xtfrontier — Stochastic frontier models for panel data

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*]

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

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

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

by, collect, 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.

collinear and coeflegend do 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*); 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: noci, 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 options are available with xtfrontier but are not shown in the dialog box:

collinear, 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, xtfrontier fits a production frontier model.

constraints(*constraints*); 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: noci, 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 options are available with `xtfrontier` but are not shown in the dialog box:

`collinear`, `coeflegend`; see [R] **Estimation options**.

Remarks and examples

Remarks are presented under the following headings:

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 $f(\mathbf{z}_{it}, \boldsymbol{\beta})$. In a world without error or inefficiency, in time t, the ith firm would produce

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

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

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

where ξ_{it} is the level of efficiency for firm i at time t; ξ_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 $f(\mathbf{z}_{it}, \boldsymbol{\beta})$. 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 $f(\mathbf{z}_{it}, \boldsymbol{\beta})$. 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

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

Taking the natural log of both sides yields

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

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

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

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 μ and variance σ^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 ith panel, η is the decay parameter, $u_i \overset{\text{iid}}{\sim} N^+(\mu, \sigma_u^2)$, $v_{it} \overset{\text{iid}}{\sim} N(0, \sigma_v^2)$, and u_i and v_{it} are distributed independently of each other and the covariates in the model. Specifying the `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 $u_{it} = u_i$, $u_i \overset{\text{iid}}{\sim} N^+(\mu, \sigma_u^2)$, $v_{it} \overset{\text{iid}}{\sim} N(0, \sigma_v^2)$, and u_i and v_{it} are distributed independently of each other and the covariates in the model. 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 https://www.stata-press.com/data/r19/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
Log likelihood = -1472.6069                          Prob > chi2      = 0.0000
```

lnwidgets	Coefficient	Std. err.	z	P>|z|	[95% conf.	interval]
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 σ_v^2. `sigma_u2` is the estimate of σ_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 γ must be between 0 and 1, the optimization is parameterized in terms of the logit of γ, and this estimate is reported as `lgtgamma`. Because σ_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 μ.

◁

❑ Technical note

Our simulation results indicate that this estimator requires relatively large samples to achieve any reasonable degree of precision in the estimates of μ and σ_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 $u_i \overset{\text{iid}}{\sim} N^+(\mu, \sigma_u^2)$.

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
Log likelihood = -1472.5289                     Prob > chi2      = 0.0000
```

