

**xtfrontier** — Stochastic frontier models for panel data[Description](#)[Menu](#)[Options for time-invariant model](#)[Remarks and examples](#)[Methods and formulas](#)[Also see](#)[Quick start](#)[Syntax](#)[Options for time-varying decay model](#)[Stored results](#)[References](#)

## Description

`xtfrontier` fits stochastic production or cost frontier models for panel data where the disturbance term is a mixture of an inefficiency term and the idiosyncratic error. `xtfrontier` can fit a time-invariant model, in which the inefficiency term is assumed to have a truncated-normal distribution, or a time-varying decay model, in which the inefficiency term is modeled as a truncated-normal random variable multiplied by a function of time.

`xtfrontier` expects that the dependent variable and independent variables are on the natural logarithm scale; this transformation must be performed before estimation takes place.

## Quick start

Stochastic production frontier regression of `lny` on `lnx1` and `lnx2` with time-invariant inefficiency using `xtset` data

```
xtfrontier lny lnx1 lnx2, ti
```

Stochastic cost frontier regression of `lny` on `lnx1` and `lnx2` with time-invariant inefficiency

```
xtfrontier lny lnx1 lnx2, ti cost
```

Time-varying decay model for production

```
xtfrontier lny lnx1 lnx2, tvd
```

## Menu

Statistics > Longitudinal/panel data > Frontier models

## Syntax

### Time-invariant model

```
xtfrontier depvar [indepvars] [if] [in] [weight], ti [ti_options]
```

### Time-varying decay model

```
xtfrontier depvar [indepvars] [if] [in] [weight], tvd [tvd_options]
```

| <i>ti_options</i>                         | Description  |
|---|--|
| <hr/>                                     |  |
| Model                                     |  |
| <u>noconstant</u>                         | suppress constant term   |
| <u>ti</u>                                 | use time-invariant model   |
| <u>cost</u>                               | fit cost frontier model  |
| <u>constraints</u> ( <i>constraints</i> ) | apply specified linear constraints   |
| <u>collinear</u>                          | keep collinear variables   |
| SE  |  |
| <u>vce</u> ( <i>vcetype</i> )             | <i>vcetype</i> may be <code>oim</code> , <code>bootstrap</code> , or <code>jackknife</code>  |
| Reporting                                 |  |
| <u>level</u> (#)                          | set confidence level; default is <code>level(95)</code>  |
| <u>nocnsreport</u>                        | do not display constraints   |
| <u>display_options</u>                    | control columns and column formats, row spacing, line width, display of omitted variables and base and empty cells, and factor-variable labeling |
| Maximization                              |  |
| <u>maximize_options</u>                   | control the maximization process; seldom used  |
| <u>coeflegend</u>                         | display legend instead of statistics   |

---

| <i>tvd_options</i>                           | Description  |
|--|--|
| Model  |  |
| <code>noconstant</code>                      | suppress constant term   |
| <code>tvd</code>                             | use time-varying decay model   |
| <code>cost</code>                            | fit cost frontier model  |
| <code>constraints(<i>constraints</i>)</code> | apply specified linear constraints   |
| <code>collinear</code>                       | keep collinear variables   |
| SE   |  |
| <code>vce(<i>vcetype</i>)</code>             | <i>vcetype</i> may be <code>oim</code> , <code>bootstrap</code> , or <code>jackknife</code>  |
| Reporting                                    |  |
| <code>level(#)</code>                        | set confidence level; default is <code>level(95)</code>  |
| <code>nocnsreport</code>                     | do not display constraints   |
| <code>display_options</code>                 | control columns and column formats, row spacing, line width, display of omitted variables and base and empty cells, and factor-variable labeling |
| Maximization                                 |  |
| <code>maximize_options</code>                | control the maximization process; seldom used  |
| <code>coeflegend</code>                      | display legend instead of statistics   |

A panel variable must be specified. For `xtfrontier`, `tvd`, a time variable must also be specified. Use `xtset`; see [XT] `xtset`.

`indepvars` may contain factor variables; see [U] 11.4.3 Factor variables.

`devar` and `indepvars` may contain time-series operators; see [U] 11.4.4 Time-series varlists.

`by`, `fp`, and `statsby` are allowed; see [U] 11.1.10 Prefix commands.

