Farming efficiency evaluation

Kasem Kunasri, MSc

Institute for Environmental Economics and World Trade
Leibniz University Hannover

Outlines

- Introduction farming efficiency
- Data extraction and preparation
- Final dataset
- How to estimate farming efficiency

Introduction farming efficiency

- Farming efficiency assesses the economic performance of a farm under resource constraints.
- It helps understand how efficiently farms utilize available inputs to achieve optimal production.
- There are two important components in farm efficiency analysis (Hoang & Nguyen, 2019)

1) Measuring efficiency scores

- Estimates the efficiency level of farms and identifies variations across different farms.

2) Identifying determinants of efficiency:

- Analyzes factors influencing efficiency to provide practical recommendations for farmers and policymakers.
- Helps in identifying strategies to improve efficiency and optimize resource use.

Introduction farming efficiency

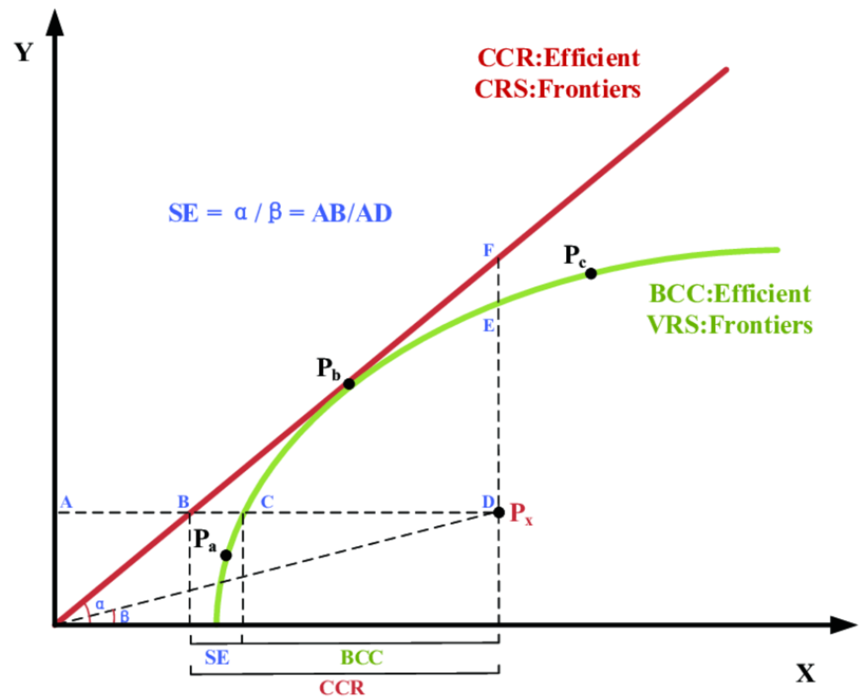
- Efficiency measurement approaches
  - Technical Efficiency: How well farms convert inputs into outputs.
    - *Example:* A farm using precision farming technology can maximize crop yield with minimal input waste.
  - Allocative Efficiency: The ability to use inputs at the lowest cost.
    - *Example:* A farmer selecting cost-effective fertilizers and irrigation methods to maximize profit margins.
  - Economic Efficiency: Combination of technical and allocative efficiency.
    - *Example:* A farm adopting mechanized plowing and strategic crop rotation to increase both yield and profitability.

Introduction farming efficiency

- There are two common methods for measuring farming efficiency

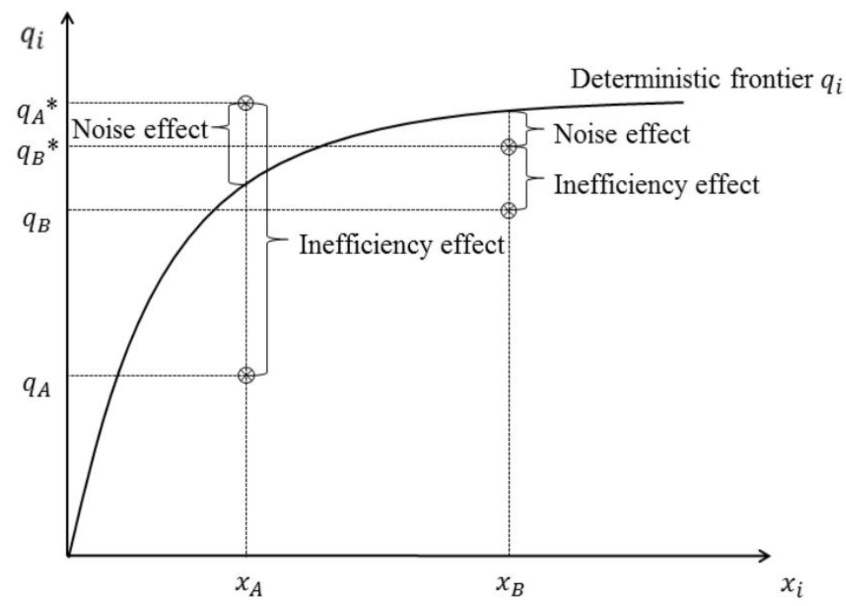
1) Data Envelopment Analysis (DEA):

   - DEA is a non-parametric method that compares the efficiency scores of farms without assuming a specific functional form .
   - It constructs an efficient frontier based on the best-performing farms and measures how far others deviate from this optimal efficiency.
   - Example: If a farm uses more inputs than another farm producing the same output, DEA identifies it as less efficient.



Introduction farming efficiency

- Stochastic Frontier Analysis (SFA):
  - Definition: A parametric approach that assumes a specific production function and considers random errors that may affect farm output.
  - How it works: SFA separates inefficiency effects from random shocks (e.g., weather, pests) that can impact production.
  - Example: A farm may have lower output due to drought, not inefficiency—SFA helps distinguish these effects



The stochastic frontier model for efficiency analysis (source: coelli et al., 2005)

Introduction farming efficiency

Comparison:

Aspect	DEA	SFA
Type	Non-parametric	Parametric
Error Consideration	No error term	Separates inefficiency & randomness
Best For	Benchmarking farms	Analyzing farm efficiency under uncertainty
Flexibility	No assumption on production function	Requires specifying a functional form

Conclusion

- **DEA** is useful for **benchmarking farms** and identifying the best-performing ones.
- **SFA** is preferred when considering **external shocks** that affect efficiency.

