Title

> **xthtaylor** — Hausman–Taylor estimator for error-components models

Description

xthtaylor fits a random-effects model for panel data in which some of the covariates are correlated with the unobserved individual-level random effects. The command implements the Hausman–Taylor estimator by default, but the Amemiya–MaCurdy estimator is available for balanced panels.

Quick start

Hausman–Taylor model of y as a function of time-varying exogenous variable x1, time-invariant binary variable a, and time-varying endogenous variable x2 using xtset data

```
xthtaylor y x1 x2 a, endog(x2)
```

Same as above, and verify that a is the only time-invariant variable in the model

```
xthtaylor y x1 x2 a, endog(x2) constant(a)
```

Add time-invariant x3 as an endogenous covariate, but do not verify that a and x3 are the only time-invariant variables

```
xthtaylor y x1 x2 a x3, endog(x2 x3)
```

Same as above, but use Amemiya–MaCurdy estimator for balanced panels

```
xthtaylor y x1 x2 a x3, endog(x2 x3) am
```

Menu

Statistics > Longitudinal/panel data > Endogenous covariates > Hausman–Taylor regression (RE)

Syntax

> xthtaylor *depvar* *indepvars* [*if*] [*in*] [*weight*] , <u>e</u>ndog(*varlist*) [*options*]

options	Description
Model	
<u>no</u>constant	suppress constant term
* <u>e</u>ndog(*varlist*)	explanatory variables in *indepvars* to be treated as endogenous
<u>const</u>ant(*varlist*_{ti})	independent variables that are constant within panel
<u>vary</u>ing(*varlist*_{tv})	independent variables that are time varying within panel
<u>ama</u>curdy	fit model based on Amemiya and MaCurdy estimator
SE/Robust	
vce(*vcetype*)	*vcetype* may be conventional, <u>r</u>obust, <u>c</u>luster *clustvar*, <u>boot</u>strap, or <u>jack</u>knife
Reporting	
<u>l</u>evel(*#*)	set confidence level; default is level(95)
<u>sm</u>all	report small-sample statistics

* endog(*varlist*) is required.

A panel variable must be specified. For xthtaylor, amacurdy, a time variable must also be specified. Use xtset; see [XT] **xtset**.

depvar, *indepvars*, and all *varlist*s may contain time-series operators; see [U] **11.4.4 Time-series varlists**.

by, collect, statsby, and xi are allowed; see [U] **11.1.10 Prefix commands**.

iweights and fweights are allowed unless the amacurdy option is specified. Weights must be constant within panel; see [U] **11.1.6 weight**.

See [U] **20 Estimation and postestimation commands** for more capabilities of estimation commands.

Options

Model

noconstant; see [R] **Estimation options**.

endog(*varlist*) specifies that a subset of explanatory variables in *indepvars* be treated as endogenous variables, that is, the explanatory variables that are assumed to be correlated with the unobserved random effect. endog() is required.

constant(*varlist*_{ti}) specifies the subset of variables in *indepvars* that are time invariant, that is, constant within panel. By using this option, you assert not only that the variables specified in *varlist*_{ti} are time invariant but also that all other variables in *indepvars* are time varying. If this assertion is false, xthtaylor does not perform the estimation and will issue an error message. xthtaylor automatically detects which variables are time invariant and which are not. However, users may want to check their understanding of the data and specify which variables are time invariant and which are not.

varying(*varlist*_{tv}) specifies the subset of variables in *indepvars* that are time varying. By using this option, you assert not only that the variables specified in *varlist*_{tv} are time varying but also that all other variables in *indepvars* are time invariant. If this assertion is false, xthtaylor does not perform the estimation and will issue an error message. xthtaylor automatically detects which variables are time varying and which are not. However, users may want to check their understanding of the data and specify which variables are time varying and which are not.

amacurdy specifies that the Amemiya–MaCurdy estimator be used. This estimator uses extra instruments to gain efficiency at the cost of additional assumptions on the data-generating process. This option may be specified only for samples containing balanced panels, and weights may not be specified. The panels must also have a common initial period.

⌐ SE/Robust ¬

vce(*vcetype*) specifies the type of standard error reported, which includes types that are derived from asymptotic theory (conventional), that are robust to some kinds of misspecification (robust), that allow for intragroup correlation (cluster *clustvar*), and that use bootstrap or jackknife methods (bootstrap, jackknife); see [XT] *vce_options*.

vce(conventional), the default, uses the conventionally derived variance estimator for this Hausman–Taylor model.

