

TVTIE

Time-varying true individual-effects stochastic frontier models

Stata user guide and worked examples

Levent Kutlu
School of Economics and Finance
The University of Texas Rio Grande Valley

Command version 1.2.0 · September 2026

Install and begin

```
net install tvtie, from("https://leventkutlu.com/stata/tvtie") replace
help tvtie
help tvtie_postestimation
```

TVTIE requires Stata 16 or later and has no additional package dependencies. The examples in this guide were run in Stata/MP 19 on Windows. An Internet connection is needed to install the command and load the hosted example data.

The command estimates the model of Kutlu, Tran, and Tsionas (2019). This guide explains its Stata syntax and interpretation, including extensions available in this implementation. The examples use simulated data and do not reproduce the paper's empirical tables.

Guide contents

Pages	Topic
2–3	Model, individual effects, and heterogeneity rank
4–5	Endogenous terms, instruments, and estimation controls
6–8	Cost and production frontiers: examples 1–8
9–10	Heterogeneity, endogeneity, and individual-trend tests
11–12	Predictions, saved results, troubleshooting, and resources

2. Model and interpretation

TVTIE separates individual heterogeneity, two-sided frontier noise, and nonnegative inefficiency. Individual heterogeneity may vary over time through a set of observed basis variables. Each panel has its own coefficients on this basis.

```
y_it = H_it alpha_i + X_it beta - s h_it u_i* + v_it  
s = 1 for production; s = -1 for cost  
h_it = exp(U_it phi / 2); sigma_u^2 = exp(phi_0)
```

The same nonnegative latent component u_i^* applies throughout a panel. The scaling function h_{it} allows its contribution to change over time. The columns of H_{it} describe individual heterogeneity; X_{it} contains frontier regressors; U_{it} contains the variables specified in `uhet()`, abbreviated `u()`.

What the reported equations mean

Equation or quantity	Interpretation
frontier	Coefficients beta on frontier regressors.
usigma	Coefficients in $\log(\sigma_u^2 h_{it}^2) = \phi_0 + U_{it} \phi$. A scale intercept is always estimated.
Insig2r	Log variance of the frontier noise conditional on the reduced-form disturbances.
eta	Endogenous correction coefficients. Joint $\eta = 0$ is the null of the endogeneity test.
kappa	For <code>distribution(tnormal)</code> , μ/σ_u for the underlying Gaussian before truncation.

Distribution and frontier direction

Production is the default: inefficiency lowers output relative to the frontier. The cost option reverses that sign. Half-normal inefficiency is the default. `distribution(tnormal)` estimates the location of the underlying normal distribution before truncation at zero. Both specifications give nonnegative inefficiency.

Projection and the likelihood

An orthogonal projection within each panel removes $H_{it} \alpha_i$ before likelihood maximization. Consequently, regressors and scaling variation must survive that projection. Raw observations and transformed observations differ: the latter total is the sum of T_i minus the heterogeneity rank over retained panels. Likelihood values from different heterogeneity spaces cannot be compared by an ordinary likelihood-ratio test.

3. Individual effects and rank

Specification	Individual-effect basis
Default, or trend(2)	Individual intercept, linear time trend, and quadratic time trend. Rank 3.
trend(0)	Individual intercept only. Rank 1.
heterogeneity(t w)	Individual intercept plus t and w. Rank 3 if the columns are independent in every panel.
heterogeneity(t w) nohetconstant	Only t and w. Rank 2 if independent in every panel.
nohet	No individual effects. A common frontier intercept is ordinarily estimated.

“Heterogeneity rank per panel = 3” means that three linearly independent basis columns are removed within each panel. Each panel has its own three coefficients. It is not the number of panels, latent classes, or inefficiency components.

The basis need not consist of time trends

The heterogeneity() option accepts other numeric covariates. These variables need adequate within-panel variation and must form a full-rank basis in every retained panel. A panel-constant covariate duplicates an automatic individual intercept, so it cannot supply a separate basis column. Avoid a basis that absorbs a frontier regressor or all fitted inefficiency scaling variation.