Introduction farming efficiency

Types of Production Functions in Efficiency Analysis:

- Cobb-Douglas Production Function:

$$Y = AX_1^{\beta_1} X_2^{\beta_2} ... X_n^{\beta_n} e^{(v-u)} \qquad (1)$$

- Translog Production Function:

$$\ln Y = \beta_0 + \sum \beta_i \ln X_i + \frac{1}{2} \sum \sum \beta_{ij} \ln X_i \ln X_j + (v-u) \qquad (2)$$

- CES (Constant Elasticity of Substitution) Production Function:

$$Y = A \left[\delta X_1^{-\rho} + (1-\delta) X_2^{-\rho}\right]^{-1/\rho} e^{(v-u)} \qquad (3)$$

Introduction farming efficiency

Production Function	Key Characteristics	Use Cases
Cobb-Douglas	Assumes **constant elasticity of substitution** among inputs; commonly used in agricultural efficiency studies.	Analyzing **returns to scale** and **input contributions** in agriculture.
Translog	More **flexible** than Cobb-Douglas; allows for **variable elasticities of substitution**.	Suitable when **input interactions** need explicit modeling.
CES (Constant Elasticity of Substitution)	Allows for **different degrees of input substitutability**.	Used in **technology adoption** and efficiency studies.

Introduction farming efficiency

- **Stochastic Frontier Model:**

The SFA model can be represented as *(Pitt and Lee, 1981; Schmidt and Sickles, 1984).*

$$q_{it} = \alpha + f(x_{it};\beta) - \mu_i + v_{it} \qquad (4)$$

Where

q_{it}: the farm output (i.e,. in the logarithm form) produced by firm i in year t

α : a common intercept

$f(x_{it};\beta)$: the production technology

x_{it} : the vector of input (i.e. in the logarithm form)

β : the associated vector of technology parameters to be estimated

v_{it} : a random two-sided noise term

$$v_{it} \sim N^{+}(0, \delta_{\omega}^{2})$$

μ_i is the non-negative inefficiency term

Introduction farming efficiency

- Limitations of standard fixed-and random-effects models
- Key issues
  - Constant inefficiency over time
    - Standard models assume inefficiency does not change, which is unrealistic.
  - Homoscedasticity assumption
    - The two-sided error term (v_{it})and inefficiency term (μ_i) are assumed to have constant variance
    - May lead to biased and inconsistent parameter estimates.

- Improvements in the Literature:
  Kumbhakar & Heshmati (1995):
  - Proposed a two-component inefficiency model:
  - Persistent inefficiency (time-invariant).
  - Time-varying inefficiency (changes over time).
  - Further refined models to address time-varying inefficiency *(Ahn et al., 2000):*

$$q_{it} = \alpha + f(x_{it};\beta) + \Omega_i + \mu_{it} + v_{it} \qquad (5)$$

Where

v_{it} is the noise

$\Omega_i \geq 0$ represents persistent inefficiency,

$\mu_{it} \geq 0$ represent time varying residual inefficiency.

Introduction farming efficiency

- Extensions to Address Firm Heterogeneity in inefficiency models
  - Debate on firm effects and inefficiency
  - Abdulai & Tietje (2007): Argued that interpreting firm effects as inefficiency requires assuming away time-invariant heterogeneity.
  - Kumbhakar et al. (2014): No strong economic justification for treating firm effects as inefficiency.

- Refinements in Stochastic Frontier Models

  Greene (2005) Proposed two models:

  True fixed-effects model:

  Separates firm heterogeneity from inefficiency.

$$q_{it} = \alpha_i + f(x_{it}; \beta) + \mu_{it} + v_{it} \tag{6}$$

  True random-effects model:

  Treats firm effects separately from inefficiency.

  Provides more flexibility in estimation.

Introduction farming efficiency

$$q_{it} = \alpha_i + f(x_{it};\beta) + \mu_{it} + v_{it} \tag{6}$$

- Challenges in True fixed-effects and random-effects models
- Key Issues with True Fixed-Effects Model (Greene, 2005)
  - α_i represents firm-specific heterogeneity.
  - Bias in Inefficiency Estimates
    - When **T<10**, inefficiency estimates suffer from severe bias.
    - Abdulai & Tietje (2007) warn that this may lead to unreliable results.
- True Random-Effects Model (Greene, 2005)

$$q_{it} = \alpha + \omega_i + f(x_{it};\beta) - \mu_{it} + v_{it} \tag{7}$$

ω_i is a firm- specific random effect distributed as $\omega_{it} \sim N^+(0, \delta_\omega^2)$

μ_{it} is the time-varying inefficiency component, distributed as $\left(\mu_{it} \sim N^+(0, \delta_\omega^2) =\right.$

Introduction farming efficiency

So, We follow the translog specification from Nguyen et al., (2021) that can be written as:

$$lnY_{it} = \alpha + \omega_i + \sum_m \beta_m lnx_{itm} + \frac{1}{2}\sum_m \sum_n \beta_{mn} lnx_{itm} lnx_{itn} - \mu_{it} + v_{it}$$

- Translog Functional Form for Efficiency Analysis
- Why Translog?
- More flexible than the Cobb-Douglas functional form.
- Allows for elasticity of substitution between inputs.
- Frequently used in stochastic frontier models.