Specifying vce(robust) is equivalent to specifying vce(cluster *panelvar*); see *xtpoisson, re and the robust VCE estimator* in *Methods and formulas* of [XT] **xtpoisson**.

⌐ Reporting ¬

level(*#*); see [R] **Estimation options**.

small specifies that the *p*-values from the Wald tests in the output and all subsequent Wald tests obtained via test use t and F distributions instead of the large-sample normal and χ^2 distributions. By default, the *p*-values are obtained using the normal and χ^2 distributions.

Remarks and examples

If you have not read [XT] **xt**, please do so.

Consider a random-effects model of the form

$$y_{it} = \mathbf{X}_{1it}\boldsymbol{\beta}_1 + \mathbf{X}_{2it}\boldsymbol{\beta}_2 + \mathbf{Z}_{1i}\boldsymbol{\delta}_1 + \mathbf{Z}_{2i}\boldsymbol{\delta}_2 + \mu_i + \epsilon_{it}$$

where

$\mathbf{X}_{1it}$ is a $1 \times k_1$ vector of observations on exogenous, time-varying variables assumed to be uncorrelated with μ_i and ϵ_{it};

$\mathbf{X}_{2it}$ is a $1 \times k_2$ vector of observations on endogenous, time-varying variables assumed to be (possibly) correlated with μ_i but orthogonal to ϵ_{it};

$\mathbf{Z}_{1i}$ is a $1 \times g_1$ vector of observations on exogenous, time-invariant variables assumed to be uncorrelated with μ_i and ϵ_{it};

$\mathbf{Z}_{2i}$ is a $1 \times g_2$ vector of observations on endogenous, time-invariant variables assumed to be (possibly) correlated μ_i but orthogonal to ϵ_{it};

μ_i is the unobserved, panel-level random effect that is assumed to have zero mean and finite variance σ_μ^2 and to be independent and identically distributed (i.i.d.) over the panels;

ϵ_{it} is the idiosyncratic error that is assumed to have zero mean and finite variance σ_ϵ^2 and to be i.i.d. over all the observations in the data;

$\boldsymbol{\beta}_1, \boldsymbol{\beta}_2, \boldsymbol{\delta}_1,$ and $\boldsymbol{\delta}_2$ are $k_1 \times 1$, $k_2 \times 1$, $g_1 \times 1$, and $g_2 \times 1$ coefficient vectors, respectively; and

$i = 1, \ldots, n$, where n is the number of panels in the sample and, for each i, $t = 1, \ldots, T_i$.

Because $\mathbf{X}_{2it}$ and $\mathbf{Z}_{2i}$ may be correlated with μ_i, the simple random-effects estimators—xtreg, re and xtreg, mle—are generally not consistent for the parameters in this model. Because the within estimator, xtreg, fe, removes the μ_i by mean-differencing the data before estimating β_1 and β_2, it is consistent for these parameters. However, in the process of removing the μ_i, the within estimator also eliminates the $\mathbf{Z}_{1i}$ and the $\mathbf{Z}_{2i}$. Thus it cannot estimate δ_1 nor δ_2. The Hausman–Taylor and Amemiya–MaCurdy estimators implemented in xthtaylor are designed to resolve this problem.

The within estimator consistently estimates β_1 and β_2. Using these estimates, we can obtain the within residuals, called $\widehat{d}_i$. Intermediate, albeit consistent, estimates of δ_1 and δ_2—called $\widehat{\delta}_{1\mathrm{IV}}$ and $\widehat{\delta}_{2\mathrm{IV}}$, respectively—are obtained by regressing the within residuals on $\mathbf{Z}_{1i}$ and $\mathbf{Z}_{2i}$, using $\mathbf{X}_{1it}$ and $\mathbf{Z}_{1i}$ as instruments. The order condition for identification requires that the number of variables in $\mathbf{X}_{1it}$, k_1, be at least as large as the number of elements in $\mathbf{Z}_{2i}$, g_2 and that there be sufficient correlation between the instruments and $\mathbf{Z}_{2i}$ to avoid a weak-instrument problem.

The within estimates of β_1 and β_2 and the intermediate estimates $\widehat{\delta}_{1\mathrm{IV}}$ and $\widehat{\delta}_{2\mathrm{IV}}$ can be used to obtain sets of within and overall residuals. These two sets of residuals can be used to estimate the variance components (see *Methods and formulas* for details).

