BBVA Research

Evaluating the out-of-sample prediction performance of panel data models

12th Spanish STATA Conference, Madrid, Spain,
October 17, 2019

Alfonso Ugarte-Ruiz

Creating Opportunities

Contents

Motivation

- Evaluating the forecasting/prediction accuracy of a statistical model is becoming increasingly common and essential in a broad range of practical applications (e.g. macroeconomics variables forecasting for regulatory purposes, machine-learning and big-data techniques, etc.)

- However, the available applications that we are aware of, have concentrated on only one type of data structure per application/case, either time-series or unstructured/cross-section/pooled data.

- The evaluation of the prediction performance of a panel-data statistical model ideally should take into account the two dimensions inherent in a panel, the time-series dimension and the cross-section (individuals) dimension.

- To the best of our knowledge there is no automatic procedure in Stata to evaluate the out-of-sample performance of a model in a time-series dimension.

- Additionally, the available procedures that perform **cross-validation** exercises (e.g. *crossfold, cvauroc*) usually play with all the observations when separating the in- and out-of-samples, without taking into account if such observations could belong to different individuals or are subsequent observations from the same individual.

- The latter could be problematic if one wants to fit a dynamic or a Fixed-Effects model, or could simply make the results more difficult to analyze in a panel data framework.

- Moreover, it is usually convenient (and also common practice) to express the performance of a model in relative terms to another alternative estimation method.

- For instance, when evaluating the forecasting accuracy in a time-series framework, the RMSE of a model is usually compared to the RMSE of a "naïve" forecast in which the last observation of the in-sample period is used as a direct forecast for the out-of-sample observations.

- But, what would be the "naïve" forecast if you just randomly take out observations?

- We also think in the panel data case a more useful exercise would be one analogous to cross-validation, but using individuals instead of observations.

General features of the new procedures

- We have developed 4 new commands that allow evaluating the out-of-sample prediction performance of panel-data models in their time-series and cross-individual dimensions separately, and have also developed separate procedures for different types of dependent variables, either continuous or dichotomous variables (*xtoos_t, xtoos_i, xtoos_bin_t* and *xtoos_bin_i*).

- The time-series procedures (*xtoos_t, xtoos_bin_t*) exclude a number of time periods defined by the user from the estimation sample for each individual in the panel.

- Correspondingly, the cross-individual procedures (*xtoos_i, xtoos_bin_i*) exclude a group of individuals (e.g. countries) defined by the user from the estimation sample (including all their observations throughout time).

- Then for the remaining (in-sample) subsamples they fit the specified models and use the resulting parameters to forecast/predict the dependent variable (or the probability of a positive outcome) in the unused periods or individuals (out-of-sample).

- The unused time-periods or individuals sets are then recursively reduced by one period in every subsequent step in the time-series case, or in a random or ordered fashion in the cross-individuals one, and the estimation and forecasting evaluation repeated, until there are no more periods ahead or more individuals that could be left out and evaluated.

- In the continuous cases the model's forecasting performance is reported both in absolute terms (RMSE) and also relative to an alternative “naïve” prediction and the relative performance expressed by means of an U-Theil ratio.

- In the binary dependent variable case, the performance is evaluated based on the area under the receiver operator characteristic statistic (AUROC) evaluated in both the training sample and the out-of-sample.

- The procedures' options and characteristics are flexible enough to allow the following:

1. Choosing different estimation methods
2. Choosing between a naïve prediction or an AR1 model as the alternative/comparison model
3. Choosing the estimation method of the AR1 model
4. Using dynamic specifications (lags of the dependent variable). It automatically handles **dynamic forecasting**
5. Choosing dynamic methods (**xtabond/xtdpdsys**)
6. Could be used automatically in a dataset with only time-series observations
7. Using data with different time frequencies, i.e. annual, quarterly, monthly and undefined time-periods
8. Evaluating the model's performance of one particular individual or a defined group of individuals instead of the whole panel
9. Choosing between within (FE), random (RE) or dummy variables estimation
10. To include, or not, the estimated individual component (intercept) in the prediction

Continuous case, time-series dimension: *xtoos_t*

- *xtoos_t* reports the specified model's forecasting performance, both in absolute terms (RMSE) and also relative to an alternative model by means of an U-Theil ratio (ratio of corresponding RMSEs).

- The default estimation method is *xtreg*

- By default, the alternative method is a "naive" prediction in which the last observation of the in-sample period is used directly as a forecast without any change. The procedure also allows to use an AR1 model as the alternative model for the comparison.

- If the sample is unbalanced, it automatically discards those individuals with observations that start within the defined out-of-sample periods.