Data extraction and coding

Variables	Data file	Variable at the household level
Y: Output_crop_value	crop_hh	_x42034
X1: Land_cultivated	crop_hh	_x42005
X2: Labor_farm	mem_hh	_x21014
X3: Cost_labor_hired	crop_hh	x42021 + _x42019 + _x42024 + _x42026 + _x42028
X4:Cost_seed	crop_hh	_x42020
X5: Cost_weeding	crop_hh	_x42022
X6: Cost_land_prep	crop_hh	_x42018
X7: Cost_fertilizer	crop_hh	_x42023
X8: Cost_pesticides	crop_hh	_x42025
X9: Cost_harvest	crop_hh	_x42027
X10: Cost_irrigation	crop_hh	_x42029

Data Extraction and Preparation

Step 1: Data Preparation and Loading

- Load the first dataset into Stata
  - Open the Stata do-file: *gen_variables.do*
  - Set the directory and load crop data for 2007-2017

- Can follow the do file named gen_variables

```
global mydir "E:\TVSEP data updated 2022"
cd "E:\TVSEP data updated 2022\data workshop\dataset_crop"


***************************************
* Load and summarize 2007 crop data
***************************************

clear all
set more off  // Prevents Stata from pausing output

* Load dataset
use "crop2007-02.dta", clear



* Summarize data
describe
summarize
```

Data extraction and coding

Step 2: Generating key variables
Defining key farming variables in Stata

Can follow the do file named gen_variables

```
// Calculate net income from crop production
gen Value_Crop_per_crop = _x42010n * _x42016n
replace Value_Crop_per_crop = 0 if Value_Crop_per_crop == .

// Generate area planted
gen crop_land = _x42005
replace crop_land = 0 if crop_land == .
gen crop_land_ha = crop_land

// Generate expenditure variables
gen cost_seed = _x42020n
replace cost_seed = 0 if cost_seed == .

gen cost_hand_weed = _x42022n
replace cost_hand_weed = 0 if cost_hand_weed == .

gen cost_preparation = _x42018n
replace cost_preparation = 0 if cost_preparation == .

gen cost_fertilizer = _x42023n
replace cost_fertilizer = 0 if cost_fertilizer == .

gen cost_pesticides = _x42025n
replace cost_pesticides = 0 if cost_pesticides == .

gen cost_harvesting = _x42027n
replace cost_harvesting = 0 if cost_harvesting == .

gen cost_irrigation = _x42029n
replace cost_irrigation = 0 if cost_irrigation == .

egen cost_hiredlabor = rowtotal(_x42021n _x42019n _x42024n _x42026n _x42028n)
replace cost_hiredlabor = 0 if cost_hiredlabor == .
```

Data extraction and coding

Step 3: Converting and aggregating data

- Standardizing financial values
- Convert financial variables to Purchasing Power Parity (PPP) USD for better comparability.
  - Why PPP Conversion?
  - PPP accounts for differences in price levels across regions.
  - Ensures financial figures are internationally comparable.
- Aggregating Data at Household and Location Level
  - Summarize financial and land-use data at the household level
  - Why Collapse Data?
  - Aggregates data from individual crops to household-level values.
  - Reduces redundancy for analysis.

```
** Convert values to PPP USD **
local source  Value_Crop_per_crop cost_seed cost_hand_weed cost_preparation ///
cost_fertilizer cost_pesticides cost_harvesting cost_irrigation cost_hiredlabor
foreach x of varlist `source' {
    gen P`x' = `x' * 0.06
    label var P`x' "`: var label `x'' (PPP USD)"
}

** Collapse data by household and location **
collapse (sum)  PValue_Crop_per_crop Pcost_seed Pcost_hand_weed ///
Pcost_preparation Pcost_fertilizer Pcost_pesticides Pcost_harvesting ///
Pcost_irrigation Pcost_hiredlabor  crop_land_ha, by(QID _x10001 _x10002 _x10003 _x10004)

** Rename variables **
ren _x10001 prov
ren _x10002 distr
ren _x10003 subdistr
ren _x10004 vill
```

```
.
. ** Summary statistics **
. sum  PValue_Crop_per_crop Pcost_seed Pcost_hand_weed Pcost_preparation ///
>  Pcost_fertilizer Pcost_pesticides Pcost_harvesting Pcost_irrigation Pcost_hiredl
```

Variable	Obs	Mean	Std. Dev.	Min	Max
PValue_Cro~p	1,806	2924.252	6071.494	0	146400
Pcost_seed	1,806	81.11915	386.2635	0	6000
Pcost_hand~d	1,806	12.06156	87.32752	0	1830
Pcost_prep~n	1,806	216.7815	340.1519	0	6060
Pcost_fert~r	1,806	380.4219	462.6739	0	6000
Pcost_pest~s	1,806	37.66123	265.093	0	9174
Pcost_harv~g	1,806	117.1067	307.5117	0	4500
Pcost_irri~n	1,806	18.45452	176.3345	0	6960
Pcost_hire~r	1,806	368.9085	1140.085	0	18729.6
crop_land_ha	1,806	2.880582	2.570743	.0003984	23.2

```
.
. ** Add year variable and reorder **
. gen year = 2007

. order year, after(QID)

.
```