The estimated variance components can then be used to perform a GLS transform on each of the variables. For what follows, define the general notation $\breve{w}_{it}$ to represent the GLS transform of the variable w_{it}, $\overline{w}_i$ to represent the within-panel mean of w_{it}, and $\widetilde{w}_{it}$ to represent the within transform of w_{it}. With this notational convention, the Hausman–Taylor (1981) estimator of the coefficients of interest can be obtained by the instrumental-variables regression

$$\breve{y}_{it} = \breve{\mathbf{X}}_{1it}\beta_1 + \breve{\mathbf{X}}_{2it}\beta_2 + \breve{\mathbf{Z}}_{1i}\delta_1 + \breve{\mathbf{Z}}_{2i}\delta_2 + \breve{\mu}_i + \breve{\epsilon}_{it} \tag{1}$$

using $\widetilde{\mathbf{X}}_{1it}$, $\widetilde{\mathbf{X}}_{2it}$, $\overline{\mathbf{X}}_{1i}$, $\overline{\mathbf{X}}_{2i}$, and $\mathbf{Z}_{1i}$ as instruments.

For the instruments to be valid, this estimator requires that $\overline{\mathbf{X}}_{1i.}$ and $\mathbf{Z}_{1i}$ be uncorrelated with the random-effect μ_i. More precisely, the instruments are valid when

$$\mathrm{plim}_{n\to\infty}\frac{1}{n}\sum_{i=1}^{n}\overline{\mathbf{X}}_{1i.}\mu_i = 0$$

and

$$\mathrm{plim}_{n\to\infty}\frac{1}{n}\sum_{i=1}^{n}\mathbf{Z}_{1i}\mu_i = 0$$

Amemiya and MaCurdy (1986) place stricter requirements on the instruments that vary within panels to obtain a more efficient estimator. Specifically, Amemiya and MaCurdy (1986) assume that $\mathbf{X}_{1it}$ is orthogonal to μ_i in every period; that is, $\mathrm{plim}_{n\to\infty}1/n\sum_{i=1}^{n}\mathbf{X}_{1it}\mu_i = 0$ for $t = 1, \ldots, T$. With this restriction, they derive the Amemiya–MaCurdy estimator as the instrumental-variables regression of (1) using instruments $\widetilde{\mathbf{X}}_{1it}$, $\widetilde{\mathbf{X}}_{2it}$, $\mathbf{X}_{1it}^*$, and $\mathbf{Z}_{1i}$. The order condition for the Amemiya–MaCurdy estimator is now $Tk_1 > g_2$. xthtaylor uses the Amemiya–MaCurdy estimator when the amacurdy option is specified.

Although the estimators implemented in xthtaylor and xtivreg (see [XT] **xtivreg**) use the method of instrumental variables, each command is designed for different problems. The estimators implemented in xtivreg assume that a subset of the explanatory variables in the model are correlated with the idiosyncratic error ϵ_{it}. In contrast, the Hausman–Taylor and Amemiya–MaCurdy estimators that are implemented in xthtaylor assume that some of the explanatory variables are correlated with the individual-level random effects, u_i, but that none of the explanatory variables are correlated with the idiosyncratic error, ϵ_{it}.

▷ Example 1

This example replicates the results of Baltagi and Khanti-Akom (1990, table II, column HT) using 595 observations on individuals over 1976–1982 that were extracted from the Panel Study of Income Dynamics (PSID). In the model, the log-transformed wage `lwage` is assumed to be a function of how long the person has worked for a firm, `wks`; binary variables indicating whether a person lives in a large metropolitan area or in the south, `smsa` and `south`; marital status is `ms`; years of education, `ed`; a quadratic of work experience, `exp` and `exp2`; occupation, `occ`; a binary variable indicating employment in a manufacture industry, `ind`; a binary variable indicating that wages are set by a union contract, `union`; a binary variable indicating gender, `fem`; and a binary variable indicating whether the individual is African American, `blk`.

We suspect that the time-varying variables `exp`, `exp2`, `wks`, `ms`, and `union` are all correlated with the unobserved individual random effect. We can inspect these variables to see if they exhibit sufficient within-panel variation to serve as their own instruments.

```
. use https://www.stata-press.com/data/r18/psidextract
. xtsum exp exp2 wks ms union
```