`fweights` and `iweights` are allowed; see [U] 11.1.6 weight. Weights must be constant within panel.

`coeflegend` does not appear in the dialog box.

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

## Options for time-invariant model

### Model

`noconstant`; see [R] estimation options.

`ti` specifies that the parameters of the time-invariant technical inefficiency model be estimated.

`cost` specifies that the frontier model be fit in terms of a cost function instead of a production function. By default, `xtfrontier` fits a production frontier model.

`constraints(constraints)`, `collinear`; see [R] estimation options.

### SE

`vce(vcetype)` specifies the type of standard error reported, which includes types that are derived from asymptotic theory (`oim`) and that use bootstrap or jackknife methods (`bootstrap`, `jackknife`); see [XT] `vce_options`.

### Reporting

---

`level(#)`; see [R] [estimation options](#).

`nocnsreport`; see [R] [estimation options](#).

`display_options`: `nocl`, `nopvalues`, `noomitted`, `vsquish`, `noemptycells`, `baselevels`, `allbaselevels`, `nofvlabel`, `fvwrap(#)`, `fvwrapon(style)`, `cformat(%fmt)`, `pformat(%fmt)`, `sformat(%fmt)`, and `nolstretch`; see [R] [estimation options](#).

### Maximization

---

`maximize_options`: `difficult`, `technique(algorithm_spec)`, `iterate(#)`, `[no]log`, `trace`, `gradient`, `showstep`, `hessian`, `showtolerance`, `tolerance(#)`, `ltolerance(#)`, `nrtolerance(#)`, `nonrtolerance`, and `from(init_specs)`; see [R] [maximize](#). These options are seldom used.

The following option is available with `xtfreontier` but is not shown in the dialog box:

`coeflegend`; see [R] [estimation options](#).

## Options for time-varying decay model

### Model

---

`noconstant`; see [R] [estimation options](#).

`tvd` specifies that the parameters of the time-varying decay model be estimated.

`cost` specifies that the frontier model be fit in terms of a cost function instead of a production function. By default, `xtfreontier` fits a production frontier model.

`constraints(constraints)`, `collinear`; see [R] [estimation options](#).

### SE

---

`vce(vcetype)` specifies the type of standard error reported, which includes types that are derived from asymptotic theory (`oim`) and that use bootstrap or jackknife methods (`bootstrap`, `jackknife`); see [XT] [vce\\_options](#).

### Reporting

---

`level(#)`; see [R] [estimation options](#).

`nocnsreport`; see [R] [estimation options](#).

`display_options`: `nocl`, `nopvalues`, `noomitted`, `vsquish`, `noemptycells`, `baselevels`, `allbaselevels`, `nofvlabel`, `fvwrap(#)`, `fvwrapon(style)`, `cformat(%fmt)`, `pformat(%fmt)`, `sformat(%fmt)`, and `nolstretch`; see [R] [estimation options](#).

### Maximization

---

`maximize_options`: `difficult`, `technique(algorithm_spec)`, `iterate(#)`, `[no]log`, `trace`, `gradient`, `showstep`, `hessian`, `showtolerance`, `tolerance(#)`, `ltolerance(#)`, `nrtolerance(#)`, `nonrtolerance`, and `from(init_specs)`; see [R] [maximize](#). These options are seldom used.

The following option is available with `xtfrontier` but is not shown in the dialog box: `coeflegend`; see [R] [estimation options](#).

## Remarks and examples

[stata.com](http://stata.com)

Remarks are presented under the following headings:

- [Introduction](#)
- [Time-invariant model](#)
- [Time-varying decay model](#)

## Introduction

Stochastic production frontier models were introduced by [Aigner, Lovell, and Schmidt \(1977\)](#) and [Meeusen and van den Broeck \(1977\)](#). Since then, stochastic frontier models have become a popular subfield in econometrics; see [Kumbhakar and Lovell \(2000\)](#) for an introduction. `xtfrontier` fits two stochastic frontier models with distinct specifications of the inefficiency term and can fit both production- and cost-frontier models.