Data extraction and coding

Step 4 : make panel data set

- Combining Multiple Years of Data
- *append using "……."*
- Merging: one-to-one
- *Merge 1:1 QID year using…..*

```
*********************************************
// Append all datasets 2007-2017
*********************************************

clear all
use "$mydir\data workshop\crop_2007.dta", replace
destring prov distr subdistr vill, replace

append using "$mydir\data workshop\crop_2010.dta"
append using "$mydir\data workshop\crop_2013.dta"
append using "$mydir\data workshop\crop_2017.dta"

// Merge with labor data
merge 1:1 QID year using "$mydir\data workshop\labor_farm2007-2017.dta"

// Drop unmatched cases
drop if _merge == 2

drop _merge
*********************************************
```

```
.
. // Merge with labor data
. merge 1:1 QID year using "$mydir\data workshop\labor_farm2007-2017.dta"

    Result                           # of obs.
    -----------------------------------------
    not matched                         1,401
        from master                         0  (_merge==1)
        from using                      1,401  (_merge==2)

    matched                             6,800  (_merge==3)
    -----------------------------------------

.
. // Drop unmatched cases
. drop if _merge != 3
(1,401 observations deleted)

.
end of do-file
```

```
. sum

    Variable |        Obs        Mean    Std. Dev.       Min        Max
-------------+---------------------------------------------------------
        prov |      6,800    35.44279    6.153904         31         48
       distr |      6,800    3554.294    614.1632       3101       4812
    subdistr |      6,800    355436.7    61415.73     310103     481201
        vill |      6,800    3.55e+07     6141573   3.10e+07   4.81e+07
         QID |          0
-------------+---------------------------------------------------------
        year |      6,800    2011.501     3.64008       2007       2017
PValue_Cro~p |      6,800    3697.073    6359.483          0     146400
  Pcost_seed |      6,800    91.92504    681.3553          0      40848
Pcost_hand~d |      6,800    35.09289    286.2388          0    18079.2
Pcost_prep~n |      6,800    206.3718    334.7261          0    7320.57
-------------+---------------------------------------------------------
Pcost_fert~r |      6,800      542.06    983.6691          0   30662.72
Pcost_pest~s |      6,800    38.76231    203.6834          0       9174
Pcost_harv~g |      6,800    224.6179    452.5627          0   13311.34
Pcost_irri~n |      6,800    17.78416    223.2193          0      14904
Pcost_hire~r |      6,800    386.8527    1559.209          0   77540.44
-------------+---------------------------------------------------------
crop_land_ha |      6,800    3.022186    3.413844          0     100.56
  labor_farm |      6,800    2.298529    1.177321          0         13
```

Final dataset

Step 5: log-transforming variables

- **Clear all**
- **use "final_dataset2007-2017.dta" ,clear**
- **sort QID year**

Why log transformation?

- Helps **normalize skewed data** for better statistical analysis.
- Reduces the impact of outliers
- Allows interpretation in terms of **elasticities** in regression models.

	QID	year	vill	PValue_Cro~p	Pcost_seed
1	3101030202	2007	31010302	3600	0
2	3101030202	2010	31010302	907.92	0
3	3101030202	2013	31010302	5851.064	39.68
4	3101030202	2017	31010302	2790.277	0
5	3101030204	2007	31010302	1764	0
6	3101030204	2010	31010302	4171.776	0
7	3101030204	2013	31010302	1222.144	95.232
8	3101030204	2017	31010302	1048.828	0
9	3101030205	2007	31010302	768	0
10	3101030205	2010	31010302	2933.28	0
11	3101030205	2013	31010302	2269.696	0
12	3101030205	2017	31010302	5628.45	71.7
13	3101030207	2007	31010302	2352	0
14	3101030207	2010	31010302	890.46	0
15	3101030207	2013	31010302	1190.4	297.6
16	3101030207	2017	31010302	1070.72	0
17	3101030208	2007	31010302	3259.2	0
18	3101030208	2010	31010302	4354.99	0
19	3101030208	2013	31010302	4648.512	0
20	3101030208	2017	31010302	2030.544	76.48
21	3101030210	2007	31010302	6720	504
22	3101030210	2010	31010302	94237.44	0
23	3101030210	2013	31010302	30389.92	0
24	3101030212	2007	31010302	34272	0
25	3101030212	2010	31010302	12657.92	0
26	3101030212	2013	31010302	98.20799	0
27	3101030901	2013	31010309	0	0
28	3101030902	2013	31010309	7.44	0
29	3101030902	2017	31010309	1876.748	23.9
30	3101030903	2007	31010309	1638.566	72
31	3101030903	2010	31010309	2092.872	0

Final dataset

- **Step 6: Preparing the Final Dataset for Analysis**
- Key Transformations Applied
  - Standardization: Converted all cost and revenue variables to per hectare (Ha) basis.
  - Log Transformations: Applied logarithmic transformation to key variables to improve statistical properties.
- **Code**

```
 local source PValue_Crop_per ….
foreach x of varlist `source' {
  replace `x' = `x' * 0.16
  lab var `x' "`: var lab `x"(PPP USD/Ha)"
  }
*
// Generating log-transformed variables for production function
gen Y1 = ln(PValue_Crop_per_crop + 0.001)  // Log of crop output value
gen x1 = ln(crop_land_ha + 0.001)    // Log of land cultivated
gen x2 = ln(labor_farm)                // Log of farm labor
gen x3 = ln(Pcost_hiredlabor + 0.001)   // Log of hired labor cost
gen x4 = ln(Pcost_seed + 0.001)          // Log of seed cost
gen x5 = ln(Pcost_hand_weed + 0.001)       // Log of weeding cost
gen x6 = ln(Pcost_preparation + 0.001)     // Log of land preparation cost
gen x7 = ln(Pcost_fertilizer + 0.001)    // Log of fertilizer cost
gen x8 = ln(Pcost_pesticides + 0.001)    // Log of pesticide cost
gen x9 = ln(Pcost_harvesting + 0.001)      // Log of harvest cost
gen x10 = ln(Pcost_irrigation + 0.001)   // Log of irrigation cost
```