Variable		Mean	Std. dev.	Min	Max		Observations	
exp	overall	19.85378	10.96637	1	51	N =	4165
	between		10.79018	4	48	n =	595
	within		2.00024	16.85378	22.85378	T =	7
exp2	overall	514.405	496.9962	1	2601	N =	4165
	between		489.0495	20	2308	n =	595
	within		90.44581	231.405	807.405	T =	7
wks	overall	46.81152	5.129098	5	52	N =	4165
	between		3.284016	31.57143	51.57143	n =	595
	within		3.941881	12.2401	63.66867	T =	7
ms	overall	.8144058	.3888256	0	1	N =	4165
	between		.3686109	0	1	n =	595
	within		.1245274	−.0427371	1.671549	T =	7
union	overall	.3639856	.4812023	0	1	N =	4165
	between		.4543848	0	1	n =	595
	within		.1593351	−.4931573	1.221128	T =	7

We are also going to assume that the exogenous variables `occ`, `south`, `smsa`, `ind`, `fem`, and `blk` are instruments for the endogenous, time-invariant variable `ed`. The output below indicates that although `fem` appears to be a weak instrument, the remaining instruments are probably sufficiently correlated to identify the coefficient on `ed`. (See Baltagi and Khanti-Akom [1990] for more discussion.)

```
. correlate fem blk occ south smsa ind ed
(obs=4,165)
```

	fem	blk	occ	south	smsa	ind	ed
fem	1.0000						
blk	0.2086	1.0000					
occ	−0.0847	0.0837	1.0000				
south	0.0516	0.1218	0.0413	1.0000			
smsa	0.1044	0.1154	−0.2018	−0.1350	1.0000		
ind	−0.1778	−0.0475	0.2260	−0.0769	−0.0689	1.0000	
ed	−0.0012	−0.1196	−0.6194	−0.1216	0.1843	−0.2365	1.0000

We will assume that the correlations are strong enough and proceed with the estimation. The output below gives the Hausman–Taylor estimates for this model.

```
. xthtaylor lwage occ south smsa ind exp exp2 wks ms union fem blk ed,
> endog(exp exp2 wks ms union ed)
```

Hausman–Taylor estimation	Number of obs	=	4,165
Group variable: id	Number of groups	=	595

Obs per group:

min =	7
avg =	7
max =	7

Random effects u_i ~ i.i.d.	Wald chi2(12)	=	6891.87
	Prob > chi2	=	0.0000

lwage	Coefficient	Std. err.	z	P>|z|	[95% conf. interval]	
TVexogenous						
occ	-.0207047	.0137809	-1.50	0.133	-.0477149	.0063055
south	.0074398	.031955	0.23	0.816	-.0551908	.0700705
smsa	-.0418334	.0189581	-2.21	0.027	-.0789906	-.0046761
ind	.0136039	.0152374	0.89	0.372	-.0162608	.0434686
TVendogenous						
exp	.1131328	.002471	45.79	0.000	.1082898	.1179758
exp2	-.0004189	.0000546	-7.67	0.000	-.0005259	-.0003119
wks	.0008374	.0005997	1.40	0.163	-.0003381	.0020129
ms	-.0298508	.01898	-1.57	0.116	-.0670508	.0073493
union	.0327714	.0149084	2.20	0.028	.0035514	.0619914
TIexogenous						
fem	-.1309236	.126659	-1.03	0.301	-.3791707	.1173234
blk	-.2857479	.1557019	-1.84	0.066	-.5909179	.0194221
TIendogenous						
ed	.137944	.0212485	6.49	0.000	.0962977	.1795902
_cons	2.912726	.2836522	10.27	0.000	2.356778	3.468674
sigma_u	.94180304					
sigma_e	.15180273					
rho	.97467788	(fraction of variance due to u_i)				

Note: TV refers to time varying; TI refers to time invariant.

The estimated σ_μ and σ_ϵ are 0.9418 and 0.1518, respectively, indicating that a large fraction of the total error variance is attributed to μ_i. The z statistics indicate that several the coefficients may not be significantly different from zero. Whereas the coefficients on the time-invariant variables fem and blk have relatively large standard errors, the standard error for the coefficient on ed is relatively small.

Baltagi and Khanti-Akom (1990) also present evidence that the efficiency gains of the Amemiya–MaCurdy estimator over the Hausman–Taylor estimator are small for these data. This point is especially important given the additional restrictions that the estimator places on the data-generating process. The output below replicates the Baltagi and Khanti-Akom (1990) results from column AM of table II.

```
. xthtaylor lwage occ south smsa ind exp exp2 wks ms union fem blk ed,
> endog(exp exp2 wks ms union ed) amacurdy
```

Amemiya–MaCurdy estimation   Number of obs   =   4,165
Group variable: id   Number of groups   =   595

Time variable: t   Obs per group:
                                                             min =   7
                                                             avg =   7
                                                             max =   7

