Title

> **frontier** — Stochastic frontier models

Description	Quick start	Menu	Syntax
Options	Remarks and examples	Stored results	Methods and formulas
References	Also see		

Description

frontier fits stochastic production or cost frontier models; the default is a production frontier model. It provides estimators for the parameters of a linear model with a disturbance that is assumed to be a mixture of two components, which have a strictly nonnegative and symmetric distribution, respectively. frontier can fit models in which the nonnegative distribution component (a measurement of inefficiency) is assumed to be from a half-normal, exponential, or truncated-normal distribution. See Kumbhakar and Lovell ([2000](2000)) for a detailed introduction to frontier analysis.

Quick start

Cobb–Douglas production frontier model of lny1 as a function of lnx1 and lnx2
        frontier lny1 lnx1 lnx2

Same as above, but use exponential instead of half-normal distribution for the inefficiency term
        frontier lny1 lnx1 lnx2, distribution(exponential)

Include x3 as an explanatory variable in the idiosyncratic error variance function
        frontier lny1 lnx1 lnx2, vhet(x3)

Same as above, and include x4 as an explanatory variable in the technical inefficiency variance function
        frontier lny1 lnx1 lnx2, vhet(x3) uhet(x4)

Conditional mean model with the mean modeled as a linear function of x3
        frontier lny1 lnx1 lnx2, distribution(tnormal) cm(x3)

Cost frontier model of y2 as a function of lnx1 and lnx2
        frontier y2 lnx1 lnx2, distribution(tnormal) cost

Menu

Statistics > Linear models and related > Frontier models

Syntax

> frontier *depvar* [*indepvars*] [*if*] [*in*] [*weight*] [, *options*]

options	Description
Model	
<u>noconst</u>ant	suppress constant term
distribution(hnormal)	half-normal distribution for the inefficiency term
distribution(exponential)	exponential distribution for the inefficiency term
distribution(tnormal)	truncated-normal distribution for the inefficiency term
ufrom(*matrix*)	specify untransformed log likelihood; only with d(tnormal)
cm(*varlist* [, <u>noconst</u>ant])	fit conditional mean model; only with d(tnormal); use noconstant to suppress constant term
Model 2	
constraints(*constraints*)	apply specified linear constraints
<u>u</u>het(*varlist* [, <u>noconst</u>ant])	explanatory variables for technical inefficiency variance function; use noconstant to suppress constant term
<u>v</u>het(*varlist* [, <u>noconst</u>ant])	explanatory variables for idiosyncratic error variance function; use noconstant to suppress constant term
cost	fit cost frontier model; default is production frontier model
SE/Robust	
* vce(*vcetype*)	*vcetype* may be oim, <u>r</u>obust, <u>cl</u>uster *clustvar*, opg, <u>boot</u>strap, or jackknife
Reporting	
<u>l</u>evel(*#*)	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

---

* vce(robust) and vce(cluster *clustvar*) may not be specified with distribution(tnormal).

indepvars and *varlist* may contain factor variables; see [U] **11.4.3 Factor variables**.

bootstrap, by, collect, fp, jackknife, rolling, and statsby are allowed; see [U] **11.1.10 Prefix commands**.

Weights are not allowed with the bootstrap prefix; see [R] **bootstrap**.

fweights, iweights, and pweights are allowed; see [U] **11.1.6 weight**.

collinear and coeflegend do not appear in the dialog box.

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

Options

> Model

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

distribution(*distname*) specifies the distribution for the inefficiency term as half-normal (hnormal), exponential, or truncated-normal (tnormal). The default is hnormal.

ufrom(*matrix*) specifies a $1 \times K$ matrix of untransformed starting values when the distribution is truncated-normal (tnormal). frontier can estimate the parameters of the model by maximizing either the log likelihood or a transformed log likelihood (see *Methods and formulas*). frontier automatically transforms the starting values before passing them on to the transformed log likelihood. The matrix must have the same number of columns as there are parameters to estimate.

cm(*varlist* [, noconstant]) may be used only with distribution(tnormal). Here frontier will fit a conditional mean model in which the mean of the truncated-normal distribution is modeled as a linear function of the set of covariates specified in *varlist*. Specifying noconstant suppresses the constant in the mean function.

> Model 2

constraints(*constraints*); see [R] **Estimation options**.