	QID	year	Y1	x1	x2	x3	x4	x5	
1	3101030202	2007	6.35611	1.463487	1.098612	-6.907755	-6.907755	-6.907755	3.9
2	3101030202	2010	4.978582	1.764902	1.098612	3.094718	-6.907755	-6.907755	-6.9
3	3101030202	2013	6.841798	1.444799	.6931472	3.38651	1.848423	-6.907755	4.9
4	3101030202	2017	6.101317	1.764902	.6931472	2.676367	-6.907755	-6.907755	4.8
5	3101030204	2007	5.642761	1.63335	0	3.00375	-6.907755	-6.907755	2.9
6	3101030204	2010	6.503517	.8758854	.6931472	5.011601	-6.907755	-6.907755	-6.9
7	3101030204	2013	5.275785	.8758854	.6931472	3.352378	2.7238	-6.907755	.86
8	3101030204	2017	5.122853	.8758854	0	1.929229	-6.907755	-6.907755	3
9	3101030205	2007	4.811216	.1142211	0	3.503062	-6.907755	-6.907755	1.5
10	3101030205	2010	6.151297	.4706284	.6931472	4.460775	-6.907755	-6.907755	-6.9
11	3101030205	2013	5.894823	.4706284	1.386294	-.2299164	-6.907755	-6.907755	2.
12	3101030205	2017	6.803009	1.386544	0	-.2668342	2.439996	-6.907755	4.4
13	3101030207	2007	5.930442	-.9137939	.6931472	5.041697	-6.907755	-6.907755	2.6
14	3101030207	2010	4.959164	.4706284	.6931472	3.37838	-6.907755	-6.907755	-6.9
15	3101030207	2013	5.249468	.4706284	.6931472	1.378514	3.86319	-6.907755	3.
16	3101030207	2017	5.143511	.5193888	.6931472	2.034575	-6.907755	-6.907755	2.0
17	3101030208	2007	6.256658	.247641	.6931472	4.883536	-6.907755	-6.907755	3.1
18	3101030208	2010	6.546497	1.258745	.6931472	4.460775	-6.907755	-6.907755	-6.9
19	3101030208	2013	6.611722	1.370622	.6931472	3.680872	-6.907755	-6.907755	-6.9
20	3101030208	2017	5.783481	.247641	.6931472	2.6497	2.504529	-6.907755	3.2
21	3101030210	2007	6.980263	2.079566	.6931472	3.050268	4.390007	-6.907755	4.9
22	3101030210	2010	9.620992	2.549523	.6931472	3.416789	-6.907755	-6.907755	4.9
23	3101030210	2013	8.489285	2.81427	0	6.276104	-6.907755	4.831427	-6.9
24	3101030212	2007	8.609503	2.099367	.6931472	2.744253	-6.907755	-6.907755	-6.9
25	3101030212	2010	7.613457	2.1383	.6931472	4.23252	-6.907755	-6.907755	4.5
26	3101030212	2013	2.75457	-2.709051	1.386294	-6.907755	-6.907755	-6.907755	-6.9
27	3101030901	2013	-6.907755	-4.769102	.	-6.907755	-6.907755	-6.907755	-6.9
28	3101030902	2013	.175129	-4.50986	.	-6.907755	-6.907755	-6.907755	-6.9
29	3101030902	2017	5.704717	-.0397809	0	.4256599	1.341558	.4256599	.64
30	3101030903	2007	5.568999	.4706284	.6931472	5.592184	2.444171	.8876563	3.2
31	3101030903	2010	5.813714	.4706285	.6931472	5.768824	-6.907755	-6.907755	2.8

Final dataset

- **Step 7 : Correlation test**
- Correlate (corr) - displays the correlation matrix or covariance matrix for a group of variables. It measure the direction and strength of linear relationship between two quantitative variables
- Values range from -1 to 1:

  1 → Perfect positive correlation

  0 → No correlation

  -1 → Perfect negative correlation

```
corr Y1 x1 x2 x3 x4 x5 x6 x7 x8 x9 x10
(obs=6,558)
```

	Y1	x1	x2	x3	x4	x5	x6	x7	x8	x9
Y1	1.0000									
x1	0.5348	1.0000								
x2	0.1524	0.2205	1.0000							
x3	0.2369	0.3189	0.0259	1.0000						
x4	0.0386	0.1239	0.0254	0.1444	1.0000					
x5	0.0970	0.1195	-0.0050	0.2444	0.1729	1.0000				
x6	0.1763	0.3087	0.0607	0.0023	0.0507	-0.0102	1.0000			
x7	0.3922	0.5266	0.1276	0.2535	0.0837	0.0823	0.2400	1.0000		
x8	0.1646	0.2056	0.0535	0.1879	0.2000	0.0845	0.0789	0.2443	1.0000	
x9	0.2448	0.2815	-0.0026	0.0605	0.0603	0.0091	0.2300	0.2395	0.1864	1.0000
x10	0.0701	0.0832	0.0708	0.0530	0.0972	-0.0167	-0.0139	0.0673	0.0577	-0.0237

	x10
x10	1.0000

Final dataset

Step 8 : Multicollinearity check (VIF)

- 1) Running the Regression Model
- 2. Checking for Multicollinearity (VIF Analysis)

Rule of Thumb:
VIF > 10 → High multicollinearity issue (needs attention).
VIF 5-10 → Moderate multicollinearity.
VIF < 5 → Low or no concern.

Findings:
Mean VIF = 1.21 → No serious multicollinearity concerns.
Highest VIF is for x1 (1.66), which is well below the threshold of 5.

```
reg Y1 x1 x2 x3 x4 x5 x6 x7 x8 x9 x10
```

Source	SS	df	MS
Model	7433.10659	10	743.310659
Residual	15855.656	6,547	2.42182007
Total	23288.7626	6,557	3.55174052

Number of obs	=	6,558
F(10, 6547)	=	306.92
Prob > F	=	0.0000
R-squared	=	0.3192
Adj R-squared	=	0.3181
Root MSE	=	1.5562

Y1	Coef.	Std. Err.	t	P>|t|	[95% Conf.	Interval]
x1	.7603374	.0245111	31.02	0.000	.7122876	.8083871
x2	.1820377	.044307	4.11	0.000	.0951815	.268894
x3	.0259336	.0047418	5.47	0.000	.0166381	.0352291
x4	-.0223695	.0046889	-4.77	0.000	-.0315612	-.0131777
x5	.0165975	.0060699	2.73	0.006	.0046986	.0284965
x6	-.0018549	.0053691	-0.35	0.730	-.0123801	.0086703
x7	.1117839	.0107239	10.42	0.000	.0907615	.1328063
x8	.0104218	.0047169	2.21	0.027	.0011751	.0196685
x9	.0348702	.0040023	8.71	0.000	.0270243	.0427161
x10	.0162368	.0061159	2.65	0.008	.0042475	.028226
_cons	4.621302	.0755032	61.21	0.000	4.473291	4.769313

```
. vif
```