lnwidgets	Coefficient	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 η 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 Pitblado, Poi, and Gould (2024). 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
Log likelihood = -1472.6069                     Prob > chi2       = 0.0000
 ( 1)  [eta]_cons = 0
```

lnwidgets	Coefficient	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	
e(N)	number of observations
e(N_g)	number of groups
e(k)	number of parameters
e(k_eq)	number of equations in e(b)
e(k_eq_model)	number of equations in overall model test
e(k_dv)	number of dependent variables
e(df_m)	model degrees of freedom
e(ll)	log likelihood
e(g_min)	minimum number of observations per group
e(g_avg)	average number of observations per group
e(g_max)	maximum number of observations per group
e(sigma2)	sigma2
e(gamma)	gamma
e(Tcon)	1 if panels balanced, 0 otherwise
e(sigma_u)	standard deviation of technical inefficiency
e(sigma_v)	standard deviation of random error
e(chi2)	χ^2
e(p)	*p*-value for model test
e(rank)	rank of e(V)
e(ic)	number of iterations
e(rc)	return code
e(converged)	1 if converged, 0 otherwise
Macros	
e(cmd)	xtfrontier
e(cmdline)	command as typed
e(depvar)	name of dependent variable
e(ivar)	variable denoting groups
e(tvar)	variable denoting time within groups
e(function)	production or cost
e(model)	ti, after time-invariant model; tvd, after time-varying decay model
e(wtype)	weight type
e(wexp)	weight expression
e(title)	title in estimation output
e(chi2type)	Wald; type of model χ^2 test
e(vce)	*vcetype* specified in vce()
e(opt)	type of optimization
e(which)	max or min; whether optimizer is to perform maximization or minimization
e(ml_method)	type of ml method
e(user)	name of likelihood-evaluator program
e(technique)	maximization technique
e(properties)	b V
e(predict)	program used to implement predict
e(asbalanced)	factor variables fvset as asbalanced
e(asobserved)	factor variables fvset as asobserved
Matrices	
e(b)	coefficient vector
e(Cns)	constraints matrix
e(ilog)	iteration log (up to 20 iterations)
e(gradient)	gradient vector
e(V)	variance–covariance matrix of the estimators
Functions	
e(sample)	marks estimation sample

In addition to the above, the following is stored in `r()`:

Matrices
`r(table)` matrix containing the coefficients with their standard errors, test statistics, p-values, and confidence intervals

Note that results stored in `r()` are updated when the command is replayed and will be replaced when any r-class command is run after the estimation command.

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 $\sigma_S = (\sigma_u^2 + \sigma_v^2)^{1/2}$, $\gamma = \sigma_u^2/\sigma_S^2$, $\epsilon_{it} = y_{it} - \mathbf{x}_{it}\boldsymbol{\beta}$, $\eta_{it} = \exp\{-\eta(t - T_i)\}$, $\tilde{z} = \mu/\left(\gamma\sigma_S^2\right)^{1/2}$, $\Phi()$ 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 η, μ, σ_v, and σ_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. https://doi.org/10.1016/0304-4076(77)90052-5.

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. https://doi.org/10.1007/BF00158774.

———. 1995. A model for technical inefficiency effects in a stochastic frontier production function for panel data. *Empirical Economics* 20: 325–332. https://doi.org/10.1007/BF01205442.

Belotti, F., S. Daidone, G. Ilardi, and V. Atella. 2013. Stochastic frontier analysis using Stata. *Stata Journal* 13: 719–758.

Caudill, 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. https://doi.org/10.2307/1392525.

Cococcioni, M., M. Grazzi, L. Li, and F. Ponchio. 2022. A toolbox for measuring heterogeneity and efficiency using zonotopes. *Stata Journal* 22: 25–59.

Coelli, T. J. 1995. Estimators and hypothesis tests for a stochastic frontier function: A Monte Carlo analysis. *Journal of Productivity Analysis* 6: 247–268. https://doi.org/10.1007/BF01076978.

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. https://doi.org/10.1007/b136381.

Fé, E., and R. Hofler. 2020. sfcount: Command for count-data stochastic frontiers and underreported and overreported counts. *Stata Journal* 20: 532–547.

Karakaplan, M. U. 2017. Fitting endogenous stochastic frontier models in Stata. *Stata Journal* 17: 39–55.

———. 2022. Panel stochastic frontier models with endogeneity. *Stata Journal* 22: 643–663.

Kumbhakar, S. C., and C. A. K. Lovell. 2000. *Stochastic Frontier Analysis*. Cambridge: Cambridge University Press. https://doi.org/10.1017/CBO9781139174411.

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. https://doi.org/10.2307/2525757.

Pitblado, J. S., B. P. Poi, and W. W. Gould. 2024. *Maximum Likelihood Estimation with Stata*. 5th ed. College Station, TX: Stata Press.

Rovigatti, G., and V. Mollisi. 2018. Theory and practice of total-factor productivity estimation: The control function approach using Stata. *Stata Journal* 18: 618–662.

Wang, D., K. Du, and N. Zhang. 2022. Measuring technical efficiency and total factor productivity change with undesirable outputs in Stata. *Stata Journal* 22: 103–124.

Zellner, A., and N. S. Revankar. 1969. Generalized production functions. *Review of Economic Studies* 36: 241–250. https://doi.org/10.2307/2296840.

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