Socioeconomic Drivers of Food Waste Over Time: A Comparative Evaluation of Panel Stochastic Frontier Models for Indirect Quantification in Chinese Households

Rui Liu, Texas A&M University, liurui@tamu.edu

Emiliano Lopez Barrera, Texas A&M University, elopezba@tamu.edu

Selected Paper prepared for presentation at the 2024 Agricultural & Applied Economics Association Annual Meeting, New Orleans, LA; July 28-30, 2024

Abstract

Efforts to achieve the United Nations' Sustainable Development Goals necessitate a global commitment to halve food loss and waste within the next decade. Reducing household food waste not only enhances environmental sustainability but also promotes economic efficiency and food security. In this context, the issue of consumer food waste in China—a country that accounts for nearly a quarter of global food production—remains largely unexplored. This study advances the existing literature on indirect quantification methods for consumer food waste by specifically applying Stochastic Frontier Analysis (SFA) to panel data. This approach facilitates a detailed examination of the socioeconomic factors influencing household food waste over time, providing comprehensive insights into the underlying dynamics. Employing balanced panel data from the China Health and Nutrition Surveys, our findings indicate an average household food waste rate of approximately 20%. We also explore the impact of heterogeneous family characteristics on food waste, revealing that households with refrigerators, a higher number of members and children, and those located in southern areas tend to have higher levels of waste. Conversely, no significant associations were found between household income and living in rural versus urban areas with respect to food waste rates. Our results offer evidence-based insights for program interventions aimed at reducing food waste.

Keywords: Food Waste, Sustainable Development Goals, Stochastic Frontier Analysis, Food Security.

1. Introduction

Addressing food loss and waste is critical to achieving the United Nations' Sustainable Development Goal of ensuring sustainable consumption and production patterns (United Nations, 2015). According to the FAO report, global food waste amounts to 1.03 billion tons annually, accounting for about 17% of global food production. China is a populous country and agricultural-dominated country. Global food production was 2722 billion kg, of which China's output was 664 billion kg, accounting for nearly a quarter of global food production in 2019 (FAO et al, 2020). Though total grain production has doubled over the past 40 years (Liu and Zhou, 2021), food security in China is threatened by urbanization, climate change, and limited water and land resources (Lam et al., 2013, Ghose, 2014). To reinforce food security, it's essential not only to focus on increasing food production but also to minimize food loss and waste from the consumer demand side.

Some research estimates the average food waste varies among Chinese provinces ranging from 12 to 33 kg/cap/yr, with a carbon footprint from 30 to 96 kg CO2e/cap/yr using a Bayesian Belief Network (BBN) model (Song et al., 2018). As China's economy grows and urbanization continues, food wastage, particularly in the restaurant sector, is projected to rise further (Gustavsson et al., 2011). Indeed, China's economy is projected to significantly shape global food waste trends in the coming decades, intensifying pressures on resource use (Lopez Barrera & Hertel 2021). The literature primarily identifies three methods for measuring household food waste: questionnaires and interviews, direct weighing, and inferring from secondary data, often employing a combination of these approaches for enhanced accuracy. For instance, the China Health and Nutrition Survey (CHNS) database shows 16 kg/capita of household food waste annually, equivalent to carbon, water, and ecological footprints of 40 kg CO2 eq., 18 m3, and

173 gm2, respectively (Song et al., 2015). A direct-weighing method used in rural households in Shandong Province, China shows that the average amount of food waste generation is 7.25 kg per capita per rural household (Li et al., 2021), The urban restaurant's food waste in China is 11 kg/cap/year in 2015 (Wang et al., 2017) using direct weighing and field surveys. A survey-based method combined with the municipal Solid Waste data in Shenzhen City, China estimating the quantity of avoidable household food waste (Zhang et al., 2018). An alternative indirect method for estimating household food waste outside of China employs the Stochastic Production Frontier Approach (SFA), wherein food waste is characterized as input inefficiency, as demonstrated in studies by Yu and Jaenicke (2020) and Smith and Landry (2020). They treated household food consumption as a production process that transforms various types of food inputs, quantified by the gram weights of food acquired, into chemical energy necessary for human metabolic processes and physical activities. Food waste is subsequently identified as input inefficiency within this production framework, analyzed using a cross-sectional stochastic production frontier model as outlined by Aigner, Lovell, and Schmidt (1977) and Jondrow et al. (1982).