> By default, when fitting the truncated-normal model or the conditional mean model, frontier maximizes a transformed log likelihood. When constraints are applied, frontier will maximize the untransformed log likelihood with constraints defined in the untransformed metric.

uhet(*varlist* [, noconstant]) specifies that the technical inefficiency component is heteroskedastic, with the variance function depending on a linear combination of $varlist_u$. Specifying noconstant suppresses the constant term from the variance function. This option may not be specified with distribution(tnormal).

vhet(*varlist* [, noconstant]) specifies that the idiosyncratic error component is heteroskedastic, with the variance function depending on a linear combination of $varlist_v$. Specifying noconstant suppresses the constant term from the variance function. This option may not be specified with distribution(tnormal).

cost specifies that frontier fit a cost frontier model.

> SE/Robust

vce(*vcetype*) specifies the type of standard error reported, which includes types that are derived from asymptotic theory (oim, opg), 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 [R] *vce_option*.

> vce(robust) and vce(cluster *clustvar*) may not be specified with distribution(tnormal).

> Reporting

level(*#*), 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: <u>difficult</u>, <u>technique</u>(*algorithm_spec*), <u>iterate</u>(#), [no]<u>log</u>, <u>trace</u>,
  <u>gradient</u>, <u>showstep</u>, <u>hessian</u>, <u>showtolerance</u>, <u>tolerance</u>(#), <u>ltolerance</u>(#),
  <u>nrtolerance</u>(#), <u>nonrtolerance</u>, and <u>from</u>(*init_specs*); see [R] **Maximize**. These options are
  seldom used.

  Setting the optimization type to technique(bhhh) resets the default *vcetype* to vce(opg).

The following options are available with frontier but are not shown in the dialog box:

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

Remarks and examples

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. Kumbhakar and Lovell (2000) provide a good introduction.

frontier fits three stochastic frontier models with distinct parameterizations of the inefficiency
term and can fit stochastic production or cost frontier models.

Let's review the nature of the stochastic frontier problem. Suppose that a producer has a production
function $f(\mathbf{z}_i, \boldsymbol{\beta})$. In a world without error or inefficiency, the ith firm would produce

$$q_i = f(\mathbf{z}_i, \boldsymbol{\beta})$$

Stochastic frontier analysis assumes that each firm potentially produces less than it might due to
a degree of inefficiency. Specifically,

$$q_i = f(\mathbf{z}_i, \boldsymbol{\beta})\xi_i$$

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

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

$$q_i = f(\mathbf{z}_i, \boldsymbol{\beta})\xi_i \exp(v_i)$$

Taking the natural log of both sides yields

$$\ln(q_i) = \ln\{f(\mathbf{z}_i, \boldsymbol{\beta})\} + \ln(\xi_i) + v_i$$

Assuming that there are k inputs and that the production function is linear in logs, defining
$u_i = -\ln(\xi_i)$ yields

$$\ln(q_i) = \beta_0 + \sum_{j=1}^{k} \beta_j \ln(z_{ji}) + v_i - u_i \tag{1}$$

Because u_i is subtracted from $\ln(q_i)$, restricting $u_i \geq 0$ implies that $0 < \xi_i \leq 1$, as specified above.

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

$$\ln(c_i) = \beta_0 + \beta_q \ln(q_i) + \sum_{j=1}^{k} \beta_j \ln(p_{ji}) + v_i + u_i \tag{2}$$

where q_i is output, z_{ji} are input quantities, c_i is cost, and the p_{ji} are input prices.

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

❏ Technical note

The model that `frontier` actually fits is of the form

$$y_i = \beta_0 + \sum_{j=1}^{k} \beta_j x_{ji} + v_i - s u_i$$

where

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

so, in the context of the discussion above, $y_i = \ln(q_i)$, and $x_{ji} = \ln(z_{ji})$ for a production function; and for a cost function, $y_i = \ln(c_i)$, and the x_{ji} are the $\ln(p_{ji})$ and $\ln(q_i)$. You must take the natural logarithm of the data before fitting a stochastic frontier production or cost model. `frontier` performs no transformations on the data.