Let's review the nature of the stochastic frontier problem. Suppose that a producer has a production function  $f(\mathbf{z}_{it}, \beta)$ . In a world without error or inefficiency, in time  $t$ , the  $i$ th firm would produce

$$q_{it} = f(\mathbf{z}_{it}, \beta)$$

A fundamental element of stochastic frontier analysis is that each firm potentially produces less than it might because of a degree of inefficiency. Specifically,

$$q_{it} = f(\mathbf{z}_{it}, \beta)\xi_{it}$$

where  $\xi_{it}$  is the level of efficiency for firm  $i$  at time  $t$ ;  $\xi_i$  must be in the interval  $(0, 1]$ . If  $\xi_{it} = 1$ , the firm is achieving the optimal output with the technology embodied in the production function  $f(\mathbf{z}_{it}, \beta)$ . When  $\xi_{it} < 1$ , the firm is not making the most of the inputs  $\mathbf{z}_{it}$  given the technology embodied in the production function  $f(\mathbf{z}_{it}, \beta)$ . Because the output is assumed to be strictly positive (that is,  $q_{it} > 0$ ), the degree of technical efficiency is assumed to be strictly positive (that is,  $\xi_{it} > 0$ ).

Output is also assumed to be subject to random shocks, implying that

$$q_{it} = f(\mathbf{z}_{it}, \beta)\xi_{it}\exp(v_{it})$$

Taking the natural log of both sides yields

$$\ln(q_{it}) = \ln\{f(\mathbf{z}_{it}, \beta)\} + \ln(\xi_{it}) + v_{it}$$

Assuming that there are  $k$  inputs and that the production function is linear in logs, defining  $u_{it} = -\ln(\xi_{it})$  yields

$$\ln(q_{it}) = \beta_0 + \sum_{j=1}^k \beta_j \ln(z_{jit}) + v_{it} - u_{it} \tag{1}$$

Because  $u_{it}$  is subtracted from  $\ln(q_{it})$ , restricting  $u_{it} \geq 0$  implies that  $0 < \xi_{it} \leq 1$ , as specified above.

Kumbhakar and Lovell (2000) provide a detailed version of this derivation, and they show that performing an analogous derivation in the dual cost function problem allows us to specify the problem as

$$\ln(c_{it}) = \beta_0 + \beta_q \ln(q_{it}) + \sum_{j=1}^k \beta_j \ln(p_{jit}) + v_{it} - su_{it} \quad (2)$$

where  $q_{it}$  is output, the  $z_{jit}$  are input quantities,  $c_{it}$  is cost, the  $p_{jit}$  are input prices, and

$$s = \begin{cases} 1, & \text{for production functions} \\ -1, & \text{for cost functions} \end{cases}$$

Intuitively, the inefficiency effect is required to lower output or raise expenditure, depending on the specification.

## □ Technical note

The model that `xtfrontier` actually fits has the form

$$y_{it} = \beta_0 + \sum_{j=1}^k \beta_j x_{jit} + v_{it} - su_{it}$$

so in the context of the discussion above,  $y_{it} = \ln(q_{it})$  and  $x_{jit} = \ln(z_{jit})$  for a production function; for a cost function,  $y_{it} = \ln(c_{it})$ , the  $x_{jit}$  are the  $\ln(p_{jit})$ , and  $\ln(q_{it})$ . You must perform the natural logarithm transformation of the data before estimation to interpret the estimation results correctly for a stochastic frontier production or cost model. `xtfrontier` does not perform any transformations on the data. □

As shown above, the disturbance term in a stochastic frontier model is assumed to have two components. One component is assumed to have a strictly nonnegative distribution, and the other component is assumed to have a symmetric distribution. In the econometrics literature, the nonnegative component is often referred to as the *inefficiency term*, and the component with the symmetric distribution as the *idiosyncratic error*.

Equation (2) is a variant of a panel-data model in which  $v_{it}$  is the idiosyncratic error and  $u_{it}$  is a time-varying panel-level effect. Much of the literature on this model has focused on deriving estimators for different specifications of the  $u_{it}$  term. Kumbhakar and Lovell (2000) provide a survey of this literature.

`xtfrontier` provides estimators for two different specifications of  $u_{it}$ . To facilitate the discussion, let  $N^+(\mu, \sigma^2)$  denote the truncated-normal distribution, which is truncated at zero with mean  $\mu$  and variance  $\sigma^2$ , and let  $\overset{\text{iid}}{\sim}$  stand for independent and identically distributed.