- Performance results are broken down and reported in two different ways:

  1) According to the last period included in the estimation sample.
  2) According to the length of the forecasting horizon.

Syntax

```
xtoos_t depvar [indepvars] [if],
        [indate(string)] [cdate(string)] [met(string)] [mcomp(string)] [evalopt(varname)]
        [fe] [xbu] [dum] [opar] [lags(numlist)] [hgraph(numlist)]
        [model_options]
```

- Use of *xtoos_t* to evaluate the prediction perfomance **between periods 15 and 20** (out of 20 total periods in the sample, T=20, N=5)

```
. webuse invest2, clear
. xtset company time
. xtoos_t invest market stock, indate(15) cdate(20)
```

```
Out of sample evaluation according to last in-sample date
```

	RMSE_ous	RMSE_Al~s	UTheil	N
15	197.5926	268.0517	.7371436	25
16	213.4479	264.6098	.8066516	20
17	218.0811	242.715	.898507	15
18	215.6059	237.3739	.9082966	10
19	189.9464	115.6006	1.643126	5
Summary	208.0358	250.6152	.8301006	75

```
Out of sample evaluation according to forecasting horizon
```

	RMSE_ous	RMSE_Al~s	UTheil	N
1	178.7174	112.7054	1.585704	25
2	202.2296	212.2442	.9528156	20
3	222.5415	287.1967	.774875	15
4	241.209	369.5434	.6527217	10
5	246.8993	421.7185	.5854599	5
Summary	208.0358	250.6152	.8301006	75

- Use of *xtoos_t* to evaluate the prediction perfomance between periods 15 and 20, but restricting the evaluation only to company # 1

```
. gen company1=company==1
. xtoos_t invest market stock, indate(15) cdate(20) evalopt(company1)

Out of sample evaluation according to last in-sample date
```

	RMSE_ous	RMSE_Al~s	UTheil	N
15	265.9022	563.972	.4714813	5
16	304.6125	553.1732	.5506639	4
17	319.0211	533.3001	.5982018	3
18	326.3651	512.5205	.6367845	2
19	243.6763	182.2999	1.336678	1
Summary	294.6642	530.7943	.5551382	15

```
Out of sample evaluation according to forecasting horizon
```

	RMSE_ous	RMSE_Al~s	UTheil	N
1	170.677	220.3413	.7746027	5
2	257.4415	435.1569	.5916061	4
3	333.4687	601.2803	.5545977	3
4	414.2776	797.9501	.5191773	2
5	463.4559	931.6	.4974838	1
Summary	294.6642	530.7943	.5551382	15

- Use of *xtoos_t* using as **estimation method** the command **xtregar,** and using **xtabond** to estimate an AR1 model as the **comparison model**

```
. xtoos_t invest market stock, indate(15) cdate(20) met(xtregar) mcomp(xtabond)
```

Out of sample evaluation according to last in-sample date

	RMSE_ous	RMSE_Al~s	UTheil	N
15	225.784	397.9647	.5673469	25
16	231.1073	409.8826	.5638378	20
17	234.0517	378.393	.6185415	15
18	230.5379	305.2529	.7552358	10
19	192.5137	111.7499	1.722719	5
Summary	227.4836	373.4778	.6090955	75

Out of sample evaluation according to forecasting horizon

	RMSE_ous	RMSE_Al~s	UTheil	N
1	184.2658	170.6574	1.079741	25
2	214.7247	330.6099	.6494805	20
3	243.642	447.3137	.5446781	15
4	276.5614	537.3367	.5146893	10
5	301.6278	575.9547	.5237005	5
Summary	227.4836	373.4778	.6090955	75

- Use of *xtoos_t* using **Fixed-Effects (within)** estimator, and **including** the estimated individual components in the prediction

```
. xtoos_t invest market stock, indate(15) cdate(20) fe xbu
```

Out of sample evaluation according to last in-sample date

	RMSE_ous	RMSE_Al~s	UTheil	N
15	131.2454	268.0517	.4896273	25
16	146.6338	264.6098	.554151	20
17	153.0876	242.715	.6307299	15
18	156.1486	237.3739	.6578172	10
19	145.2602	115.6006	1.25657	5
Summary	144.2926	250.6152	.5757536	75

Out of sample evaluation according to forecasting horizon

	RMSE_ous	RMSE_Al~s	UTheil	N
1	107.0292	112.7054	.949637	25
2	134.963	212.2442	.6358857	20
3	158.8733	287.1967	.5531866	15
4	182.9497	369.5434	.4950697	10
5	198.7604	421.7185	.4713107	5
Summary	144.2926	250.6152	.5757536	75