Two different constants

nohetconstant omits the automatically added individual intercept. It does not remove a constant explicitly listed in heterogeneity(). By contrast, noconstant controls the common frontier intercept when it is estimable. That common intercept is already absorbed when the individual-effect basis spans a constant.

Time normalization and short panels

With the automatic intercept, polynomial time is centered and scaled for numerical stability. estat heterogeneity returns coefficients on this transformed basis. Its center and scale are e(time_center) and e(time_scale). Without an intercept, time is scaled but not centered, because centering would change the restricted model.

Each retained panel needs more observations than the basis rank. dropshort excludes panels with no remaining residual dimension; it does not repair collinearity or other rank failures. Unbalanced panels are supported. Duplicate panel-time pairs are rejected.

4. Endogenous terms and instruments

Declare each primitive endogenous covariate in `en()`. The command jointly estimates a reduced-form equation for each primitive variable. Endogeneity can occur in the frontier or in the inefficiency scaling equation. `i()` lists the excluded instruments.

One endogenous variable and its square

```
tvttie y c.x##c.x, nohet u(u1) en(x) i(z)
```

Here `x` is the primitive endogenous variable. Continuous factor notation identifies `x` squared as a function of `x`. The command fits one reduced-form equation and excludes both endogenous frontier terms from its automatic included instrument set. Only `z` is supplied as an excluded instrument. It does not generate `z` squared.

```
generate double xsq=x^2  
tvttie y x xsq, nohet u(u1) en(x) endof(xsq) i(z)
```

This manually generated specification is equivalent for estimation. `endof()` abbreviates `endofunctions()`. Naming the variable `xsq` does not tell Stata how it was generated, so `en(x)` alone is insufficient when a separately named square appears in the model. List manually generated endogenous functions explicitly.

Interactions and bootstrap reconstruction

Continuous interactions involving an endogenous variable are recognized in the same way. For bootstrap tests, use continuous factor notation. The bootstrap must reconstruct the transformed term whenever a new primitive endogenous variable is drawn; it rejects manually declared `endof()` variables whose formulas are unavailable.

Included regressors and exclusion restrictions

By default, reduced forms include distinct frontier and scaling covariates, excluding declared endogenous variables and their functions. The excluded instruments are then added. `rfvars()` replaces the automatic included reduced-form regressors. It does not generate transformations or establish exclusion restrictions.

At least one distinct excluded instrument per primitive endogenous variable must survive projection. The complete projected instrument matrix must have full column rank. Instrument validity remains a modeling assumption. Neither the endogeneity test nor a large first-stage *F* statistic proves it.

5. Estimation controls

tvttie depvar [indepvars] [if] [in], options

Option	Default or role
id(varname), time(varname)	Use the panel and time variables declared with xtset unless overridden.
uhet(varlist), or u(varlist)	Covariates in the log squared inefficiency scale. Do not include a constant variable.
vce(oim)	Inverse observed information of the full joint likelihood; default.
vce(robust)	Sandwich covariance using panel scores.
vce(cluster clustvar)	Scores summed within larger clusters. Each panel must belong wholly to one cluster.
starts(#), seed(#)	8 optimization starts; seed 190119. The user's random-number state is restored.
iterate(#), tolerance(#)	500 BFGS iterations per start; tolerance 1e-7. Newton refinement follows.
from(matname)	First starting vector in the parameter order of e(b).
nolog, level(#)	Suppress progress output; control the displayed confidence level.

Covariance choice

The robust and cluster covariances use no finite-sample multiplier. They require more independent panels or larger clusters than full estimated parameters. Their justification relies on many independent units and adequate identification. They do not correct a misspecified likelihood or supply weak-instrument-robust inference. The parametric bootstrap commands require vce(oim).

Checking the solution

The command chooses the highest likelihood among accepted stationary endpoints. It checks the scaled full score and the observed information. More starts can reveal another optimum, but do not guarantee a global maximum. Inspect estat diagnostics when results are sensitive to starting values or a specification fails.

Only numeric variables are accepted. The frontier supports continuous factor notation, but not categorical factors, time-series operators, or weights. Generate lags and custom basis terms before estimation, and account for any resulting missing observations.