❏

Different specifications of the u_i and the v_i terms give rise to distinct models. `frontier` provides estimators for the parameters of three basic models in which the idiosyncratic component, v_i, is assumed to be independently $N(0, \sigma_v)$ distributed over the observations. The basic models differ in their specification of the inefficiency term, u_i, as follows:

> `exponential`: the u_i are independently exponentially distributed with variance σ_u^2
>
> `hnormal`: the u_i are independently half-normally $N^+(0, \sigma_u^2)$ distributed
>
> `tnormal`: the u_i are independently $N^+(\mu, \sigma_u^2)$ distributed with truncation point at 0

For half-normal or exponential distributions, `frontier` can fit models with heteroskedastic error components, conditional on a set of covariates. For a truncated-normal distribution, `frontier` can also fit a conditional mean model in which the mean is modeled as a linear function of a set of covariates.

▷ Example 1: The half-normal and the exponential models

For our first example, we demonstrate the half-normal and exponential models by reproducing a study found in Greene ([2003](#), 505), which uses data originally published in Zellner and Revankar ([1969](#)). In this study of the transportation equipment manufacturing industry, observations on value added, capital, and labor are used to estimate a Cobb–Douglas production function. The variable `lnv` is the log-transformed value added, `lnk` is the log-transformed capital, and `lnl` is the log-transformed labor. OLS estimates are compared with those from stochastic frontier models using both the half-normal and exponential distribution for the inefficiency term.

```
. use https://www.stata-press.com/data/r18/greene9
(Transportation equipment manufacturing industry)
. regress lnv lnk lnl
```

Source	SS	df	MS		Number of obs	=	25
					F(2, 22)	=	397.54
Model	44.1727741	2	22.086387		Prob > F	=	0.0000
Residual	1.22225984	22	.055557265		R-squared	=	0.9731
					Adj R-squared	=	0.9706
Total	45.3950339	24	1.89145975		Root MSE	=	.23571

lnv	Coefficient	Std. err.	t	P>|t|	[95% conf. interval]	
lnk	.2454281	.1068574	2.30	0.032	.0238193	.4670368
lnl	.805183	.1263336	6.37	0.000	.5431831	1.067183
_cons	1.844416	.2335928	7.90	0.000	1.359974	2.328858

```
. frontier lnv lnk lnl
Iteration 0:  Log likelihood =  2.3357572
Iteration 1:  Log likelihood =  2.4673009
Iteration 2:  Log likelihood =  2.4695125
Iteration 3:  Log likelihood =  2.4695222
Iteration 4:  Log likelihood =  2.4695222
```

Stoc. frontier normal/half-normal model

Log likelihood = 2.4695222

Number of obs = 25
Wald chi2(2) = 743.71
Prob > chi2 = 0.0000

lnv	Coefficient	Std. err.	z	P>|z|	[95% conf. interval]	
lnk	.2585478	.098764	2.62	0.009	.0649738	.4521218
lnl	.7802451	.1199399	6.51	0.000	.5451672	1.015323
_cons	2.081135	.281641	7.39	0.000	1.529128	2.633141
/lnsig2v	−3.48401	.6195353	−5.62	0.000	−4.698277	−2.269743
/lnsig2u	−3.014599	1.11694	−2.70	0.007	−5.203761	−.8254368
sigma_v	.1751688	.0542616			.0954514	.3214633
sigma_u	.2215073	.1237052			.074134	.6618486
sigma2	.0797496	.0426989			−.0039388	.163438
lambda	1.264536	.1678684			.9355204	1.593552

LR test of sigma_u=0: chibar2(01) = 0.43   Prob >= chibar2 = 0.256

```
. predict double u_h, u
```

```
. frontier lnv lnk lnl, distribution(exponential)

Iteration 0:   Log likelihood =  2.7270659
Iteration 1:   Log likelihood =  2.8551532
Iteration 2:   Log likelihood =  2.8604815
Iteration 3:   Log likelihood =  2.8604897
Iteration 4:   Log likelihood =  2.8604897
```

Stoc. frontier normal/exponential model
Log likelihood = 2.8604897

Number of obs =   25
Wald chi2(2)   = 845.68
Prob > chi2   = 0.0000

lnv	Coefficient	Std. err.	z	P>|z|	[95% conf. interval]	
lnk	.2624859	.0919988	2.85	0.004	.0821717	.4428002
lnl	.7703795	.1109569	6.94	0.000	.5529079	.9878511
_cons	2.069242	.2356159	8.78	0.000	1.607444	2.531041
/lnsig2v	-3.527598	.4486176	-7.86	0.000	-4.406873	-2.648324
/lnsig2u	-4.002457	.9274575	-4.32	0.000	-5.820241	-2.184674
sigma_v	.1713925	.0384448			.1104231	.2660258
sigma_u	.1351691	.0626818			.0544692	.3354317
sigma2	.0476461	.0157921			.016694	.0785981
lambda	.7886525	.087684			.616795	.9605101