- Which is equivalent to the use of *xtoos_t* using **dummy variables** per individual **and including their estimated values** in the prediction

```
. xtoos_t invest market stock, indate(15) cdate(20) dum
```

- Use of *xtoos_t* using **Fixed-Effects (within)** estimator, **without including** the estimated individual components in the prediction

```
. xtoos_t invest market stock, indate(15) cdate(20) fe

 Out of sample evaluation according to last in-sample date

             |  RMSE_ous  RMSE_Al~s     UTheil          N
-------------+----------------------------------------------
          15 |  200.6561   268.0517   .7485726         25
          16 |  217.1251   264.6098    .820548         20
          17 |  220.5621    242.715   .9087287         15
          18 |   216.974   237.3739     .91406         10
          19 |  189.1054   115.6006   1.635851          5
-------------+----------------------------------------------
     Summary |   210.673   250.6152   .8406234         75

 Out of sample evaluation according to forecasting horizon

             |  RMSE_ous  RMSE_Al~s     UTheil          N
-------------+----------------------------------------------
           1 |  179.4424   112.7054   1.592138         25
           2 |  204.1317   212.2442   .9617777         20
           3 |  225.9538   287.1967   .7867564         15
           4 |  246.3197   369.5434   .6665514         10
           5 |  252.1064   421.7185   .5978073          5
-------------+----------------------------------------------
     Summary |   210.673   250.6152   .8406234         75
```

- Which is equivalent to the use of *xtoos_t* using **dummy variables** per individual without including their estimated values in the prediction

```
. xtoos_t invest market stock, indate(15) cdate(20) dum opar
```

- Use of *xtoos_t* including lags of the dependent variable in the specification

```
. xtoos_t invest market stock, indate(15) cdate(20) lags(1/2)
```

Out of sample evaluation according to last in-sample date

	RMSE_ous	RMSE_Al~s	UTheil	N
15	276.2091	268.0517	1.030432	25
16	239.8666	264.6098	.9064915	20
17	196.2107	242.715	.8083996	15
18	176.9739	237.3739	.7455494	10
19	85.2597	115.6006	.7375368	5
Summary	230.5075	250.6152	.9197667	75

Out of sample evaluation according to forecasting horizon

	RMSE_ous	RMSE_Al~s	UTheil	N
1	96.5169	112.7054	.8563648	25
2	182.2716	212.2442	.8587828	20
3	257.5752	287.1967	.89686	15
4	348.5573	369.5434	.9432106	10
5	418.9478	421.7185	.9934299	5
Summary	230.5075	250.6152	.9197667	75

- Use of **xtoos_t** using a **dynamic model method**, either xtabond or xtdpdsys. In this case, the default specification includes one lag of the dependent variable

```
. xtoos_t invest market stock, indate(15) cdate(20) lags(2) met(xtabond)
```