6. Cost frontiers and predictions

1. Exogenous cost frontier

The cost frontier uses two scaling covariates and the default quadratic individual trends. The hosted data are simulated.

```
use "https://leventkutlu.com/stata/tvtie/tvtie_example.dta", clear
xtset id period
tvtie y_exo x1 x2, u(u1 u2) cost nolog
estat diagnostics
```

2. Endogenous cost frontier

Declare primitive endogenous covariates in `en()` and the excluded instruments in `i()`. The reported endogeneity test is a joint Wald test.

```
use "https://leventkutlu.com/stata/tvtie/tvtie_example.dta", clear
xtset id period
tvtie y x1 x2, u(u1 u2) en(x2 u2) i(z3 z4) cost nolog
estat endogeneity
estat firststage, detail
estat components
```

3. Predictions and individual coefficients

After example 2, compare the two efficiency summaries. Posterior quantiles condition on fitted parameters; they are not confidence limits.

```
predict double te, efficiency
predict double mean_te, meante
predict double te_q025, tequantile(.025)
predict double te_q975, tequantile(.975)
predict double individual_effect, fe
summarize te mean_te te_q025 te_q975
estat heterogeneity
matrix panel_coefficients=r(alpha)
matrix panel_identifiers=r(panel_id)
```

7. Choosing individual heterogeneity

4. Individual intercepts and user-specified heterogeneity

A `trend(0)` basis contains individual intercepts. Heterogeneity variables need not be powers of time. The custom basis below uses time and `u1`, with an automatic individual intercept. Removing that automatic intercept leaves two basis columns.

```
use "https://leventkutlu.com/stata/tvtie/tvtie_example.dta", clear
xtset id period
tvtie y_exo x1 x2, u(u1 u2) cost trend(0) nolog
tvtie y_exo x1 x2, u(u1 u2) cost heterogeneity(time u1) nolog
tvtie y_exo x1 x2, u(u1 u2) cost heterogeneity(time u1) nohetconstant nolog
```

5. An explicit individual intercept

An explicitly supplied constant remains in the basis when `nohetconstant` is specified. With one, time, and time squared, the heterogeneity rank per panel is three.

```
use "https://leventkutlu.com/stata/tvtie/tvtie_example.dta", clear
xtset id period
generate double one=1
generate double time2=time^2
tvtie y_exo x1 x2, u(u1 u2) cost heterogeneity(one time time2) nohetconstant nolog
```

Choosing a substantive specification

Custom basis variables describe how individual heterogeneity may change. They should follow the economic specification and data structure. Increasing basis rank removes more within-panel variation and leaves fewer transformed observations. A basis with good numerical rank can still remove the variation needed to identify frontier or scaling parameters.

The custom basis in example 4 illustrates syntax. It is not a claim that this alternative matches the simulated data-generating process. The explicit constant in example 5 reproduces the span of the default quadratic basis; recovered coefficients may use different time coordinates.

8. A nonlinear endogenous frontier

6. A nonlinear endogenous production frontier

Continuous factor notation identifies the square as a function of x . There is one primitive endogenous equation. Only z is added as the excluded instrument; no powers of z are generated.

```
use "https://leventkutlu.com/stata/tvtie/tvtie_nonlinear.dta", clear
xtset id period
tvtie y c.x##c.x, nohet u(u1) en(x) i(z) nolog
estat firststage, detail
```

7. A manually generated endogenous square

For a named variable xsq , declare its status with `endof(xsq)`, an abbreviation of `endofunctions(xsq)`. Use example 6 when requesting a bootstrap, so the square can be regenerated in every draw.

```
use "https://leventkutlu.com/stata/tvtie/tvtie_nonlinear.dta", clear
xtset id period
generate double xsq=x^2
tvtie y x xsq, nohet u(u1) en(x) endof(xsq) i(z) nolog
```

8. Panel-robust covariance

The robust covariance uses independent panel scores. It does not make the endogeneity test robust to weak instruments or repair a misspecified likelihood.

```
use "https://leventkutlu.com/stata/tvtie/tvtie_nonlinear.dta", clear
xtset id period
tvtie y c.x##c.x, nohet u(u1) en(x) i(z) vce(robust) nolog
```

Selected output from example 6

Quantity	Estimate
Frontier coefficient on x	0.6260762
Frontier coefficient on x squared	0.1309804
eta coefficient for x	0.3564504

These estimates come from the hosted simulated production dataset, with 400 observations and 50 panels. Examples 6 and 7 gave matching coefficients and log likelihoods. The values illustrate reproducibility of the syntax; they are not evidence about empirical size or power.