Consider the simplest specification in which the inefficiency term  $u_{it}$  is a time-invariant truncated-normal random variable. In the time-invariant model,  $u_{it} = u_i$ ,  $u_i \overset{\text{iid}}{\sim} N^+(\mu, \sigma_u^2)$ ,  $v_{it} \overset{\text{iid}}{\sim} N(0, \sigma_v^2)$ , and  $u_i$  and  $v_{it}$  are distributed independently of each other and the covariates in the model. Specifying the `ti` option causes `xtfrontier` to estimate the parameters of this model.

In the Battese–Coelli (1992) parameterization of time effects, the inefficiency term is modeled as a truncated-normal random variable multiplied by a specific function of time. In the time-varying decay specification,

$$u_{it} = \exp\{-\eta(t - T_i)\}u_i$$

where  $T_i$  is the last period in the  $i$ th panel,  $\eta$  is the decay parameter,  $u_i \stackrel{iid}{\sim} N^+(\mu, \sigma_u^2)$ ,  $v_{it} \stackrel{iid}{\sim} N(0, \sigma_v^2)$ , and  $u_i$  and  $v_{it}$  are distributed independently of each other and the covariates in the model. Specifying the tvd option causes xtfrontier to estimate the parameters of this model.

## Time-invariant model

### ► Example 1

xtfrontier, ti provides maximum likelihood estimates for the parameters of the time-invariant decay model. In this model, the inefficiency effects are modeled as  $u_{it} = u_i$ ,  $u_i \stackrel{iid}{\sim} N^+(\mu, \sigma_u^2)$ ,  $v_{it} \stackrel{iid}{\sim} N(0, \sigma_v^2)$ , and  $u_i$  and  $v_{it}$  are distributed independently of each other and the covariates in the model. In this example, firms produce a product called a widget, using a constant-returns-to-scale technology. We have 948 observations—91 firms, with 6–14 observations per firm. Our dataset contains variables representing the quantity of widgets produced, the number of machine hours used in production, the number of labor hours used in production, and three additional variables that are the natural logarithm transformations of the three aforementioned variables.

We fit a time-invariant model using the transformed variables:

```
. use http://www.stata-press.com/data/r15/xtfrontier1
. xtfrontier lnwidgets lnmachines lnworkers, ti
Iteration 0:  log likelihood = -1473.8703
Iteration 1:  log likelihood = -1473.0565
Iteration 2:  log likelihood = -1472.6155
Iteration 3:  log likelihood = -1472.607
Iteration 4:  log likelihood = -1472.6069

Time-invariant inefficiency model      Number of obs      =      948
Group variable: id                    Number of groups   =      91
Obs per group:
    min =      6
    avg =     10.4
    max =     14

Wald chi2(2)      =     661.76
Prob > chi2      =     0.0000

Log likelihood = -1472.6069
```

|            | Coef.    | Std. Err. | z     | P> z  | [95% Conf. Interval] |          |
|------------|----------|-----------|-------|-------|----------------------|----------|
| lnwidgets  |          |           |       |       |                      |          |
| lnmachines | .2904551 | .0164219  | 17.69 | 0.000 | .2582688             | .3226415 |
| lnworkers  | .2943333 | .0154352  | 19.07 | 0.000 | .2640808             | .3245858 |
| _cons      | 3.030983 | .1441022  | 21.03 | 0.000 | 2.748548             | 3.313418 |
| /mu        | 1.125667 | .6479217  | 1.74  | 0.082 | -.144236             | 2.39557  |
| /lnsigma2  | 1.421979 | .2672745  | 5.32  | 0.000 | .898131              | 1.945828 |
| /lgtgamma  | 1.138685 | .3562642  | 3.20  | 0.001 | .4404204             | 1.83695  |
| sigma2     | 4.145318 | 1.107938  |       |       | 2.455011             | 6.999424 |
| gamma      | .7574382 | .0654548  |       |       | .6083592             | .8625876 |
| sigma_u2   | 3.139822 | 1.107235  |       |       | .9696821             | 5.309962 |
| sigma_v2   | 1.005496 | .0484143  |       |       | .9106055             | 1.100386 |