Out of sample evaluation according to last in-sample date

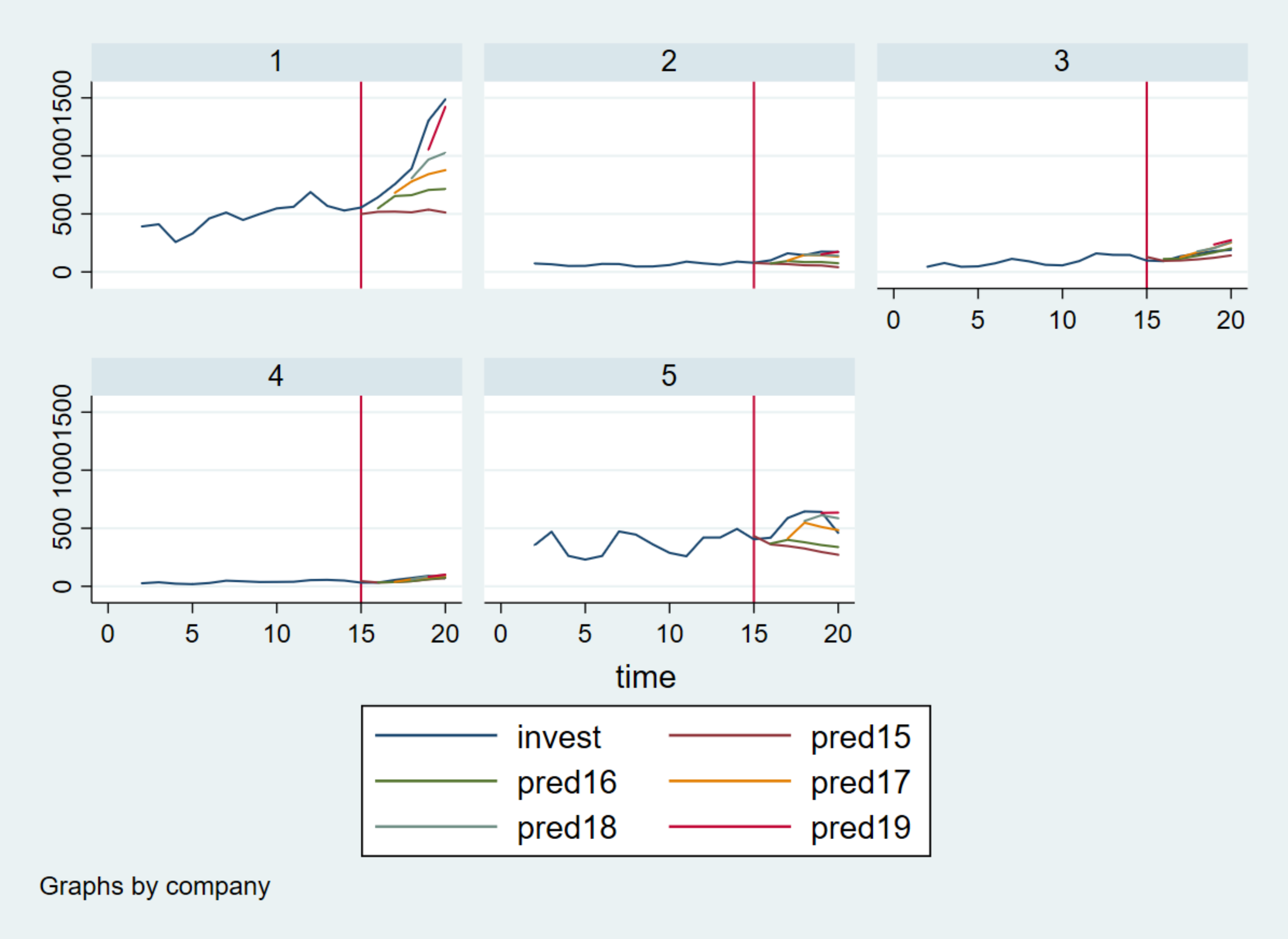
	RMSE_ous	RMSE_Al~s	UTheil	N
15	162.2756	268.0517	.6053894	25
16	158.3917	264.6098	.5985858	20
17	133.3588	242.715	.5494461	15
18	111.5392	237.3739	.4698881	10
19	139.9054	115.6006	1.210248	5
Summary	148.2853	250.6152	.591685	75

Out of sample evaluation according to forecasting horizon

	RMSE_ous	RMSE_Al~s	UTheil	N
1	110.7552	112.7054	.9826967	25
2	150.9671	212.2442	.7112896	20
3	171.9631	287.1967	.5987642	15
4	174.3669	369.5434	.4718441	10
5	166.7588	421.7185	.3954268	5
Summary	148.2853	250.6152	.591685	75

- Use of *xtoos_t* to draw a "hair" graph with all the model forecasts at each forecasting horizons for individuals 1 to 5

```
. xtoos_t invest market stock, indate(15) cdate(20) lags(1) hgraph(1/5)
```



Continuous case, cross-individuals dimension: *xtoos_i*

- *xtoos_i* reports the specified model's forecasting performance, both in absolute terms (RMSE) and also relative to an alternative model by means of an U-Theil ratio.

- The default estimation method is *xtreg*

- By default, the alternative model is a "naive" prediction in which **the mean of all in-sample individuals at every time-period is used as a prediction** for the excluded ones. The procedure also allows to use an AR1 model as the alternative model for the comparison.

- It also reports several in-sample and out-of-sample statistics of both the specified and the comparison models.

- The individuals excluded (out-of-sample) could be:

    1. random subsamples of size n; if the whole sample contains N individuals, then N/n subsamples without repeated individuals are extracted and evaluated. Moreover, the sampling process could be repeated r times, similar to “bootstrapping”

    2. an ordered partition of the sample in subsamples of size k; if the whole sample contains N individuals, then N/k ordered subsamples are formed and evaluated, similar to K-fold cross-validation, but using individuals instead of observations.

    3. a particular individual or a particular group (e.g. country or a region).