Variable	VIF	1/VIF
x1	1.66	0.601385
x7	1.46	0.684591
x3	1.22	0.819313
x6	1.16	0.863366
x9	1.15	0.867183
x8	1.14	0.878450
x5	1.09	0.915732
x4	1.09	0.919415
x2	1.06	0.940370
x10	1.03	0.974745
Mean VIF	1.21	

IUW Leibniz Universität Hannover

- // Generate Quadratic Terms (Squared Inputs)

```
// Generate Quadratic Terms (Squared Inputs)
foreach var in x1 x2 x3 x4 x5 x6 x7 x8 x9 x10 {
    gen `var'_sq = `var'^2
}
```

- // Generate Interaction Terms (Cross-Products of Inputs)

```
//Generate Interaction Terms (Cross-Products of Inputs)
 gen x1_x2 = x1*x2
 gen x1_x3 = x1*x3
 gen x1_x4 = x1*x4
 gen x1_x5 = x1*x5
 gen x1_x6 = x1*x6
 gen x1_x7 = x1*x7
 gen x1_x8 = x1*x8
 gen x1_x9 = x1*x9
 gen x1_x10 = x1*x10
 gen x2_x3 = x2*x3
 gen x2_x4 = x2*x4
 gen x2_x5 = x2*x5
 gen x2_x6 = x2*x6
 gen x2_x7 = x2*x7
 gen x2_x8 = x2*x8
 gen x2_x9 = x2*x9
 gen x2_x10 = x2*x10
 gen x3_x4 = x3*x4
```

	QID	x1_sq	x2_sq	x3_sq	x4_sq	x5_sq	x6_sq
1	3.101e+09	2.141794	1.206949	47.71708	47.71708	47.71708	15.52948
2	3.101e+09	3.114879	1.206949	9.577276	47.71708	47.71708	47.71708
3	3.101e+09	2.087444	.480453	11.46845	3.416669	47.71708	24.61934
4	3.101e+09	3.114879	.480453	7.162939	47.71708	47.71708	23.43371
5	3.101e+09	2.667831	0	9.022514	47.71708	47.71708	8.731802
6	3.101e+09	.7671752	.480453	25.11615	47.71708	47.71708	47.71708
7	3.101e+09	.7671752	.480453	11.23844	7.419087	47.71708	.7531749
8	3.101e+09	.7671752	0	3.721924	47.71708	47.71708	9.816316
9	3.101e+09	.0130465	0	12.27144	47.71708	47.71708	2.461209
10	3.101e+09	.2214911	.480453	19.89852	47.71708	47.71708	47.71708
11	3.101e+09	.2214911	1.921812	.0528616	47.71708	47.71708	7.643121
12	3.101e+09	1.922505	0	.0712005	5.953583	47.71708	19.48342
13	3.101e+09	.8350192	.480453	25.4187	47.71708	47.71708	7.114477
14	3.101e+09	.2214911	.480453	11.41345	47.71708	47.71708	47.71708
15	3.101e+09	.2214911	.480453	1.900301	14.92424	47.71708	14.92424
16	3.101e+09	.2697647	.480453	4.139495	47.71708	47.71708	4.139495
17	3.101e+09	.0613261	.480453	23.84893	47.71708	47.71708	10.10029
18	3.101e+09	1.584439	.480453	19.89852	47.71708	47.71708	47.71708
19	3.101e+09	1.878605	.480453	13.54882	47.71708	47.71708	47.71708
20	3.101e+09	.0613261	.480453	7.020912	6.272668	47.71708	10.84347
21	3.101e+09	4.324597	.480453	9.304133	19.27216	47.71708	24.01811
22	3.101e+09	6.500069	.480453	11.67444	47.71708	47.71708	24.51218
23	3.101e+09	7.920117	0	39.38948	47.71708	23.34269	47.71708
24	3.101e+09	4.407341	.480453	7.530927	47.71708	47.71708	47.71708
25	3.101e+09	4.572327	.480453	17.91422	47.71708	47.71708	20.60476
26	3.101e+09	7.338955	1.921812	47.71708	47.71708	47.71708	47.71708
27	3.101e+09	22.74434	.	47.71708	47.71708	47.71708	47.71708
28	3.101e+09	20.33884	.	47.71708	47.71708	47.71708	47.71708
29	3.101e+09	.0015825	0	.1811863	1.799779	.1811863	.4207763
30	3.101e+09	.2214911	.480453	31.27252	5.973974	.7879338	10.74474
31	3.101e+09	.2214912	.480453	33.27933	47.71708	47.71708	8.246002

How to estimate farming efficiency

- // Calculate CRE household-level means for inputs

```
// Calculate CRE Household-Level Means for Inputs
foreach var in x1 x2 x3 x4 x5 x6 x7 x8 x9 x10 {
    bysort QID: egen mean_`var' = mean(`var')
}
```

- // Set panel data structure

```
// Set Panel Data Structure
destring QID, replace
xtset QID year
```

```
. // Set Panel Data Structure
. destring QID, replace
QID: all characters numeric; replaced as double

. xtset QID year
       panel variable:  QID (unbalanced)
        time variable:  year, 2007 to 2017, but with gaps
                delta:  1 unit

.
```