LR test of sigma_u=0: chibar2(01) = 1.21   Prob >= chibar2 = 0.135

```
. predict double u_e, u

. list state u_h u_e
```

	state	u_h	u_e
1.	Alabama	.2011338	.14592865
2.	California	.14480966	.0972165
3.	Connecticut	.1903485	.13478797
4.	Florida	.51753139	.5903303
5.	Georgia	.10397912	.07140994
6.	Illinois	.12126696	.0830415
7.	Indiana	.21128212	.15450664
8.	Iowa	.24933153	.20073081
9.	Kansas	.10099517	.06857629
10.	Kentucky	.05626919	.04152443
11.	Louisiana	.20332731	.15066405
12.	Maine	.22263164	.17245793
13.	Maryland	.13534062	.09245501
14.	Massachusetts	.15636999	.10932923
15.	Michigan	.15809566	.10756915
16.	Missouri	.10288047	.0704146
17.	NewJersey	.09584337	.06587986
18.	NewYork	.27787793	.22249416
19.	Ohio	.22914231	.16981857
20.	Pennsylvania	.1500667	.10302905
21.	Texas	.20297875	.14552218
22.	Virginia	.14000132	.09676078
23.	Washington	.11047581	.07533251
24.	WestVirginia	.15561392	.11236153
25.	Wisconsin	.14067066	.0970861

The parameter estimates and the estimates of the inefficiency terms closely match those published in Greene (2003, 505), but the standard errors of the parameter estimates are estimated differently (see the technical note below).

The output from frontier includes estimates of the standard deviations of the two error components, σ_v and σ_u, which are labeled sigma_v and sigma_u, respectively. In the log likelihood, they are parameterized as $\ln\sigma_v^2$ and $\ln\sigma_u^2$, and these estimates are labeled /lnsig2v and /lnsig2u in the output. frontier also reports two other useful parameterizations. The estimate of the total error variance, $\sigma_S^2 = \sigma_v^2 + \sigma_u^2$, is labeled sigma2, and the estimate of the ratio of the standard deviation of the inefficiency component to the standard deviation of the idiosyncratic component, $\lambda = \sigma_u/\sigma_v$, is labeled lambda.

At the bottom of the output, frontier reports the results of a test that there is no technical inefficiency component in the model. This is a test of the null hypothesis $H_0 : \sigma_u^2 = 0$ against the alternative hypotheses $H_1 : \sigma_u^2 > 0$. If the null hypothesis is true, the stochastic frontier model reduces to an OLS model with normal errors. However, because the test lies on the boundary of the parameter space of σ_u^2, the standard likelihood-ratio test is not valid, and a one-sided generalized likelihood-ratio test must be constructed; see Gutierrez, Carter, and Drukker (2001). For this example, the output shows LR = 0.43 with a p-value of 0.256 for the half-normal model and LR = 1.21 with a p-value of 0.135 for the exponential model. There are several possible reasons for the failure to reject the null hypothesis, but the fact that the test is based on an asymptotic distribution and the sample size was 25 is certainly a leading candidate among those possibilities.

◁

❑ Technical note

frontier maximizes the log-likelihood function of a stochastic frontier model by using the Newton–Raphson method, and the estimated variance–covariance matrix is calculated as the inverse of the negative Hessian (matrix of second partial derivatives); see [R] **ml**. When comparing the results with those published using other software, be aware of the difference in the optimization methods, which may result in different, yet asymptotically equivalent, variance estimates.

❑

▷ Example 2: Models with heteroskedasticity

Often the error terms may not have constant variance. frontier allows you to model heteroskedasticity in either error term as a linear function of a set of covariates. The variance of either the technical inefficiency or the idiosyncratic component may be modeled as

$$\sigma_i^2 = \exp(\mathbf{w}_i\boldsymbol{\delta})$$

The default constant included in $\mathbf{w}_i$ may be suppressed by appending a noconstant option to the list of covariates. Also, you can simultaneously specify covariates for both σ_{u_i} and σ_{v_i}.