- If in option 1, n=1, or in option 2, k=1, both would be equivalent to “Leave-one-out cross-validation (LOOCV)”

Syntax

```
xtoos_i depvar [indepvars] [if],
        [ous(integer)] [rsmpl(integer)] [ksmpl(integer)]
        [evalopt(varname)] [met] [mcomp(string)] [fe] [dum] [lags(numlist)]
        [hgraph] [model_options]
```

- Use of *xtoos_i* to evaluate the prediction performance for 20 random subsamples of 40 individuals (rsmpl() and ous()) and ordered subsamples of also 40 individuals (ksmpl())

```
. webuse abdata, clear

. xtoos_i n w l.w k l.k ys l.ys, ous(40) rsmpl(20) ksmpl(40)
```

Out of sample evaluation: Random sampling

	RMSE_in	RMSE_ous	R2_in	R2_ous	RMSE_Al~n	RMSE_Al~s	R2_Alt_in	R2_Alt_~s	Utheil_~t	N
Summary	.5687863	.5790578	.8189515	.8094784	1.319233	1.337636	.0260438	-.0166604	.4328965	236.2735

Out of sample evaluation: Ordered partition

	RMSE_in	RMSE_ous	R2_in	R2_ous	RMSE_Al~n	RMSE_Al~s	R2_Alt_in	R2_Alt_~s	Utheil_~t	N
Summary	.5679359	.5890679	.8193944	.791674	1.318939	1.347322	.0259486	-.0898223	.4372138	230.0393

- Use of *xtoos_i* to evaluate the prediction performance restricting the **evaluation only to first 6 individuals**, and **no random sampling**

```
. gen id1to6=id<=6

. xtoos_i n w l.w k l.k ys l.ys, ous(40) rsmpl(0) ksmpl(40) evalopt(id1to6)
```

Out of sample evaluation: Ordered partition

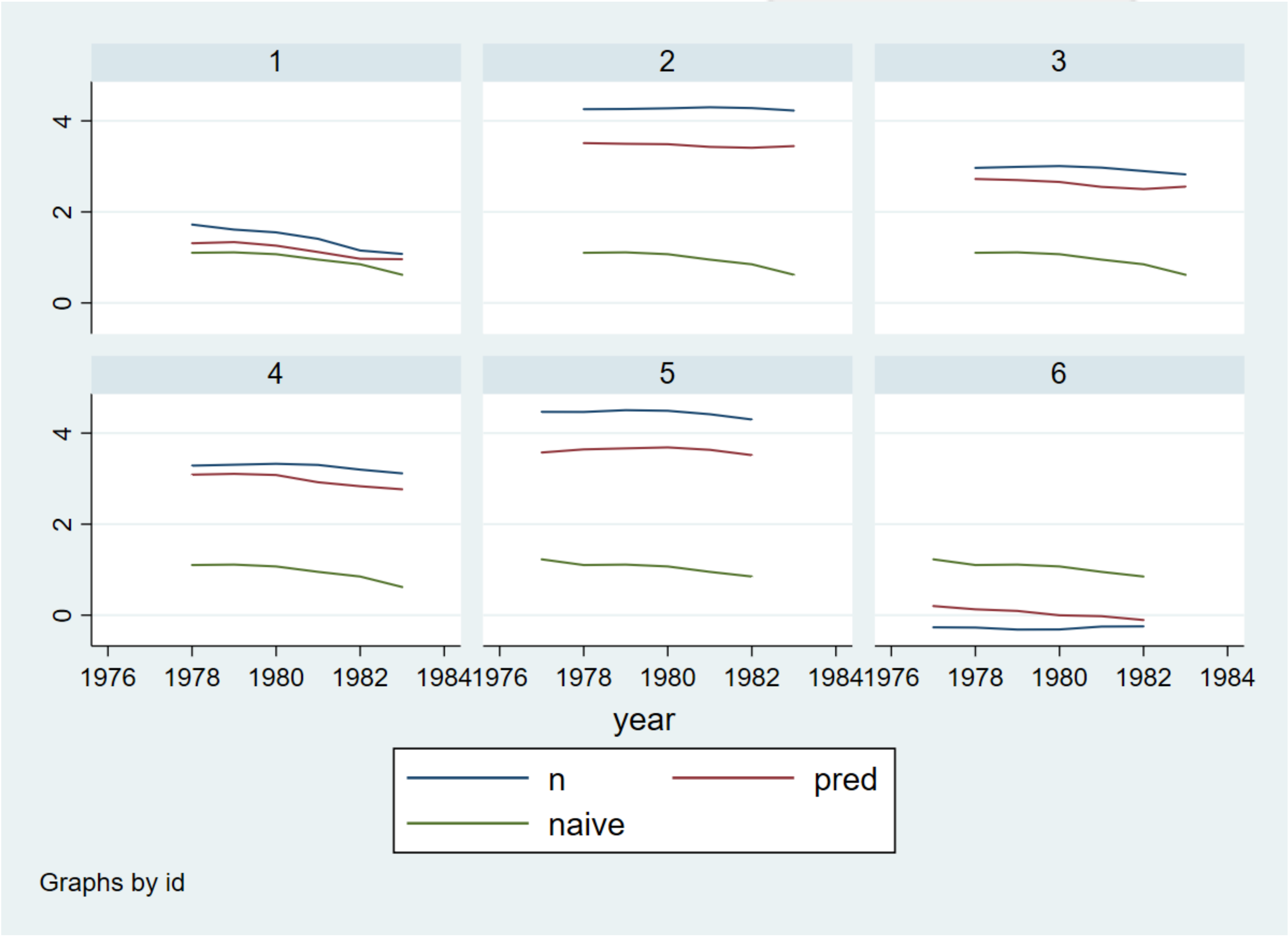
	RMSE_in	RMSE_ous	R2_in	R2_ous	RMSE_Al~n	RMSE_Al~s	R2_Alt_in	R2_Alt_~s	Utheil_~t	N
Summary	.5679359	.5890679	.8193944	.791674	1.318939	1.347322	.0259486	-.0898223	.4372138	230.0393