Previous studies on household food waste in China have largely relied on direct methods such as weighing food waste in smaller households or survey-based methods in larger samples, which may suffer from recall bias and lack of nationwide representation. Our study first compares four different SFA methods for estimating household output-oriented food usage inefficiency and then advances this field by employing R.W. Shephard's Input Distance Function (IDF) from "Cost and Production Functions," Princeton Univ. Press, Princeton, 1953 with True random-effect model (TRE) using balanced panel data to estimate food waste and get heterogeneous household waste behaviors. By conceptualizing household food consumption as a production process where food inputs are converted into chemical energy necessary for human metabolic processes and physical activities. Unlike traditional Stochastic Frontier Analysis (SFA) production function, which assesses food usage efficiency based on household members' total energy expenditure (TEE)—an output influenced by diet but not immediately reflective of short-term dietary changes—IDF offers a nuanced perspective by considering how much food is required based on households' TEE. IDF also doesn't assume the same inefficiency rate of each input transferring from output-oriented inefficiency. This indirect measurement method not only reduces biases inherent in direct and survey methods but also provides a more accurate and focused evaluation of input-oriented technical inefficiencies directly, epitomized by food waste.

Our study enhances the existing literature in three distinct ways. First, it advances the methodologies for indirectly quantifying household food waste by integrating Stochastic Frontier Analysis (SFA) into a panel data setting. This approach provides a deeper understanding of the dynamic relationship between household characteristics and food waste over time. By applying this methodology to a balanced panel from the CHNS database, our research offers new insights into the indirect quantification of household food waste within China, an economy projected to significantly shape global food waste trends in the coming decades. Finally, our study explores the relationship between household heterogeneity and varying levels of food waste, thereby providing a comprehensive national perspective previously lacking in previous studies of Chinese food waste. Through this analysis, we identify critical socioeconomic factors associated with increased household food waste in China. These insights are vital for policymakers and support the development of targeted strategies to reduce food waste at the household level.

The rest of the paper is organized as follows: the next section presents the model specification and econometric approach, followed by the description of the data and main results, including distinguishing the determinants of food waste from household demographic characteristics.

2. Methodology

Current research on household food waste estimation in China p predominantly utilizes direct weighing methods for smaller households or interview surveys for larger samples. However, these approaches may be prone to selection bias or reliability issues. Our study applies a novel production function approach to assess input efficiency, conceptualizing household food consumption as a production process. We have employed the Translog Input Distance Function (IDF) to ascertain the minimum quantity of purchased food necessary to sustain a specified level of basal metabolic rate (BMR) and physical activity (PA).

Our methodology commences with the evaluation of four distinct stochastic frontier analysis (SFA) production models for panel data: the fixed-effect (FE) model, the random-effect (RE) model, the time-variant random-effect (TVD) model, and the true random-effect (TRE) model. Notably, these models are configured to quantify only output-oriented inefficiency. Following this initial phase, we integrate the translog input distance function with the best SFA model TRE to investigate input inefficiency, specifically in the context of household food waste. Furthermore, our analysis incorporates nine demographic variables to enhance our understanding of household food waste determinants. Compared with the study of Yu and Jaenicke (2020) and Smith and Landry (2020), our approach differentiates between time-invariant household heterogeneity and time-variant technical inefficiency and helps clarify how various determinants impact technical inefficiency in household food waste.

2.1 Four SFA Production Models Comparison

Schmidt and Sickles (1984) were pioneers in establishing a comprehensive framework for extending the cross-sectional stochastic frontier model to panel data analysis using conventional fixed-effect (FE) and random-effect (RE) models.

$$y_{it} = \beta_0 + x_{it}'\beta + v_{it} - u_i, \;\; i = 1, \dots, n, \qquad t = 2004, 2006, and\ 2009, \tag{1}$$

$$v_{it} \sim_{iid} \mathcal{N}(0, \sigma_v^2)$$

Where $y_{it} \in \mathcal{R}_+^1$ is the household $i's$ total energy expenditure (TEE) which is the sum of each member's BMR multiplied by PA level in year t. $x_{it} \in \mathcal{R}_+^p$ is energy of each purchased food group of household i in year t. v_{it} is the regular error term, while the unobserved individual heterogeneity, $u_i \geq 0$ represents the time-invariant technical inefficiency for each household.

