat{\mu}_y + \widehat{\beta}[\widehat{Ex} + \widehat{\rho}(x_t - \widehat{Ex})] \\ &= [\widehat{\mu}_y + \widehat{\beta}(1-\widehat{\rho})\widehat{Ex}] + \widehat{\beta}\widehat{\rho}x_t, \end{aligned}$$

where $\widehat{x}_{t+1} = \widehat{Ex} + \widehat{\rho}(x_t - \widehat{Ex})$. Thus (7) holds.

- Combining (5) and (7):

$$\widehat{y}_{t+1}^{IM} + \widehat{y}_{t+2}^{IM} = \underbrace{(2\widehat{\mu}_y + \widehat{\beta}(1-\widehat{\rho})\widehat{Ex})}_{=\widehat{\mu}^{IM}(2)} + \underbrace{\widehat{\beta}(1+\widehat{\rho})}_{=\widehat{\theta}^{IM}(2)}\widehat{x}_t. \tag{9}$$

- This is exactly the forecast based on our implied estimator.

- Some applied researchers like to use the following AR(1) model (instead of (8)):

$$x_t = \mu_x + \rho x_{t-1} + v_t. \tag{10}$$

- It is important to point out that it does not matter to use the AR(1) model in (8) or (10), if the interest is only in the slope $\theta(h)$ (like in the context of testing predictability).
- However, if the interest is in both the intercept and the slope (like in the context of forecast), forecast based on (8) or (10) can be very different. (The one based on (8) is often better in finite-samples.)

- Under the model (10), the iterated forecast is:

$$
\begin{aligned}
\widehat{y}_{t+2}^{IM} &\overset{\text{Step 1}}{=} \widehat{\mu}_y + \widehat{\beta}\widehat{x}_{t+1} \\
&\overset{\text{Step 2}}{=} \widehat{\mu}_y + \widehat{\beta}(\widehat{\mu}_x + \widehat{\rho}x_t) \\
&= (\widehat{\mu}_y + \widehat{\beta}\widehat{\mu}_x) + \widehat{\beta}\widehat{\rho}x_t, \qquad (11)
\end{aligned}
$$

where $\widehat{x}_{t+1} = \widehat{\mu}_x + \widehat{\rho}x_t$.

- Combining (5) and (11):

$$
\widehat{y}_{t+1}^{IM} + \widehat{y}_{t+2}^{IM} = \underbrace{(2\widehat{\mu}_y + \widehat{\beta}\widehat{\mu}_x)}_{=\widehat{\mu}^{IM}(2)} + \underbrace{\widehat{\beta}(1+\widehat{\rho})}_{=\widehat{\theta}^{IM}(2)}x_t. \qquad (12)
$$

- Again, this is exactly the forecast based on our implied estimator (if Ex is estimated under the model (10)).

- The intercept in (12) is consistent with (3), which is written as (when specifying $h = 2$):

$$
\begin{aligned}
\widehat{\mu}^{IM}(h) &= h\widehat{\mu}_y + \left(h - \textstyle\sum_{j=0}^{h-1} \widehat{\rho}^j\right) \widehat{\beta}\widehat{Ex} \\
&= 2\widehat{\mu}_y + (2 - (1 + \widehat{\rho}))\, \widehat{\beta}\frac{\widehat{\mu}_x}{1 - \widehat{\rho}} \\
&= 2\widehat{\mu}_y + \widehat{\beta}\widehat{\mu}_x.
\end{aligned}
$$

- Note that using $\widehat{Ex} = \frac{\widehat{\mu}_x}{1-\widehat{\rho}}$ is important to reconcile the implied forecast and the iterated forecast under the model (10).

Implied forecast based on which AR(1) model works better, (8) or (10)?