Random effects u_i ~ i.i.d.   Wald chi2(12)   =   6879.20
                                               Prob > chi2   =   0.0000

lwage	Coefficient	Std. err.	z	P>|z|	[95% conf. interval]	
TVexogenous						
occ	-.0208498	.0137653	-1.51	0.130	-.0478292	.0061297
south	.0072818	.0319365	0.23	0.820	-.0553126	.0698761
smsa	-.0419507	.0189471	-2.21	0.027	-.0790864	-.0048149
ind	.0136289	.015229	0.89	0.371	-.0162194	.0434771
TVendogenous						
exp	.1129704	.0024688	45.76	0.000	.1081316	.1178093
exp2	-.0004214	.0000546	-7.72	0.000	-.0005283	-.0003145
wks	.0008381	.0005995	1.40	0.162	-.0003368	.002013
ms	-.0300894	.0189674	-1.59	0.113	-.0672649	.0070861
union	.0324752	.0148939	2.18	0.029	.0032837	.0616667
TIexogenous						
fem	-.132008	.1266039	-1.04	0.297	-.380147	.1161311
blk	-.2859004	.1554857	-1.84	0.066	-.5906468	.0188459
TIendogenous						
ed	.1372049	.0205695	6.67	0.000	.0968894	.1775205
_cons	2.927338	.2751274	10.64	0.000	2.388098	3.466578
sigma_u	.94180304					
sigma_e	.15180273					
rho	.97467788	(fraction of variance due to u_i)				

Note: TV refers to time varying; TI refers to time invariant.

◁

□ Technical note

   We mentioned earlier that insufficient correlation between an endogenous variable and the instruments can give rise to a weak-instrument problem. Suppose that we simulate data for a model of the form

$$y = 3 + 3x_{1a} + 3x_{1b} + 3x_2 + 3z_1 + 3z_2 + u_i + e_{it}$$

and purposely construct the instruments so that they exhibit little correlation with the endogenous variable z_2.

```
. use https://www.stata-press.com/data/r18/xthtaylor1

. correlate ui z1 z2 x1a x1b x2 eit
(obs=10,000)
```

	ui	z1	z2	x1a	x1b	x2	eit
ui	1.0000						
z1	0.0268	1.0000					
z2	0.8777	0.0286	1.0000				
x1a	-0.0145	0.0065	-0.0034	1.0000			
x1b	0.0026	0.0079	0.0038	-0.0030	1.0000		
x2	0.8765	0.0191	0.7671	-0.0192	0.0037	1.0000	
eit	0.0060	-0.0198	0.0123	-0.0100	-0.0138	0.0092	1.0000

In the output below, weak instruments have serious consequences on the estimates produced by
xthtaylor. The estimate of the coefficient on z2 is three times larger than its true value, and its
standard error is rather large. Without sufficient correlation between the endogenous variable and
its instruments in a given sample, there is insufficient information for identifying the parameter.
Also, given the results of Stock, Wright, and Yogo (2002), weak instruments will cause serious size
distortions in any tests performed.

```
. xthtaylor yit x1a x1b x2 z1 z2, endog(x2 z2)
```

Hausman–Taylor estimation	Number of obs	=	10,000
Group variable: id	Number of groups	=	1,000

```
                                      Obs per group:
                                                   min =         10
                                                   avg =         10
                                                   max =         10
Random effects u_i ~ i.i.d.           Wald chi2(5)      =   24172.91
                                      Prob > chi2       =     0.0000
```

yit	Coefficient	Std. err.	z	P>|z|	[95% conf. interval]	
TVexogenous						
x1a	2.959736	.0330233	89.63	0.000	2.895011	3.02446
x1b	2.953891	.0333051	88.69	0.000	2.888614	3.019168
TVendogenous						
x2	3.022685	.033085	91.36	0.000	2.957839	3.08753
TIexogenous						
z1	2.709179	.587031	4.62	0.000	1.55862	3.859739
TIendogenous						
z2	9.525973	8.572966	1.11	0.266	-7.276732	26.32868
_cons	2.837072	.4276595	6.63	0.000	1.998875	3.675269

sigma_u	8.729479
sigma_e	3.1657492
rho	.88377062 (fraction of variance due to u_i)

Note: TV refers to time varying; TI refers to time invariant.