In this example, we use a sample of 756 observations of fictional firms producing a manufactured good by using capital and labor. The firms are hypothesized to use a constant returns-to-scale technology, but the sizes of the firms differ. Believing that this size variation will introduce heteroskedasticity into the idiosyncratic error term, we estimate the parameters of a Cobb–Douglas production function. To do this, we use a conditional heteroskedastic half-normal model, with the size of the firm as an explanatory variable in the variance function for the idiosyncratic error. We also perform a test of the hypothesis that the firms use a constant returns-to-scale technology.

```
. use https://www.stata-press.com/data/r18/frontier1, clear

. frontier lnoutput lnlabor lncapital, vhet(size)

Iteration 0:  Log likelihood = -1508.3692
Iteration 1:  Log likelihood =  -1501.583
Iteration 2:  Log likelihood = -1500.3942
Iteration 3:  Log likelihood = -1500.3794
Iteration 4:  Log likelihood = -1500.3794
```

Stoc. frontier normal/half-normal model   Number of obs =   756
                                                      Wald chi2(2) =   9.68
Log likelihood = -1500.3794   Prob > chi2   = 0.0079

lnoutput	Coefficient	Std. err.	z	P>|z|	[95% conf. interval]	
lnoutput						
lnlabor	.7090933	.2349374	3.02	0.003	.2486244	1.169562
lncapital	.3931345	.5422173	0.73	0.468	-.6695919	1.455861
_cons	1.252199	3.14656	0.40	0.691	-4.914946	7.419344
lnsig2v						
size	-.0016951	.0004748	-3.57	0.000	-.0026256	-.0007645
_cons	3.156091	.9265826	3.41	0.001	1.340023	4.97216
lnsig2u						
_cons	1.947487	.1017653	19.14	0.000	1.748031	2.146943
sigma_u	2.647838	.134729			2.396514	2.925518

```
. test _b[lnlabor] + _b[lncapital] = 1

 ( 1)  [lnoutput]lnlabor + [lnoutput]lncapital = 1

           chi2( 1) =     0.03
         Prob > chi2 =    0.8622
```

The output above indicates that the variance of the idiosyncratic error term is a function of firm size. Also, we failed to reject the hypothesis that the firms use a constant returns-to-scale technology.

◁

❑ Technical note

In small samples, the conditional heteroskedastic estimators will lack precision for the variance parameters and may fail to converge altogether.

❑

▷ Example 3: The truncated-normal model

Let's turn our attention to the truncated-normal model. Once again, we will use fictional data. For this example, we have 1,231 observations on the quantity of output, the total cost of production for each firm, the prices that each firm paid for labor and capital services, and a categorical variable measuring the quality of each firm's management. After taking the natural logarithm of the costs (lncost), prices (lnp_k and lnp_l), and output (lnout), we fit a stochastic cost frontier model and specify the distribution for the inefficiency term to be truncated normal.

```
. use https://www.stata-press.com/data/r18/frontier2

. frontier lncost lnp_k lnp_l lnout, distribution(tnormal) cost

Iteration 0:   Log likelihood = -2386.9523
Iteration 1:   Log likelihood = -2386.5146
Iteration 2:   Log likelihood = -2386.2704
Iteration 3:   Log likelihood = -2386.2504
Iteration 4:   Log likelihood = -2386.2493
Iteration 5:   Log likelihood = -2386.2493
```

```
Stoc. frontier normal/truncated-normal model          Number of obs =   1,231
                                                       Wald chi2(3)  =    8.82
Log likelihood = -2386.2493                            Prob > chi2   =  0.0318
```

lncost	Coefficient	Std. err.	z	P>|z|	[95% conf. interval]	
lnp_k	.3410717	.2363861	1.44	0.149	-.1222366	.80438
lnp_l	.6608628	.4951499	1.33	0.182	-.3096131	1.631339
lnout	.7528653	.3468968	2.17	0.030	.0729601	1.432771
_cons	2.602609	1.083004	2.40	0.016	.4799595	4.725259
/mu	1.095705	.881517	1.24	0.214	-.632037	2.823446
/lnsigma2	1.5534	.1873464	8.29	0.000	1.186208	1.920592
/lgtgamma	1.257862	.2589522	4.86	0.000	.7503255	1.765399
sigma2	4.727518	.8856833			3.274641	6.825001
gamma	.7786579	.0446303			.6792496	.8538846
sigma_u2	3.681119	.7503408			2.210478	5.15176
sigma_v2	1.046399	.2660035			.5250413	1.567756

```
HO: No inefficiency component: z = 5.595              Prob >= z = 0.0000
```