In addition to the coefficients, the output reports estimates for the parameters sigma\_v2, sigma\_u2, gamma, sigma2, lgtgamma, lnsigma2, and mu. sigma\_v2 is the estimate of  $\sigma_v^2$ . sigma\_u2 is the estimate of  $\sigma_u^2$ . gamma is the estimate of  $\gamma = \sigma_u^2 / \sigma_S^2$ . sigma2 is the estimate of  $\sigma_S^2 = \sigma_v^2 + \sigma_u^2$ . Because  $\gamma$  must be between 0 and 1, the optimization is parameterized in terms of the logit of  $\gamma$ , and

this estimate is reported as `lgtgamma`. Because  $\sigma_S^2$  must be positive, the optimization is parameterized in terms of  $\ln(\sigma_S^2)$ , and this estimate is reported as `lnsigma2`. Finally, `mu` is the estimate of  $\mu$ . ◀

### □ Technical note

Our simulation results indicate that this estimator requires relatively large samples to achieve any reasonable degree of precision in the estimates of  $\mu$  and  $\sigma_u^2$ . □

## Time-varying decay model

`xtfrontier`, `tvd` provides maximum likelihood estimates for the parameters of the time-varying decay model. In this model, the inefficiency effects are modeled as

$$u_{it} = \exp\{-\eta(t - T_i)\}u_i$$

where  $u_i \stackrel{\text{iid}}{\sim} N^+(\mu, \sigma_u^2)$ .

When  $\eta > 0$ , the degree of inefficiency decreases over time; when  $\eta < 0$ , the degree of inefficiency increases over time. Because  $t = T_i$  in the last period, the last period for firm  $i$  contains the base level of inefficiency for that firm. If  $\eta > 0$ , the level of inefficiency decays toward the base level. If  $\eta < 0$ , the level of inefficiency increases to the base level.

### ▷ Example 2

When  $\eta = 0$ , the time-varying decay model reduces to the time-invariant model. The following example illustrates this property and demonstrates how to specify constraints and starting values in these models.

Let's begin by fitting the time-varying decay model on the same data that were used in the previous example for the time-invariant model.



```
. xtfrontier lnwidgets lnmachines lnworkers, tvd
Iteration 0:  log likelihood = -1551.3798 (not concave)
Iteration 1:  log likelihood = -1502.2637
Iteration 2:  log likelihood = -1476.3093 (not concave)
Iteration 3:  log likelihood = -1472.9845
Iteration 4:  log likelihood = -1472.5365
Iteration 5:  log likelihood = -1472.529
Iteration 6:  log likelihood = -1472.5289

Time-varying decay inefficiency model      Number of obs      =      948
Group variable: id                          Number of groups   =      91
Time variable: t                            Obs per group:
                                                min =              6
                                                avg =             10.4
                                                max =             14

                                                Wald chi2(2)      =      661.93
                                                Prob > chi2       =      0.0000

Log likelihood = -1472.5289
```

| lnwidgets  | Coef.    | Std. Err. | z     | P> z  | [95% Conf. Interval] |          |
|------------|----------|-----------|-------|-------|----------------------|----------|
| lnmachines | .2907555 | .0164376  | 17.69 | 0.000 | .2585384             | .3229725 |
| lnworkers  | .2942412 | .0154373  | 19.06 | 0.000 | .2639846             | .3244978 |
| _cons      | 3.028939 | .1436046  | 21.09 | 0.000 | 2.74748              | 3.310399 |
| /mu        | 1.110831 | .6452809  | 1.72  | 0.085 | -.1538967            | 2.375558 |
| /eta       | .0016764 | .00425    | 0.39  | 0.693 | -.0066535            | .0100064 |
| /lnsigma2  | 1.410723 | .2679485  | 5.26  | 0.000 | .885554              | 1.935893 |
| /lgtgamma  | 1.123982 | .3584243  | 3.14  | 0.002 | .4214828             | 1.82648  |
| sigma2     | 4.098919 | 1.098299  |       |       | 2.424327             | 6.930228 |
| gamma      | .7547265 | .0663495  |       |       | .603838              | .8613419 |
| sigma_u2   | 3.093563 | 1.097606  |       |       | .9422943             | 5.244832 |
| sigma_v2   | 1.005356 | .0484079  |       |       | .9104785             | 1.100234 |

The estimate of  $\eta$  is close to zero, and the other estimates are not too far from those of the time-invariant model.