❑

▷ Example 2

Now let's consider why we might want to specify the constant(*varlist*$_{ti}$) option. For this example, we will use simulated data. In the output below, we fit a model over the full sample. Note the placement in the output of the coefficient on the exogenous variable x1c.

```
. use https://www.stata-press.com/data/r18/xthtaylor2
. xthtaylor yit x1a x1b x1c x2 z1 z2, endog(x2 z2)
Hausman-Taylor estimation                     Number of obs      =      10,000
Group variable: id                            Number of groups   =       1,000
                                              Obs per group:
                                                            min =          10
                                                            avg =          10
                                                            max =          10
Random effects u_i ~ i.i.d.                   Wald chi2(6)       =    10341.63
                                              Prob > chi2        =      0.0000
```

yit	Coefficient	Std. err.	z	P>|z|	[95% conf. interval]
TVexogenous						
x1a	3.023647	.0570274	53.02	0.000	2.911875	3.135418
x1b	2.966666	.0572659	51.81	0.000	2.854427	3.078905
x1c	.2355318	.123502	1.91	0.057	-.0065276	.4775912
TVendogenous						
x2	14.17476	3.128385	4.53	0.000	8.043234	20.30628
TIexogenous						
z1	1.741709	.4280022	4.07	0.000	.9028398	2.580578
TIendogenous						
z2	7.983849	.6970903	11.45	0.000	6.617577	9.350121
_cons	2.146038	.3794179	5.66	0.000	1.402393	2.889684
sigma_u	5.6787791					
sigma_e	3.1806188					
rho	.76120931	(fraction of variance due to u_i)				

Note: TV refers to time varying; TI refers to time invariant.

Now suppose that we want to fit the model using only the first eight periods. Below, x1c now appears under the TIexogenous heading rather than the TVexogenous heading because x1c is time invariant in the subsample defined by t<9.

```
. xthtaylor yit x1a x1b x1c x2 z1 z2 if t<9, endog(x2 z2)
```

Hausman–Taylor estimation	Number of obs	=	8,000
Group variable: id	Number of groups	=	1,000

```
                                       Obs per group:
                                                    min =           8
                                                    avg =           8
                                                    max =           8
Random effects u_i ~ i.i.d.            Wald chi2(6)      =    15354.87
                                       Prob > chi2       =      0.0000
```

yit	Coefficient	Std. err.	z	P>|z|	[95% conf. interval]	
TVexogenous						
x1a	3.051966	.0367026	83.15	0.000	2.98003	3.123901
x1b	2.967822	.0368144	80.62	0.000	2.895667	3.039977
TVendogenous						
x2	.7361217	3.199764	0.23	0.818	−5.5353	7.007543
TIexogenous						
x1c	3.215907	.5657191	5.68	0.000	2.107118	4.324696
z1	3.347644	.5819756	5.75	0.000	2.206992	4.488295
TIendogenous						
z2	2.010578	1.143982	1.76	0.079	−.231586	4.252742
_cons	3.257004	.5295828	6.15	0.000	2.219041	4.294967
sigma_u	15.445594					
sigma_e	3.175083					
rho	.95945606	(fraction of variance due to u_i)				

Note: TV refers to time varying; TI refers to time invariant.

To prevent a variable from becoming time invariant, you can use either $\mathtt{constant}(varlist_{ti})$ or $\mathtt{varying}(varlist_{tv})$. $\mathtt{constant}(varlist_{ti})$ specifies the subset of variables in *varlist* that are time invariant and requires the remaining variables in *varlist* to be time varying. If you specify $\mathtt{constant}(varlist_{ti})$ and any of the variables contained in $varlist_{ti}$ are time varying, or if any of the variables not contained in $varlist_{ti}$ are time invariant, xthtaylor will not perform the estimation and will issue an error message.

```
. xthtaylor yit x1a x1b x1c x2 z1 z2 if t<9, endog(x2 z2) constant(z1 z2)
x1c not included in constant().
r(198);
```

The same thing happens when you use the $\mathtt{varying}(varlist_{tv})$ option.

◁

Stored results

xthtaylor stores the following in e():

Scalars
e(N)	number of observations
e(N_g)	number of groups
e(df_m)	model degrees of freedom
e(df_r)	residual degrees of freedom (small only)
e(g_min)	smallest group size
e(g_avg)	average group size
e(g_max)	largest group size
e(Tcon)	1 if panels balanced, 0 otherwise
e(N_clust)	number of clusters
e(sigma_u)	panel-level standard deviation
e(sigma_e)	standard deviation of ϵ_{it}
e(chi2)	χ^2
e(rho)	ρ
e(F)	model F (small only)
e(Tbar)	harmonic mean of group sizes
e(rank)	rank of e(V)