How to estimate farming efficiency

- (1) Translog stochastic frontier production estimation from the true random-effects model
- //sfpanel → Stochastic Frontier Panel Data Estimation.
  - //Y1 → Log-transformed output (dependent variable).
  - //x1, x2, ..., x10 → Main inputs.
  - //x1_sq, x2_sq, ..., x10_sq → Quadratic terms (capture non-linearity).
  - //x1_x2, ..., x9_x10 → Interaction terms (capture complementarities).
  - //model(tre) → Time-varying random effects (TRE) model.
  - //rescale base(7) → Rescales inefficiency estimates.
  - //simtype(genhalton) nsim(50) → Uses Halton sequences with 50 simulations.
  - //difficult → Helps when convergence is challenging.
  - //cluster(vill) → Adjusts standard errors at the village level. //

```
rescale:       Log simulated-likelihood = -9205.2136
rescale eq:    Log simulated-likelihood = -9199.3594
Iteration 0:   Log simulated-likelihood = -9199.3594
Iteration 1:   Log simulated-likelihood = -9181.8121
Iteration 2:   Log simulated-likelihood = -9175.0452
Iteration 3:   Log simulated-likelihood = -9175.0044
Iteration 4:   Log simulated-likelihood = -9175.0044

True random-effects model (exponential)          Number of obs      =      6800
Group variable: QID                              Number of groups   =      2027
Time variable: year                              Obs per group: min =         1
                                                                avg =       3.4
                                                                max =         4

                                                 Prob > chi2     =    0.0000
Log simulated-likelihood = -9175.0044            Wald chi2(65)   =  10690.73

Number of Randomized Halton Sequences = 50
Base for Randomized Halton Sequences  = 7
                                 (Std. Err. adjusted for 220 clusters in vill)
------------------------------------------------------------------------------
             |               Robust
          Y1 |      Coef.   Std. Err.      z    P>|z|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
Frontier     |
          x1 |   .4304821   .0787478     5.47   0.000     .2761392    .5848249
          x2 |   .2583236   .1053286     2.45   0.014     .0518834    .4647639
          x3 |   .0488283   .0121892     4.01   0.000     .0249378    .0727187
          x4 |   .0384687   .0171491     2.24   0.025     .0048572    .0720803
          x5 |   .0420833   .0169492     2.48   0.013     .0088635    .0753032
          x6 |   .0062802   .0140619     0.45   0.655    -.0212806     .033841
          x7 |   .0339523    .027645     1.23   0.219    -.0202309    .0881355
          x8 |   .0694248   .0151071     4.60   0.000     .0398153    .0990342
          x9 |   .0390661   .0110012     3.55   0.000     .0175042     .060628
         x10 |   .0473165   .0243725     1.94   0.052    -.0004527    .0950856
       x1_sq |   .0549073   .0156446     3.51   0.000     .0242444    .0855701
       x2_sq |  -.0467757   .0356601    -1.31   0.190    -.1166682    .0231167
       x3_sq |    .009052   .0009489     9.54   0.000     .0071921    .0109119
       x4_sq |   .0048941     .00258     1.90   0.058    -.0001625    .0099508
       x5_sq |   .0060243   .0021293     2.83   0.005      .001851    .0101976
       x6_sq |  -.0043473   .0014323    -3.04   0.002    -.0071545     -.00154
```

How to estimate farming efficiency

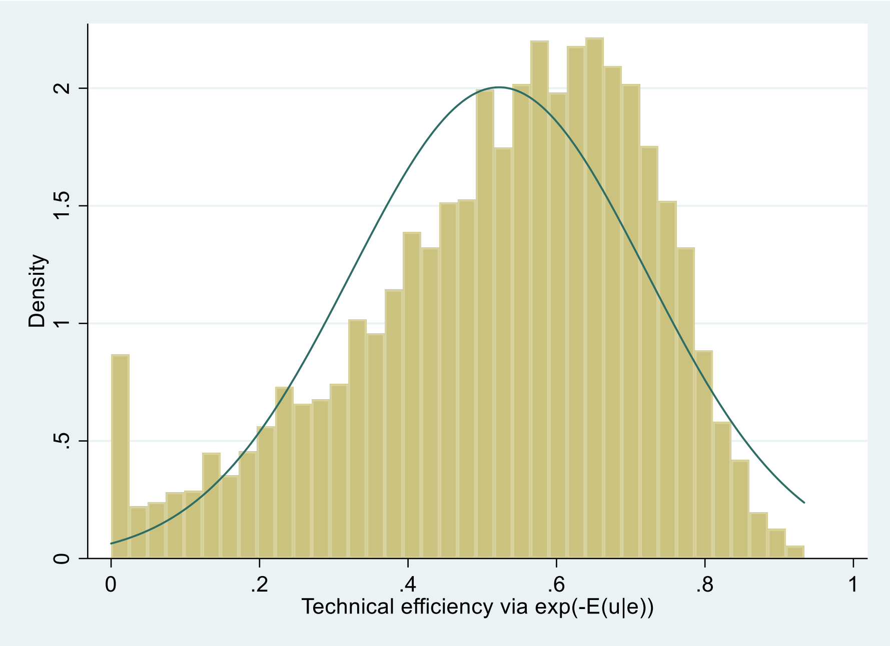
- (2) Translog stochastic frontier production estimation from the true random-effects model with Mundlak's adjustments (CRE)
- //sfpanel → Stochastic frontier panel data estimation.
  - Option are the same (1)
  - Add mean_x1, ..., mean_x10 → Household-level means