9. Testing heterogeneity and endogeneity

Question	Fit before the test	Command
Any individual heterogeneity?	nohet with a common frontier intercept	estat hettest
Time variation beyond individual intercepts?	trend(0)	estat hettest, timevarying
Joint endogeneity?	An endogenous model	estat endogeneity

The endogeneity command uses an asymptotic joint Wald test by default, and the same test appears automatically after endogenous estimation. Both heterogeneity tests use a parametric bootstrap by default. Add asymptotic to a heterogeneity test to request the ordinary F reference.

9. Heterogeneity and endogeneity bootstrap tests

Fit the nohet null before the omnibus heterogeneity test. Endogeneity requires a separate null-specific bootstrap. These examples use 99 draws for illustration; consider 999 for reported inference, subject to the Stata edition matrix limit.

```
use "https://leventkutlu.com/stata/tvtie/tvtie_nonlinear.dta", clear
xtset id period
tvtie y c.x##c.x, nohet u(u1) en(x) i(z) nolog
estat hettest, reps(99) seed(271828)
matrix heterogeneity_draws=r(draws)
estat endogeneity, bootstrap reps(99) seed(314159)
matrix endogeneity_draws=r(draws)
```

Selected bootstrap output

Test in example 9	Statistic	p-value	Failed draws
Omnibus heterogeneity	0.69900	0.84000	0 of 99
Joint endogeneity	67.91135	0.01000	0 of 99

Each test uses its own fitted null and simulated reference distribution. A single joint endogeneity procedure covers all eta coefficients; a separate procedure for each endogenous variable is unnecessary. One bootstrap draw is insufficient.

10. Individual trends and bootstrap interpretation

10. Individual trends beyond individual intercepts

Fit the individual-intercept null with `trend(0)`, then test for time variation. Bootstrap is the default. The `timevarying` option does not test whether individual intercepts exist.

```
use "https://leventkutlu.com/stata/tvtie/tvtie_example.dta", clear
xtset id period
tvtie y_exo x1 x2, u(u1 u2) cost trend(0) nolog
estat hettest, timevarying trend(2) reps(99) seed(161803)
```

This simulated cost-frontier example gave a statistic of 10.19968 and bootstrap p-value 0.01000, with no failed draws among 99 attempts. It illustrates testing a constant individual effect against individual linear and quadratic trends, not a simulation study of test calibration.

Why bootstrap these diagnostics?

The heterogeneity statistic comes from an auxiliary regression whose response uses estimated frontier coefficients and predicted inefficiency. The ordinary F reference does not incorporate all consequences of those generated quantities. The parametric bootstrap repeats the estimation and prediction under the null. It is the default for both the omnibus and individual-trend diagnostics.

The endogeneity Wald test is an asymptotic test based on the joint likelihood covariance. Its bootstrap version is optional. Neither version protects against weak instruments or a misspecified distribution. Bootstrap validity depends on the fitted null being an adequate model for the sampling process.

Null-specific draws

The omnibus null omits individual effects. The individual-trend null permits individual intercepts. The endogeneity null sets η to zero. Their draws are therefore not interchangeable. Primitive endogenous variables, latent inefficiency, and frontier noise are generated jointly under the relevant null. Exogenous covariates and instruments remain fixed. With individual effects under the null, simulation conditions on components absorbed by that basis.

Replication counts and failures

Defaults are 199 draws, seed 873921, and two optimization starts per draw. At least 19 draws are required. More draws reduce Monte Carlo uncertainty, not model misspecification. The Stata edition's matrix limit caps `reps()`, because the command returns every attempted draw in `r(draws)`.

The p-value is $(1 + \text{exceedances}) / (B + 1)$, where B is the requested draw count. If any draw fails, `r(p)` is missing. `r(p_lower)` and `r(p_upper)` bound the p-value while retaining failed draws in the denominator. Inspect `r(failed)`, these bounds, and `r(draws)`; do not silently discard failures. The original data, estimates, and random-number state are restored.