Macros
e(cmd)	xthtaylor
e(cmdline)	command as typed
e(depvar)	name of dependent variable
e(ivar)	variable denoting groups
e(tvar)	variable denoting time within groups, amacurdy only
e(TVexogenous)	exogenous time-varying variables
e(TIexogenous)	exogenous time-invariant variables
e(TVendogenous)	endogenous time-varying variables
e(TIendogenous)	endogenous time-invariant variables
e(wtype)	weight type
e(wexp)	weight expression
e(title)	Hausman-Taylor or Amemiya-MaCurdy
e(clustvar)	name of cluster variable
e(chi2type)	Wald; type of model χ^2 test
e(vce)	*vcetype* specified in vce()
e(vcetype)	title used to label Std. err.
e(properties)	b V
e(predict)	program used to implement predict

Matrices
e(b)	coefficient vector
e(V)	variance–covariance matrix of the estimators
e(V_modelbased)	model-based variance

Functions
e(sample)	marks estimation sample

Methods and formulas

Consider an error-components model of the form

$$y_{it} = \mathbf{X}_{1it}\boldsymbol{\beta}_1 + \mathbf{X}_{2it}\boldsymbol{\beta}_2 + \mathbf{Z}_{1i}\delta_1 + \mathbf{Z}_{2i}\delta_2 + \mu_i + \epsilon_{it} \tag{2}$$

for $i = 1, \ldots, n$ and, for each i, $t = 1, \ldots, T_i$, of which T_i periods are observed; n is the number of panels in the sample. The covariates in $\mathbf{X}$ are time varying, and the covariates in $\mathbf{Z}$ are time invariant. Both $\mathbf{X}$ and $\mathbf{Z}$ are decomposed into two parts. The covariates in $\mathbf{X}_1$ and $\mathbf{Z}_1$ are assumed to be uncorrelated with μ_i and e_{it}, whereas the covariates in $\mathbf{X}_2$ and $\mathbf{Z}_2$ are allowed to be correlated with μ_i but not with ϵ_{it}. Hausman and Taylor (1981) suggest an instrumental-variable estimator for this model.

For some variable w, the within transformation of w is defined as

$$\widetilde{w}_{it} = w_{it} - \overline{w}_{i.} \qquad \overline{w}_{i.} = \frac{1}{T_i} \sum_{t=1}^{T_i} w_{it}$$

Because the within estimator removes $\mathbf{Z}$, the within transformation reduces the model to

$$\widetilde{y}_{it} = \widetilde{\mathbf{X}}_{1it}\boldsymbol{\beta}_1 + \widetilde{\mathbf{X}}_{2it}\boldsymbol{\beta}_2 + \widetilde{\epsilon}_{it}$$

The within estimators $\widehat{\beta}_{1w}$ and $\widehat{\beta}_{2w}$ are consistent for $\boldsymbol{\beta}_1$ and $\boldsymbol{\beta}_2$, but they may not be efficient. Also, note that the within estimator cannot estimate $\boldsymbol{\delta}_1$ and $\boldsymbol{\delta}_2$.

From the within estimator, we can obtain an estimate of the idiosyncratic error component, σ_ϵ^2, as

$$\widehat{\sigma}_\epsilon^2 = \frac{\text{RSS}}{N - n}$$

where RSS is the residual sum of squares from the within regression and N is the total number of observations in the sample.

Using the results of the within estimation, we can define

$$\overline{d}_{it} = \overline{y}_{it} - \overline{X}_{1it}\widehat{\beta}_{1w} - \overline{X}_{2it}\widehat{\beta}_{2w}$$

where $\overline{y}_{it}$, $\overline{X}_{1it}$, and $\overline{X}_{2it}$ contain the panel level means of these variables in all observations.

Regressing $\overline{d}_{it}$ on $\mathbf{Z}_1$ and $\mathbf{Z}_2$, using $\mathbf{X}_1$ and $\mathbf{Z}_1$ as instruments, provides intermediate, consistent estimates of $\boldsymbol{\delta}_1$ and $\boldsymbol{\delta}_2$, which we will call $\widehat{\boldsymbol{\delta}}_{1\text{IV}}$ and $\widehat{\boldsymbol{\delta}}_{2\text{IV}}$.