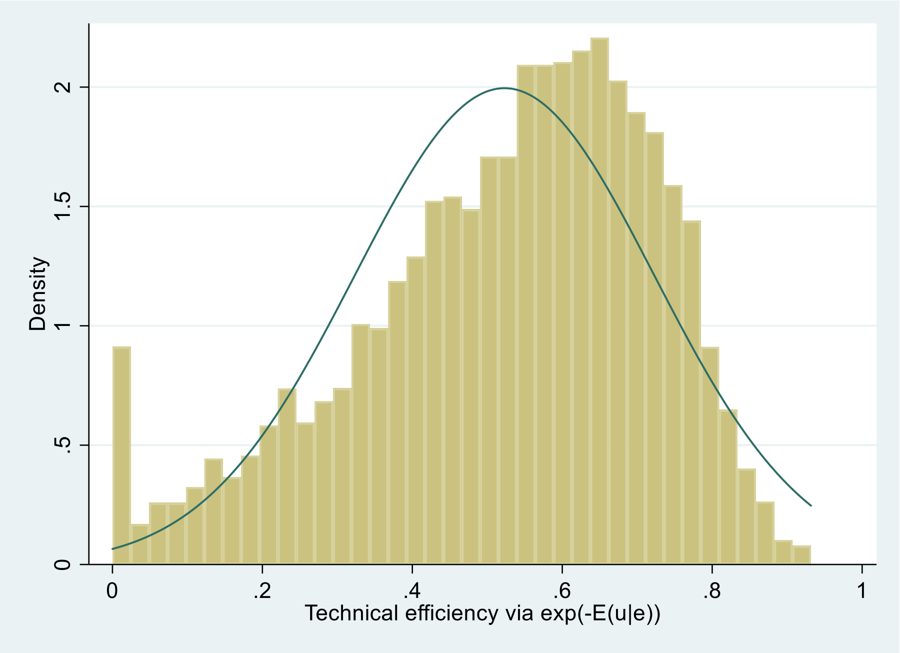
```
rescale:       Log simulated-likelihood = -9180.2489
rescale eq:    Log simulated-likelihood = -9173.7747
Iteration 0:   Log simulated-likelihood = -9173.7747
Iteration 1:   Log simulated-likelihood = -9155.5157
Iteration 2:   Log simulated-likelihood = -9149.3135
Iteration 3:   Log simulated-likelihood = -9149.2722
Iteration 4:   Log simulated-likelihood = -9149.2722

True random-effects model (exponential)              Number of obs   =     6800
Group variable: QID                                  Number of groups =    2027
Time variable: year                                  Obs per group: min =     1
                                                                    avg =   3.4
                                                                    max =     4

                                                     Prob > chi2     =   0.0000
Log simulated-likelihood = -9149.2722                Wald chi2(75)   = 12243.13

Number of Randomized Halton Sequences = 50
Base for Randomized Halton Sequences  = 7
                                 (Std. Err. adjusted for 220 clusters in vill)
------------------------------------------------------------------------------
             |               Robust
          Y1 |      Coef.   Std. Err.      z    P>|z|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
Frontier     |
          x1 |   .3755144   .0826505     4.54   0.000     .2135225    .5375064
          x2 |   .2337097   .1063848     2.20   0.028     .0251994      .44222
          x3 |   .0524738   .0122444     4.29   0.000     .0284754    .0764723
          x4 |   .0340542   .0171616     1.98   0.047     .0004181    .0676904
          x5 |   .0444125   .0170146     2.61   0.009     .0110644    .0777606
          x6 |   .0110276   .0145154     0.76   0.447    -.017422     .0394773
          x7 |   .0230094   .0278973     0.82   0.409    -.0316683    .0776872
          x8 |   .0672584   .0156651     4.29   0.000     .0365553    .0979615
          x9 |   .0360841   .0109884     3.28   0.001     .0145473     .057621
         x10 |   .0443483   .0234778     1.89   0.059    -.0016672    .0903639
       x1_sq |    .054395   .0156441     3.48   0.001     .0237332    .0850568
       x2_sq |   -.029581   .0355768    -0.83   0.406    -.0993102    .0401483
       x3_sq |   .0092046   .0009471     9.72   0.000     .0073484    .0110609
       x4_sq |   .0046807   .0025061     1.87   0.062    -.000231     .0095925
       x5_sq |   .0061076   .0021011     2.91   0.004     .0019895    .0102256
       x6_sq |  -.0041191   .0014367    -2.87   0.004    -.0069351   -.0013031
```

How to estimate farming efficiency

Predict score farming efficiency

Variable	Obs	Mean	Std. Dev.	Min	Max
farm_effic~1	6,800	.5226571	.1990606	3.71e-08	.9336818
farm_effic~2	6,800	.5229796	.1999159	3.69e-08	.9318033





References

- Abdulai, A. and Tietje, H. (2007). Estimating technical efficiency under unobserved het- erogeneity with stochastic frontier models: application to northern German dairy farms. European Review ofAgricultural Economics 34: 393–416.
- Ahn, S., Good, D. and Sickles, R. (2000). Estimation of long-run inefficiency levels: a dynamic frontier approach. Econometric Reviews 19: 461–492.
- Coelli, T. J., Rao, D. S. P., O'Donnell, C. J., & Battese, G. E. (2005).An Introduction to Efficiency and Productivity Analysis (2nd ed.). Springer.
- Greene, W. (2005). Reconsidering heterogeneity in panel data estimators of the stochastic frontier model. Journal ofEconometrics 126: 269–303.
- Huy, H. T., & Nguyen, T. T. (2019). Cropland rental market and farm technical efficiency in rural Vietnam. *Land Use Policy*, *81*, 408–423.
- Kumbhakar, S. C. and Heshmati, A. (1995). Efficiency measurement in Swedish dairy farms: an application of rotating panel data, 1976–88. American Journal ofAgricultural Economics 77: 660–674
- Kumbhakar, S. C., Lien, G. and Hardaker, J. B. (2014). Technical efficiency in competing panel data models: a study ofNorwegian grain farming. Journal ofProductivity Analysis 41: 321–337.
- Nguyen, T. T., Tran, V. T., Nguyen, T. T., & Grote, U. (2021). Farming efficiency, cropland rental market and income effect: Evidence from panel data for rural Central Vietnam. *European Review of Agricultural Economics*, *48*(1), 207–248.
- Pitt, M. M. and Lee, L. F. (1981). The measurement and sources of technical inefficiency in the Indonesian weaving industry. Journal ofDevelopment Economics 9: 43–64.
- Schmidt, P. and Sickles, R. (1984). Production frontiers and panel data. Journal ofBusiness and Economic Statistics 2: 367–374.

Thank You
For Your Attention