We can use constraint to constrain  $\eta = 0$  and obtain the same results produced by the time-invariant model. Although there is only one statistical equation to be estimated in this model, the model fits five of Stata's [R] ml equations; see [R] ml or Gould, Pitblado, and Poi (2010). The equation names can be seen by listing the matrix of estimated coefficients.

```
. matrix list e(b)
e(b) [1,7]
      lnwidgets:  lnwidgets:  lnwidgets:  lnsigma2:  lgtgamma:      mu:
      lnmachines  lnworkers   _cons      _cons      _cons      _cons
y1   .29075546   .2942412   3.0289395  1.4107233  1.1239816  1.1108307
      eta:
      _cons
y1   .00167642
```

To constrain a parameter to a particular value in any equation, except the first equation, you must specify both the equation name and the parameter name by using the syntax

```
constraint # [eqname]_b[varname] = value      or
constraint # [eqname]coefficient = value
```

where *eqname* is the equation name, *varname* is the name of the variable in a linear equation, and *coefficient* refers to any parameter that has been estimated. More elaborate specifications with

expressions are possible; see the example with constant returns to scale below, and see [R] **constraint** for general reference.

Suppose that we impose the constraint  $\eta = 0$ ; we get the same results as those reported above for the time-invariant model, except for some minute differences attributable to an alternate convergence path in the optimization.

```
. constraint 1 [eta]_cons = 0
. xtfrontier lnwidgets lnmachines lnworkers, tvd constraints(1)
Iteration 0: log likelihood = -1540.7124 (not concave)
Iteration 1: log likelihood = -1515.7726
Iteration 2: log likelihood = -1473.0162
Iteration 3: log likelihood = -1472.9223
Iteration 4: log likelihood = -1472.6254
Iteration 5: log likelihood = -1472.607
Iteration 6: log likelihood = -1472.6069

Time-varying decay inefficiency model      Number of obs      =      948
Group variable: id                          Number of groups   =      91
Time variable: t                            Obs per group:
                                             min =              6
                                             avg =             10.4
                                             max =             14

                                             Wald chi2(2)      =      661.76
                                             Prob > chi2       =      0.0000

Log likelihood = -1472.6069
( 1) [eta]_cons = 0
```

| lnwidgets  | Coef.    | Std. Err. | z     | P> z  | [95% Conf. Interval] |          |
|------------|----------|-----------|-------|-------|----------------------|----------|
| lnmachines | .2904551 | .0164219  | 17.69 | 0.000 | .2582688             | .3226414 |
| lnworkers  | .2943332 | .0154352  | 19.07 | 0.000 | .2640807             | .3245857 |
| _cons      | 3.030963 | .1440995  | 21.03 | 0.000 | 2.748534             | 3.313393 |
| /mu        | 1.125507 | .6480444  | 1.74  | 0.082 | -.1446369            | 2.39565  |
| /eta       | 0        | (omitted) |       |       |                      |          |
| /lnsigma2  | 1.422039 | .2673128  | 5.32  | 0.000 | .8981155             | 1.945962 |
| /lgtgamma  | 1.138764 | .3563076  | 3.20  | 0.001 | .4404135             | 1.837114 |
| sigma2     | 4.145565 | 1.108162  |       |       | 2.454972             | 7.000366 |
| gamma      | .7574526 | .0654602  |       |       | .6083575             | .862607  |
| sigma_u2   | 3.140068 | 1.107459  |       |       | .9694878             | 5.310649 |
| sigma_v2   | 1.005496 | .0484143  |       |       | .9106057             | 1.100386 |

## Stored results

xtfrontier stores the following in `e()`:

### Scalars

|                            |  |
|----------------------------|--|
| <code>e(N)</code>          | number of observations                       |
| <code>e(N_g)</code>        | number of groups                             |
| <code>e(k)</code>          | number of parameters                         |
| <code>e(k_eq)</code>       | number of equations in <code>e(b)</code>     |
| <code>e(k_eq_model)</code> | number of equations in overall model test    |
| <code>e(k_dv)</code>       | number of dependent variables                |
| <code>e(df_m)</code>       | model degrees of freedom                     |
| <code>e(ll)</code>         | log likelihood                               |
| <code>e(g_min)</code>      | minimum number of observations per group     |
| <code>e(g_avg)</code>      | average number of observations per group     |
| <code>e(g_max)</code>      | maximum number of observations per group     |
| <code>e(sigma2)</code>     | sigma2                                       |
| <code>e(gamma)</code>      | gamma  |
| <code>e(Tcon)</code>       | 1 if panels balanced, 0 otherwise            |
| <code>e(sigma_u)</code>    | standard deviation of technical inefficiency |
| <code>e(sigma_v)</code>    | standard deviation of random error           |
| <code>e(chi2)</code>       | $\chi^2$                                     |
| <code>e(p)</code>          | <i>p</i> -value for model test               |
| <code>e(rank)</code>       | rank of <code>e(V)</code>                    |
| <code>e(ic)</code>         | number of iterations                         |
| <code>e(rc)</code>         | return code                                  |
| <code>e(converged)</code>  | 1 if converged, 0 otherwise                  |

### Macros

|                            |  |
|----------------------------|--|
| <code>e(cmd)</code>        | xtfrontier   |
| <code>e(cmdline)</code>    | command as typed   |
| <code>e(depvar)</code>     | name of dependent variable   |
| <code>e(ivar)</code>       | variable denoting groups   |
| <code>e(tvar)</code>       | variable denoting time within groups                                     |
| <code>e(function)</code>   | production or cost   |
| <code>e(model)</code>      | ti, after time-invariant model; tvd, after time-varying decay model      |
| <code>e(wtype)</code>      | weight type  |
| <code>e(wexp)</code>       | weight expression  |
| <code>e(title)</code>      | title in estimation output   |
| <code>e(chi2type)</code>   | Wald; type of model $\chi^2$ test  |
| <code>e(vce)</code>        | <i>vctype</i> specified in <code>vce()</code>                            |
| <code>e(opt)</code>        | type of optimization   |
| <code>e(which)</code>      | max or min; whether optimizer is to perform maximization or minimization |
| <code>e(ml_method)</code>  | type of ml method  |
| <code>e(user)</code>       | name of likelihood-evaluator program                                     |
| <code>e(technique)</code>  | maximization technique   |
| <code>e(properties)</code> | b V  |
| <code>e(predict)</code>    | program used to implement predict  |
| <code>e(asbalanced)</code> | factor variables <code>fvset</code> as <code>asbalanced</code>           |
| <code>e(asobserved)</code> | factor variables <code>fvset</code> as <code>asobserved</code>           |

### Matrices

|                          |  |
|--------------------------|--|
| <code>e(b)</code>        | coefficient vector                           |
| <code>e(Cns)</code>      | constraints matrix                           |
| <code>e(ilog)</code>     | iteration log (up to 20 iterations)          |
| <code>e(gradient)</code> | gradient vector                              |
| <code>e(V)</code>        | variance-covariance matrix of the estimators |

### Functions

|                        |                         |
|------------------------|-------------------------|
| <code>e(sample)</code> | marks estimation sample |
|------------------------|-------------------------|

## Methods and formulas

`xtfrontier` fits stochastic frontier models for panel data that can be expressed as

$$y_{it} = \beta_0 + \sum_{j=1}^k \beta_j x_{jit} + v_{it} - s u_{it}$$

where  $y_{it}$  is the natural logarithm of output, the  $x_{jit}$  are the natural logarithm of the input quantities for the production efficiency problem,  $y_{it}$  is the natural logarithm of costs, the  $x_{it}$  are the natural logarithm of input prices for the cost efficiency problem, and

$$s = \begin{cases} 1, & \text{for production functions} \\ -1, & \text{for cost functions} \end{cases}$$