In addition to the coefficients, the output reports estimates for several parameters. 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 $\ln(\sigma_S^2)$, whose estimate is reported as lnsigma2. Finally, mu is the estimate of μ, the mean of the truncated-normal distribution.

In the output above, the generalized log-likelihood test for the presence of the inefficiency term has been replaced with a test based on the third moment of the OLS residuals. When $\mu = 0$ and $\sigma_u = 0$, the truncated-normal model reduces to a linear regression model with normally distributed errors. However, the distribution of the test statistic under the null hypothesis is not well established, because it becomes impossible to evaluate the log likelihood as σ_u approaches zero, prohibiting the use of the likelihood-ratio test.

However, Coelli (1995) noted that the presence of an inefficiency term would negatively skew the residuals from an OLS regression. By identifying negative skewness in the residuals with the presence of an inefficiency term, Coelli derived a one-sided test for the presence of the inefficiency term. The results of this test are given at the bottom of the output. For this example, the null hypothesis of no inefficiency component is rejected.

In the example below, we fit a truncated model and detect a statistically significant inefficiency term in the model. We might question whether the inefficiency term is identically distributed over all firms or whether there might be heterogeneity across firms. frontier provides an extension to the truncated normal model by allowing the mean of the inefficiency term to be modeled as a linear function of a set of covariates. In our dataset, we have a categorical variable that measures the quality of a firm's management. We refit the model, including the cm() option, specifying a set of

binary indicator variables representing the different categories of the quality-measurement variable as covariates.

```
. frontier lncost lnp_k lnp_l lnout, distribution(tnormal) cm(i.quality) cost

Iteration 0:  Log likelihood = -2386.9523
Iteration 1:  Log likelihood =  -2384.936
Iteration 2:  Log likelihood = -2382.3942
Iteration 3:  Log likelihood =  -2382.324
Iteration 4:  Log likelihood = -2382.3233
Iteration 5:  Log likelihood = -2382.3233

Stoc. frontier normal/truncated-normal model          Number of obs =   1,231
                                                       Wald chi2(3)  =    9.31
Log likelihood = -2382.3233                            Prob > chi2   =  0.0254
```

lncost	Coefficient	Std. err.	z	P>|z|	[95% conf. interval]	
lncost						
lnp_k	.3611204	.2359749	1.53	0.126	-.1013819	.8236227
lnp_l	.680446	.4934935	1.38	0.168	-.2867835	1.647675
lnout	.7605533	.3466102	2.19	0.028	.0812098	1.439897
_cons	2.550769	1.078911	2.36	0.018	.4361417	4.665396
mu						
quality						
2	.5056067	.3382907	1.49	0.135	-.1574309	1.168644
3	.783223	.376807	2.08	0.038	.0446947	1.521751
4	.5577511	.3355061	1.66	0.096	-.0998288	1.215331
5	.6792882	.3428073	1.98	0.048	.0073981	1.351178
_cons	.6014025	.990167	0.61	0.544	-1.339289	2.542094
/lnsigma2	1.541784	.1790926	8.61	0.000	1.190769	1.892799
/lgtgamma	1.242302	.2588968	4.80	0.000	.734874	1.749731
sigma2	4.67292	.8368852			3.289611	6.637923
gamma	.7759645	.0450075			.6758739	.8519189
sigma_u2	3.62602	.7139576			2.226689	5.025351
sigma_v2	1.0469	.2583469			.5405491	1.553251

The conditional mean model was developed in the context of panel-data estimators, and we can apply `frontier`'s conditional mean model to panel data.