Out of sample evaluation: Specific individuals: id1to6

	RMSE_in	RMSE_ous	R2_in	R2_ous	RMSE_Al~n	RMSE_Al~s	R2_Alt_in	R2_Alt_~s	Utheil_~t	N
Summary	.5769317	.5356993	.7970362	.8951593	1.262116	2.374768	.0286659	-1.060298	.2255797	36

- Use of ***xtoos_i*** to evaluate the prediction performance restricting the **evaluation only to first 6 individuals**, while drawing a **graph with the prediction for each one of those 6 individuals**

```
. xtoos_i n w l.w k l.k ys l.ys, ous(40) rsmpl(0) ksmpl(40) evalopt(id1to6) hgraph
```



Binary dependent variable case, Time-series dimension: *xtoos_bin_t*

- *xtoos_bin_t* evaluates the prediction performance based on the area under the receiver operator characteristic (ROC) statistic evaluated in both the in-sample and the out-of-sample.

- The default estimation method is *xtlogit,* but it allows to choose different estimation methods (e.g.logit, probit, xtprobit)

- *xtoos_bin_t* allows to choose different estimation methods different estimation methods (e.g. logit, probit, xtprobit) and could also be used in a time-series dataset only.

- It allows to choose the method of estimating the probability of a positive outcome that depends on the estimation method used (e.g. prob, pu0, pc1)

Syntax

```
xtoos_bin_t depvar [indepvars] [if],
        [indate(string)] [cdate(string)] [mprob(string)] [evalopt(varname)]
        [met(string)] [fe] [dum]
        [model_options]
```

- Use of *xtoos_bin_t* to evaluate the prediction performance of a FX crisis variable, **between 2015Q4 and 2018Q4** (out of a sample between 1980Q1 and 2018Q4 and 83 countries)

```
. xtoos_bin_t crisis_aq l4.inflation_bdevt l4.inflation_bmt l4.inv_gdpmt l4.inv_gdpdevt l4.gdp_gro_cycdevt l4.gdp_gro_cycmt l4.gdp_
  gro_trdevt l4.gdp_gro_trmt ///
  l4.credit_gdp_gap l4.reerdevmt l4.reerdevdevt l.libor_b l2.libor_b l.baa_spread l2.baa_spread l2.lnstintrate_devus_b l4.lnstintr
  ate_devus_b ///
  l4.ca_gdp_cycle l4.ca_gdp_tr l4.trade_gdp, indate(2015q4) cdate(2018q4) mprob(pr)

    Out of sample evaluation according to last last in-sample date
```

	ROC_in	ROC_SE_in	ROC_os	ROC_SE_os	N_in	N_os
2015q4	.7618056	.0085364	.8126019	.0364684	9240	996
2016q1	.7629634	.0084863	.7985668	.0405821	9323	913
2016q2	.7642264	.0084432	.7971519	.0427511	9406	830
2016q3	.7638412	.0084177	.8031515	.0446161	9489	747
2016q4	.7620929	.0084327	.83518	.0481357	9572	664
2017q1	.7623594	.008412	.8505102	.0499933	9655	581
2017q2	.7626354	.0084078	.9108889	.0332676	9738	498
2017q3	.764332	.0083619	.9189756	.0333101	9821	415
2017q4	.7661259	.0083113	.9223156	.0339919	9904	332
2018q1	.7664681	.0082861	.9565349	.0249399	9987	249
2018q2	.7677665	.0082283	.9489736	.0314079	10070	166
2018q3	.7683953	.0081724	.9458874	.0477669	10153	83
Summary	.7644711	.0083717	.8440162	.0402006	116358	6474