For the time-varying decay model, the log-likelihood function is derived as

$$\begin{aligned} \ln L = & -\frac{1}{2} \left( \sum_{i=1}^N T_i \right) \{ \ln(2\pi) + \ln(\sigma_S^2) \} - \frac{1}{2} \sum_{i=1}^N (T_i - 1) \ln(1 - \gamma) \\ & - \frac{1}{2} \sum_{i=1}^N \ln \left\{ 1 + \left( \sum_{t=1}^{T_i} \eta_{it}^2 - 1 \right) \gamma \right\} - N \ln \{ 1 - \Phi(-\tilde{z}) \} - \frac{1}{2} N \tilde{z}^2 \\ & + \sum_{i=1}^N \ln \{ 1 - \Phi(-z_i^*) \} + \frac{1}{2} \sum_{i=1}^N z_i^{*2} - \frac{1}{2} \sum_{i=1}^N \sum_{t=1}^{T_i} \frac{\epsilon_{it}^2}{(1 - \gamma) \sigma_S^2} \end{aligned}$$

where  $\sigma_S = (\sigma_u^2 + \sigma_v^2)^{1/2}$ ,  $\gamma = \sigma_u^2 / \sigma_S^2$ ,  $\epsilon_{it} = y_{it} - \mathbf{x}_{it}\boldsymbol{\beta}$ ,  $\eta_{it} = \exp\{-\eta(t - T_i)\}$ ,  $\tilde{z} = \mu / (\gamma \sigma_S^2)^{1/2}$ ,  $\Phi()$  is the cumulative distribution function of the standard normal distribution, and

$$z_i^* = \frac{\mu(1 - \gamma) - s\gamma \sum_{t=1}^{T_i} \eta_{it} \epsilon_{it}}{\left[ \gamma(1 - \gamma) \sigma_S^2 \left\{ 1 + \left( \sum_{t=1}^{T_i} \eta_{it}^2 - 1 \right) \gamma \right\} \right]^{1/2}}$$

Maximizing the above log likelihood estimates the coefficients  $\eta$ ,  $\mu$ ,  $\sigma_v$ , and  $\sigma_u$ .

## References

- Aigner, D. J., C. A. K. Lovell, and P. Schmidt. 1977. Formulation and estimation of stochastic frontier production function models. *Journal of Econometrics* 6: 21–37.
- Battese, G. E., and T. J. Coelli. 1992. Frontier production functions, technical efficiency and panel data: With application to paddy farmers in India. *Journal of Productivity Analysis* 3: 153–169.
- . 1995. A model for technical inefficiency effects in a stochastic frontier production function for panel data. *Empirical Economics* 20: 325–332.
- Belotti, F., S. Daidone, G. Ilardi, and V. Atella. 2013. Stochastic frontier analysis using Stata. *Stata Journal* 13: 719–758.
- Beaudill, S. B., J. M. Ford, and D. M. Gropper. 1995. Frontier estimation and firm-specific inefficiency measures in the presence of heteroscedasticity. *Journal of Business and Economic Statistics* 13: 105–111.

- Coelli, T. J. 1995. Estimators and hypothesis tests for a stochastic frontier function: A Monte Carlo analysis. *Journal of Productivity Analysis* 6: 247–268.
- Coelli, T. J., D. S. P. Rao, C. J. O’Donnell, and G. E. Battese. 2005. *An Introduction to Efficiency and Productivity Analysis*. 2nd ed. New York: Springer.
- Gould, W. W., J. S. Pitblado, and B. P. Poi. 2010. *Maximum Likelihood Estimation with Stata*. 4th ed. College Station, TX: Stata Press.
- Karakaplan, M. U. 2017. Fitting endogenous stochastic frontier models in Stata. *Stata Journal* 17: 39–55.
- Kumbhakar, S. C., and C. A. K. Lovell. 2000. *Stochastic Frontier Analysis*. Cambridge: Cambridge University Press.
- Kumbhakar, S. C., H.-J. Wang, and A. P. Horncastle. 2015. *A Practitioner’s Guide to Stochastic Frontier Analysis Using Stata*. New York: Cambridge University Press.
- Meeusen, W., and J. van den Broeck. 1977. Efficiency estimation from Cobb–Douglas production functions with composed error. *International Economic Review* 18: 435–444.
- Zellner, A., and N. S. Revankar. 1969. Generalized production functions. *Review of Economic Studies* 36: 241–250.

## Also see

- [XT] **xtfrontier postestimation** — Postestimation tools for xtfrontier
- [XT] **xtset** — Declare data to be panel data
- [R] **frontier** — Stochastic frontier models
- [U] **20 Estimation and postestimation commands**