◁

Stored results

frontier stores the following in e():

Scalars
e(N)	number of observations
e(df_m)	model degrees of freedom
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(chi2)	χ^2
e(ll)	log likelihood
e(ll_c)	log likelihood for H_0: $\sigma_u = 0$
e(z)	test for negative skewness of OLS residuals
e(sigma_u)	standard deviation of technical inefficiency
e(sigma_v)	standard deviation of v_i
e(p)	p-value for model test
e(chi2_c)	LR test statistic
e(p_z)	p-value for z
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)	frontier
e(cmdline)	command as typed
e(depvar)	name of dependent variable
e(function)	production or cost
e(wtype)	weight type
e(wexp)	weight expression
e(title)	title in estimation output
e(chi2type)	Wald; type of model χ^2 test
e(dist)	distribution assumption for u_i
e(het)	heteroskedastic components
e(u_hetvar)	*varlist* in uhet()
e(v_hetvar)	*varlist* in vhet()
e(vce)	*vcetype* specified in vce()
e(vcetype)	title used to label Std. err.
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
e(V_modelbased)	model-based variance

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

Consider an equation of the form

$$y_i = \mathbf{x}_i \boldsymbol{\beta} + v_i - s u_i$$

where y_i is the dependent variable, $\mathbf{x}_i$ is a $1 \times k$ vector of observations on the independent variables included as indent covariates, $\boldsymbol{\beta}$ is a $k \times 1$ vector of coefficients, and

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

The log-likelihood functions are as follows.

Normal/half-normal model:

$$\ln L = \sum_{i=1}^{N} \left\{ \frac{1}{2} \ln \left(\frac{2}{\pi} \right) - \ln \sigma_S + \ln \Phi \left(-\frac{s \epsilon_i \lambda}{\sigma_S} \right) - \frac{\epsilon_i^2}{2 \sigma_S^2} \right\}$$

Normal/exponential model:

$$\ln L = \sum_{i=1}^{N} \left\{ -\ln \sigma_u + \frac{\sigma_v^2}{2 \sigma_u^2} + \ln \Phi \left(\frac{-s \epsilon_i - \frac{\sigma_v^2}{\sigma_u}}{\sigma_v} \right) + \frac{s \epsilon_i}{\sigma_u} \right\}$$

Normal/truncated-normal model:

$$\ln L = \sum_{i=1}^{N} \left\{ -\frac{1}{2} \ln (2\pi) - \ln \sigma_S - \ln \Phi \left(\frac{\mu}{\sigma_S \sqrt{\gamma}} \right) \right.$$
$$\left. + \ln \Phi \left[\frac{(1-\gamma)\mu - s\gamma\epsilon_i}{\{\sigma_S^2 \gamma (1-\gamma)\}^{1/2}} \right] - \frac{1}{2} \left(\frac{\epsilon_i + s\mu}{\sigma_S} \right)^2 \right\}$$

where $\sigma_S = (\sigma_u^2 + \sigma_v^2)^{1/2}$, $\lambda = \sigma_u / \sigma_v$, $\gamma = \sigma_u^2 / \sigma_S^2$, $\epsilon_i = y_i - \mathbf{x}_i \boldsymbol{\beta}$, and $\Phi()$ is the cumulative distribution function of the standard normal distribution.

To obtain estimation for u_i, you can use either the mean or the mode of the conditional distribution $f(u|\epsilon)$.

$$E\left(u_i \mid \epsilon_i\right) = \mu_{*i} + \sigma_* \left\{ \frac{\phi(-\mu_{*i}/\sigma_*)}{\Phi(\mu_{*i}/\sigma_*)} \right\}$$

$$M\left(u_i \mid \epsilon_i\right) = \begin{cases} \mu_{*i} & \text{if } \mu_{*i} \geq 0 \\ 0 & \text{otherwise} \end{cases}$$

Then, the technical efficiency $(s = 1)$ or cost efficiency $(s = -1)$ will be estimated by

$$E_i = E\left\{ \exp(-su_i) \mid \epsilon_i \right\}$$

$$= \left\{ \frac{1 - \Phi\left(s\sigma_* - \mu_{*i}/\sigma_*\right)}{1 - \Phi\left(-\mu_{*i}/\sigma_*\right)} \right\} \exp\left(-s\mu_{*i} + \frac{1}{2}\sigma_*^2 \right)$$