Using the within estimates, $\widehat{\boldsymbol{\delta}}_{1\text{IV}}$, and $\widehat{\boldsymbol{\delta}}_{2\text{IV}}$, we can obtain an estimate of the variance of the random effect, σ_μ^2. First, let

$$\widehat{e}_{it} = \left(y_{it} - \mathbf{X}_{1it}\widehat{\beta}_{1w} - \mathbf{X}_{2it}\widehat{\beta}_{2w} - \mathbf{Z}_{1it}\widehat{\boldsymbol{\delta}}_{1\text{IV}} - \mathbf{Z}_{2it}\widehat{\boldsymbol{\delta}}_{2\text{IV}} \right)$$

Then define

$$s^2 = \frac{1}{N} \sum_{i=1}^{n} \sum_{t=1}^{T_i} \left(\frac{1}{T_i} \sum_{t=1}^{T_i} \widehat{e}_{it} \right)^2$$

Hausman and Taylor (1981) showed that, for balanced panels,

$$\text{plim}_{n\to\infty} s^2 = T\sigma_\mu^2 + \sigma_\epsilon^2$$

For unbalanced panels,

$$\text{plim}_{n\to\infty} s^2 = \overline{T}\sigma_\mu^2 + \sigma_\epsilon^2$$

where

$$\overline{T} = \frac{n}{\sum_{i=1}^{n} \frac{1}{T_i}}$$

After we plug in $\widehat{\sigma}_\epsilon^2$, our consistent estimate for σ_ϵ^2, a little algebra suggests the estimate

$$\widehat{\sigma}_\mu^2 = (s^2 - \widehat{\sigma}_\epsilon^2)(\overline{T})^{-1}$$

Define $\widehat{\theta}_i$ as

$$\widehat{\theta}_i = 1 - \left(\frac{\widehat{\sigma}_\epsilon^2}{\widehat{\sigma}_\epsilon^2 + T_i \widehat{\sigma}_\mu^2} \right)^{\frac{1}{2}}$$

With $\widehat{\theta}_i$ in hand, we can perform the standard random-effects GLS transform on each of the variables. The transform is given by

$$w_{it}^* = w_{it} - \widehat{\theta}_i \overline{w}_{i.}$$

where $\overline{w}_{i.}$ is the within-panel mean.

We can then obtain the Hausman–Taylor estimates of the coefficients in (2) and the conventional VCE by fitting an instrumental-variables regression of the GLS-transformed y_{it}^* on $\mathbf{X}_{it}^*$ and $\mathbf{Z}_{it}^*$, with instruments $\widetilde{\mathbf{X}}_{it}$, $\overline{\mathbf{X}}_{1i.}$, and $\mathbf{Z}_{1i}$.

We can obtain Amemiya–MaCurdy estimates of the coefficients in (2) and the conventional VCE by fitting an instrumental-variables regression of the GLS-transformed y_{it}^* on $\mathbf{X}_{it}^*$ and $\mathbf{Z}_{it}^*$, using $\widetilde{\mathbf{X}}_{it}$, $\breve{\mathbf{X}}_{1it}$, and $\mathbf{Z}_{1i}$ as instruments, where $\breve{\mathbf{X}}_{1it} = \mathbf{X}_{1i1}, \mathbf{X}_{1i2}, \ldots, \mathbf{X}_{1iT_i}$. The order condition for the Amemiya–MaCurdy estimator is $Tk_1 > g_2$, and this estimator is available only for balanced panels.

References

Amemiya, T., and T. E. MaCurdy. 1986. Instrumental-variable estimation of an error-components model. *Econometrica* 54: 869–880. https://doi.org/10.2307/1912840.

Baltagi, B. H. 2009. *A Companion to Econometric Analysis of Panel Data*. Chichester, UK: Wiley.

———. 2013. *Econometric Analysis of Panel Data*. 5th ed. Chichester, UK: Wiley.

Baltagi, B. H., and S. Khanti-Akom. 1990. On efficient estimation with panel data: An empirical comparison of instrumental variables estimators. *Journal of Applied Econometrics* 5: 401–406. https://doi.org/10.1002/jae.3950050408.

Hausman, J. A., and W. E. Taylor. 1981. Panel data and unobservable individual effects. *Econometrica* 49: 1377–1398. https://doi.org/10.2307/1911406.

Stock, J. H., J. H. Wright, and M. Yogo. 2002. A survey of weak instruments and weak identification in generalized method of moments. *Journal of Business and Economic Statistics* 20: 518–529. https://doi.org/10.1198/073500102288618658.

Also see

[XT] **xthtaylor postestimation** — Postestimation tools for xthtaylor

[XT] **xtivreg** — Instrumental variables and two-stage least squares for panel-data models

[XT] **xtreg** — Fixed-, between-, and random-effects and population-averaged linear models[+]

[XT] **xtset** — Declare data to be panel data

[U] **20 Estimation and postestimation commands**

For suggested citations, see the FAQ on citing Stata documentation.