Out of sample evaluation according to forecasting horizon

	ROC_in	ROC_SE_in	ROC_os	ROC_SE_os	N_in	N_os
1	.7618056	.0085364	.8406007	.	9240	996
2	.7629634	.0084863	.8313807	.	9323	913
3	.7642264	.0084432	.8196664	.	9406	830
4	.7638412	.0084177	.8248871	.	9489	747
5	.7620929	.0084327	.8399176	.	9572	664
6	.7623594	.008412	.8530721	.	9655	581
7	.7626354	.0084078	.9420602	.	9738	498
8	.764332	.0083619	.9328313	.	9821	415
9	.7661259	.0083113	.9153981	.0629233	9904	332
10	.7664681	.0082861	.9670857	.0280827	9987	249
11	.7677665	.0082283	.953713	.0374924	10070	166
12	.7683953	.0081724	.9523809	.0373951	10153	83
Summary	.7644711	.0083717	.8626034	.	116358	6474

Binary dependent variable case, cross-individuals dimension: *xtoos_bin_i*

- *xtoos_bin_i* evaluates the prediction performance based on the area under the receiver operator characteristic (AUROC) statistic evaluated in both the training sample and the out-of-sample.

- The default estimation method is *xtlogit,* but it allows to choose different estimation methods (e.g.logit, probit, xtprobit)

- It has the same options of choosing the sample to be excluded (out-of-sample) as in the continuous case (*xtoos_i*)

- It allows to choose the method for estimating the probability of a positive outcome, which depends on the estimation method used (e.g. prob, pu0, pc1)

- It also reports the AUROC for the in-sample individuals and also estimates AUROC's standard error

Syntax

```
xtoos_bin_i depvar [indepvars] [if],
        [ous(integer)] [rsmpl(integer)] [ksmpl(integer)] [mprob(string)]
        [evalopt(varname)] [met] [fe]
        [model_options]
```

- Use of *xtoos_bin_i* to evaluate the prediction based on AUROC for 1 random subsample of 20 countries (*rsmpl()* and *ous()*) and ordered subsamples of also 20 individuals (*ksmpl()*)

```
. xtoos_bin_i crisis_aq l4.inflation_bdevt l4.inflation_bmt l4.inv_gdpmt l4.inv_gdpdevt l4.gdp_gro_cycdevt l4.gdp_gro_cycmt l4.gdp_
  gro_trdevt l4.gdp_gro_trmt ///
  l4.credit_gdp_gap l4.reerdevmt l4.reerdevdevt l.libor_b l2.libor_b l.baa_spread l2.baa_spread l2.lnstintrate_devus_b l4.lnstintr
  ate_devus_b ///
  l4.ca_gdp_cycle l4.ca_gdp_tr l4.trade_gdp, k(20) o(20) mprob(pr)
```

Out of sample evaluation: Random sampling

	ROC_in	ROC_SE_os	ROC_os	ROC_SE_os	N_in	N_os
Summary	.7762528	.0092581	.742395	.0172767	7812.445	2657.555

Out of sample evaluation: Ordered partition

	ROC_in	ROC_SE_os	ROC_os	ROC_SE_os	N_in	N_os
Summary	.7752332	.0092699	.7263918	.0176432	31410	10470

- Use of *xtoos_bin_i* to evaluate the prediction performance based on AUROC for ordered subsamples of 20 individuals, and evaluating only the performance **for Indonesia**, and **no random sampling**

```
. gen idn=country=="Indonesia"
l4.ca_gdp_cycle l4.ca_gdp_tr l4.trade_gdp, k(20) o(20) r(0) evalopt(idn)
```

Out of sample evaluation: Ordered partition

	ROC_in	ROC_SE_os	ROC_os	ROC_SE_os	N_in	N_os
Summary	.7752332	.0092699	.7263918	.0176432	31410	10470

Out of sample evaluation: Specific individuals: idn

	ROC_in	ROC_SE_os	ROC_ous	ROC_SE_os	N_in	N_os
Summary	.7712875	.	.7795233	.	10315	155

Conclusions

Conclusions

- We have developed several new commands that allow evaluating the out-of-sample prediction performance of panel-data models in their time-series and cross-individual dimensions separately, with separate procedures for different types of dependent variables, either continuous or dichotomous variables (*xtoos_t, xtoos_i, xtoos_bin_t* and *xtoos_bin_i*).

- The new commands are flexible enough to allow a large number of methodological options.