11. Predictions and saved results

predict option	Meaning
xb	Frontier component $X_{it}\beta$; default. Excludes individual effects.
inefficiency	Conditional mean of nonnegative inefficiency.
efficiency	$\exp(-E[u_{it} \mid \text{transformed panel data}])$, as in the paper.
meante	$E[\exp(-u_{it}) \mid \text{transformed panel data}]$. At least efficiency, up to rounding.
fe	Recovered individual-effect contribution $H_{it}\alpha_i$.
cf	Projected reduced-form correction; zero for an exogenous model.
fitted, residuals	Plug-in fitted decomposition; observed response minus that fitted value.
uscale	$h_{it} = \exp(U_{it}\phi/2)$, excluding σ_u .
u0, u0sd	Posterior mean and standard deviation of the original panel latent component.
uquantile(p), tequantile(p)	Conditional quantiles for inefficiency or efficiency, with $0 < p < 1$.

Use `predict double` for full precision. Efficiency is on a zero-to-one scale. Quantiles condition on estimated parameters and are not confidence intervals that account for parameter estimation. Out-of-sample prediction is not supplied. The complete original estimation sample and fitted variables must remain unchanged, though row order may change.

Useful stored results

`e(b)` and `e(V)` contain full joint parameter estimates and covariance. `e(ll)`, `e(N)`, `e(N_g)`, `e(N_transformed)`, and `e(k_het)` give the likelihood and sample dimensions. `e(sample)` identifies the retained estimation sample. `e(starts)` and `e(diagnostics)` contain numerical checks. See `help tvtie` for the complete list.

```
matrix list e(b)
ereturn list
estat diagnostics
estat components
```

`estat heterogeneity` returns coefficient rows in `r(alpha)` and matching panel identifiers in `r(panel_id)`. It is not a test. Save those matrices before another `r-class` command overwrites them. `estimates store` and `estimates restore preserve` a fitted specification. Reloading estimates from disk also requires the exact original data and a correctly reconstructed `e(sample)`.

12. Data, troubleshooting, and resources

Public simulated datasets

File	Contents
tvte_example.dta	900 observations, 100 panels, 9 periods. Cost responses <code>y_exo</code> and <code>y</code> ; frontier <code>x1</code> and <code>x2</code> ; scaling <code>u1</code> and <code>u2</code> ; instruments <code>z3</code> and <code>z4</code> ; panel id and time variables.
tvte_nonlinear.dta	400 observations, 50 panels, 8 periods. Production response <code>y</code> ; endogenous <code>x</code> ; scaling covariate <code>u1</code> ; instrument <code>z</code> ; id and period.

The data contain no empirical observations from the published application. All help examples use stable HTTPS links. Run the data-loading commands in `help tvte` to open either dataset directly in Stata. The software page provides the installation command, this guide, and one do-file with three introductory examples. The do-file loads the data directly from the hosted URLs.

If a command stops

Symptom	Check
No acceptable optimum	Inspect scaling and rank, rescale poorly scaled covariates, examine diagnostics, and try additional starts. Do not treat convergence alone as proof of a correct specification.
Rank or projection error	Check variation within each panel, instrument survival, duplicate basis columns, and the number of usable periods.
Bootstrap rejected	Fit the appropriate null, use <code>vce(oim)</code> , and use continuous factor notation for nonlinear endogenous terms.
Missing bootstrap p-value	Inspect failed draws and the returned p-value bounds. More successful fits do not justify dropping failed attempts.
Data link unavailable	Check the Internet connection and copy the complete HTTPS command from the help file.

Help and reference

[Software page](#) · [Example do-file](#). In Stata, type `help tvte` or `help tvte_postestimation` for executable examples and the full command reference.

Kutlu, L., K. C. Tran, and M. G. Tsionas. 2019. A time-varying true individual effects model with endogenous regressors. *Journal of Econometrics* 211: 539–559. doi:10.1016/j.jeconom.2019.01.014.

Software author: Levent Kutlu. Copyright (C) 2026 Levent Kutlu. The software is distributed under GNU GPL version 3; see the supplied LICENSE.