

Example 14 — Predicted values

[Description](#) [Remarks and examples](#) [Also see](#)

Description

We demonstrate the use of `predict`. See [\[SEM\] Intro 7](#) and [\[SEM\] predict after sem](#).

This example picks up where the first part of [\[SEM\] Example 1](#) left off:

```
. use https://www.stata-press.com/data/r17/sem_1fmm
. sem (x1 x2 x3 x4 <- X)
```

Remarks and examples

[stata.com](#)

`predict` can create new variables containing predicted values of 1) observed endogenous variables, 2) latent variables, whether endogenous or exogenous, and 3) latent endogenous variables. In the case of latent variables, item 2 corresponds to the factor score and item 3 is the linear prediction.

Below we demonstrate 1 and 2:

```
. predict x1hat x2hat, xb(x1 x2)
. predict Xhat, latent(X)
```

You specify options on `predict` to specify what you want predicted and how. Because of the differing options, the two commands could not have been combined into one command.

Our dataset now contains three new variables. Below we compare the three variables with the original `x1` and `x2` by using first `summarize` and then `correlate`:

```
. summarize x1 x1hat x2 x2hat Xhat
```

Variable	Obs	Mean	Std. dev.	Min	Max
x1	500	99.518	14.35402	60	137
x1hat	500	99.518	9.363112	71.45533	126.7325
x2	500	99.954	14.1939	52	140
x2hat	500	99.954	9.674426	70.95827	128.0733
Xhat	500	1.03e-08	9.363112	-28.06267	27.21449

Notes:

1. Means of `x1hat` and `x1` are identical; means of `x2hat` and `x2` are identical.
2. The standard deviation of `x1hat` is less than that of `x1`; the standard deviation of `x2hat` is less than that of `x2`. Some of the variation in `x1` and `x2` is not explained by the model.
3. Standard deviations of `x1hat` and `Xhat` are equal. This is because in

$$x_1 = b_0 + b_1 X + e_1$$

coefficient b_1 was constrained to be equal to 1 because of the anchoring normalization constraint; see [Identification 2: Normalization constraints \(anchoring\)](#) in [\[SEM\] Intro 4](#).

The mean of \hat{X} in the model above is $1.03\text{e-}08$ rather than 0. Had we typed

```
. predict double Xhat, latent(X)
```

the mean would have been $-1.17\text{e-}16$.

```
. correlate x1 x1hat x2 x2hat Xhat  
(obs=500)
```

	x1	x1hat	x2	x2hat	Xhat
x1	1.0000				
x1hat	0.6705	1.0000			
x2	0.4537	0.7007	1.0000		
x2hat	0.6705	1.0000	0.7007	1.0000	
Xhat	0.6705	1.0000	0.7007	1.0000	1.0000

Notes:

1. Both $x1hat$ and $x2hat$ correlate 1 with $Xhat$. That is because both are linear functions of $Xhat$ alone.
2. That $x1hat$ and $x2hat$ correlate 1 is implied by item 1, directly above.
3. That $Xhat$, $x1hat$, and $x2hat$ all have the same correlation with $x1$ and with $x2$ is also implied by item 1, directly above.

Also see

[SEM] [Example 1](#) — Single-factor measurement model

[SEM] [Intro 7](#) — Postestimation tests and predictions

[SEM] [predict after sem](#) — Factor scores, linear predictions, etc.