where μ_{*i} and σ_* are defined for the normal/half-normal model as

$$\mu_{*i} = -s\epsilon_i \sigma_u^2 / \sigma_S^2$$

$$\sigma_* = \sigma_u \sigma_v / \sigma_S$$

for the normal/exponential model as

$$\mu_{*i} = -s\epsilon_i - \sigma_v^2 / \sigma_u$$

$$\sigma_* = \sigma_v$$

and for the normal/truncated-normal model as

$$\mu_{*i} = \frac{-s\epsilon_i \sigma_u^2 + \mu \sigma_v^2}{\sigma_S^2}$$

$$\sigma_* = \sigma_u \sigma_v / \sigma_S$$

In the half-normal and exponential models, when heteroskedasticity is assumed, the standard deviations, σ_u or σ_v, will be replaced in the above equations by

$$\sigma_i^2 = \exp(\mathbf{w}_i \boldsymbol{\delta})$$

where $\mathbf{w}$ is the vector of explanatory variables in the variance function.

In the conditional mean model, the mean parameter of the truncated normal distribution, μ, is modeled as a linear combination of the set of covariates, $\mathbf{w}$.

$$\mu = \mathbf{w}_i \boldsymbol{\delta}$$

Therefore, the log-likelihood function can be rewritten as

$$
\ln L = \sum_{i=1}^{N} \left[-\frac{1}{2} \ln (2\pi) - \ln \sigma_S - \ln \Phi \left(\frac{\mathbf{w}_i \boldsymbol{\delta}}{\sqrt{\sigma_S^2 \gamma}} \right) \right.
$$
$$
\left. + \ln \Phi \left\{ \frac{(1-\gamma)\,\mathbf{w}_i \boldsymbol{\delta} - s\gamma\epsilon_i}{\sqrt{\sigma_S^2 \gamma (1-\gamma)}} \right\} - \frac{1}{2} \left(\frac{\epsilon_i + s\mathbf{w}_i \boldsymbol{\delta}}{\sigma_S} \right)^2 \right]
$$

The z test reported in the output of the truncated-normal model is a third-moment test developed by Coelli (1995) as an extension of a test previously developed by Pagan and Hall (1983). Coelli shows that under the null of normally distributed errors, the statistic

$$
z = \frac{m_3}{\left(\frac{6m_2^3}{N} \right)^{1/2}}
$$

has a standard normal distribution, where m_3 is the third moment from the OLS regression. Because the residuals are either negatively skewed (production function) or positively skewed (cost function), a one-sided p-value is used.

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.

Badunenko, O., and P. Mozharovskyi. 2016. Nonparametric frontier analysis using Stata. *Stata Journal* 16: 550–589.

Badunenko, O., and H. Tauchmann. 2019. Simar and Wilson two-stage efficiency analysis for Stata. *Stata Journal* 19: 950–988.

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.

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

Greene, W. H. 2003. *Econometric Analysis*. 5th ed. Upper Saddle River, NJ: Prentice Hall.

Gutierrez, R. G., S. L. Carter, and D. M. Drukker. 2001. sg160: On boundary-value likelihood-ratio tests. *Stata Technical Bulletin* 60: 15–18. Reprinted in *Stata Technical Bulletin Reprints*, vol. 10, pp. 269–273. College Station, TX: Stata Press.

Hu, Y., G. Huang, and Y. Sasaki. 2023. robustpf: A command for robust estimation of production functions. *Stata Journal* 23: 86–96.

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.

Lian, Y., C. Liu, and C. F. Parmeter. 2023. Two-tier stochastic frontier analysis using Stata. *Stata Journal* 23: 197–229.

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.

Pagan, A. R., and A. D. Hall. 1983. Diagnostic tests as residual analysis. *Econometric Reviews* 2: 159–218. https://doi.org/10.1080/07311768308800039.

Petrin, A. K., B. P. Poi, and J. A. Levinsohn. 2004. Production function estimation in Stata using inputs to control for unobservables. *Stata Journal* 4: 113–123.

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

Stevenson, R. E. 1980. Likelihood functions for generalized stochastic frontier estimation. *Journal of Econometrics* 13: 57–66. https://doi.org/10.1016/0304-4076(80)90042-1.

Tauchmann, H. 2012. Partial frontier efficiency analysis. *Stata Journal* 12: 461–478.

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

[R] **frontier postestimation** — Postestimation tools for frontier

[R] **regress** — Linear regression

[XT] **xtfrontier** — Stochastic frontier models for panel data

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