- These procedures could help us in several different goals:

  i. We can asses the prediction accuracy of existing models
  ii. They should help us uncover previously ignored differences in the prediction ability of panel data models between their two inherent dimensions
  iii. Allowing us to use the out-of-sample prediction performance as a selection criteria between different models in a straightforward manner.
  iv. They can be easily incorporated into new algorithms to select among a large number of models (in fact, we have already developed various new commands in this fashion).

Appendix

Appendix

- We have also developed new commands that are analogous to the ones described here, that could help us to select among a large number or alternative models (specifications) or a large number of different explanatory variables, using the two dimensions of the prediction performance as new selection criteria.

- The command *selectmod* estimates all possible combinations (specifications) of the list of explanatory variables provided. It estimates five statistical criteria per specification (Adj R2, AIC, BIC, U-Theil in time-series, U-Theil in cross-individual), ranks each specification according to each criteria and computes a composite ranking of all five criteria. It finally sorts all possible specifications according to the selected ranking.

```
. selectmod lnhpr bcredit_gdp urban_pop_gr urban_pop_grun strintrate_b unempl vix if year>=1990 & year<=2018, ///
> ind(2010) cd(2018) met(xtregar) xbu fix(lngdprind) o(15) k(15) r(0) qui exc(example) sheet(example1) cond(!(2&3))
```

1	Model	R2_ad	AIC	BIC	Uth_TS	Uth_CS	R2_ad_r	AIC_r	BIC_r	Uth_TS_r	Uth_CS_r	Total
2	bcredit_gdp unempl	0.6618	689	710	1.7587	1.0182	9	7	9	3	37	65
3	bcredit_gdp urban_pop_gr	0.6473	686	707	1.8647	0.9909	17	1	6	14	27	65
4	bcredit_gdp urban_pop_gr unempl	0.6657	687	713	1.7482	1.0217	6	2	18	1	39	66
5	bcredit_gdp	0.6423	688	704	1.8833	0.9906	21	4	1	16	26	68
6	urban_pop_gr	0.6264	689	704	1.9282	0.9451	41	8	3	25	1	78
7	urban_pop_gr unempl	0.6401	690	711	1.8546	0.9572	28	20	16	11	10	85
8	bcredit_gdp urban_pop_grun	0.6493	690	711	1.8634	1.0023	15	14	12	12	32	85
9	bcredit_gdp strintrate_b unempl	0.6612	689	715	1.8172	1.0130	12	10	23	6	36	87
10	unempl	0.6371	691	707	1.8643	0.9574	33	29	4	13	11	90
...												
43	urban_pop_gr strintrate_b vix	0.6246	692	718	2.0618	0.9529	43	32	31	46	7	159
44	urban_pop_grun unempl vix	0.6404	695	721	1.9410	0.9695	26	47	37	28	22	160
45	bcredit_gdp urban_pop_grun strintrate_b vix	0.6502	692	723	1.9617	1.0086	14	38	42	33	35	162
46	urban_pop_grun strintrate_b vix	0.6291	692	718	2.0367	0.9615	38	37	35	45	14	169
47	urban_pop_grun strintrate_b unempl vix	0.6404	693	725	2.0003	0.9753	25	45	45	39	23	177
48	urban_pop_gr strintrate_b unempl vix	0.6388	693	725	2.0133	0.9691	32	43	44	41	20	180

Appendix

- The command ***selectvar*** estimates the same specification but changing only one variable per estimation, i.e. each variable provided in the syntax.
- It estimates seven statistics per variable (Coefficient, t-statistic, Adj. R2, AIC, BIC, U-Theil in time-series, U-Theil in cross-individual). It ranks each specification according to the last five statistical criteria and computes a composite ranking of all five criteria. It finally sorts variables according to the selected ranking.

```
. selectvar lnhpr lngdprind lngdppc lnincomepc if year>=1990 & year<=2018, ///
> ind(2010) cd(2018) met(xtregar) xbu fix(bcredit_gdp urban_pop_gr strintrate_b unempl vix) ///
> o(15) k(15) r(0) qui exc(examplevars) sheet(example1)
```

Model	coef	tstat	R2	AIC	BIC	Uth_TS	Uth_CS	R2_r	AIC_r	BIC_r	Uth_TS_r	Uth_CS_r	Total
lngdprind	0.0395	15.4157	0.6644	692	729	1.8890	1.0302	1	1	1	1	1	5
lnincomepc	0.5893	12.2113	0.6295	870	907	1.9779	1.1258	3	2	2	3	2	12
lngdppc	0.3519	8.4223	0.6305	1053	1090	1.9408	1.4135	2	3	3	2	3	13