First, assuming that u_i is a fixed variable and not correlated with v_{it} and x_{it}', fixed-effect production model is:

$$y_{it} = \alpha_i + x_{it}'\beta + v_{it} \tag{2}$$

Then the technical inefficiency can be estimated as $\hat{u_i} = \max(\hat{\alpha_i}) - \hat{\alpha_i} \geq 0$.

Next, if u_i is assumed to be random and uncorrelated with the frontier regressors and v_{it}, the Random-Effect production model is:

$$y_{it} = \beta_0^* + x_{it}'\beta + v_{it} - u_i^* = c_i + x_{it}'\beta + v_{it} \tag{3}$$

Where $\beta_0^* = \beta_0 - E(u_i)$, $u_i^* = u_i - E(u_i)$, $E(u_i) \geq 0$, $c_i = \beta_0^* - u_i^* = \beta_0 - u_i$.

Then a consistent estimator of technical inefficiency as $\hat{u_i} = \max(\hat{c_i}) - \hat{c_i} \geq 0$.

The time-invariant fixed- and random-effect in estimating the technical inefficiency is usually unrealistic for long panel data sets. Kumbhakar (1990) provided a time-varying inefficiency model with a distribution assumption in u_{it} which changes over time and across individuals. The u_i component is individual-specific, and the left component is time-varying and is common for all individuals. The estimator of inefficiency can be derived as the conditional mean from the conditional distribution, i.e., u_{it}= E($u_{it}|\varepsilon_{it}$) [see Jondrow et al. (1982) for a detailed discussion].

$$y_{it} = \beta_0 + x_{it}'\beta + v_{it} - u_{it} = \beta_0 + x_{it}'\beta + \varepsilon_{it}$$

$$u_{it} = (1 + \exp(at + bt^2))^{-1}u_i,$$

$$u_i \sim_{iid} \mathcal{N}^+(0, \sigma_u^2),$$

$$\widehat{u_{it}} = E[u_{it}|\varepsilon_{it}] \tag{4}$$

However, the technical inefficiency in the above models confounds with all time-invariant unobserved individual effects. Greene (2005a, b) proposed a "true random-effect" stochastic panel data model that disentangles unobserved individual differences α_i from transient technical efficiency. The estimated inefficiency can be derived as u_{it}= E($u_{it}|\varepsilon_{it}$).

$$y_{it} = \alpha_i + x_{it}'\beta + v_{it} - u_{it} = \alpha_i + x_{it}'\beta + \varepsilon_{it},$$

$$v_{it} \sim_{iid} \mathcal{N}(0, \sigma_{vt}^2),$$

$$u_{it} \sim_{iid} \mathcal{N}^+(0, \sigma_{uit}^2),$$

$$\widehat{u_{it}} = E[u_{it}|\varepsilon_{it}] \tag{5}$$

We got the input-oriented inefficiency from these four production models by applying the translog form production function.

2.2 Translog input Distance function

Next, we employed a True Random Effects (TRE) production model in conjunction with an Inefficiency Distribution Function (IDF) to calculate Household Input-Oriented Inefficiency. This approach eschews the assumption of uniform technical inefficiency across all food groups, acknowledging that food purchasing behavior depends on Total Energy Expenditure (TEE). IDF measures the minimum input quantities required to produce a given level of output, indicating how much a household could proportionally reduce all food purchasing while still supporting the same

TEE level. It is expressed as $D_I(y,x) = \max_{\lambda}\{\lambda | f\left(\frac{x}{\lambda}\right) \geq y\}$. Since the input distance function is homogeneous of degree one in x, it can be rewritten as $\frac{D}{x_1} = f(\frac{x_2}{x_1}, \ldots, \frac{x_j}{x_1}, y)$. After taking log of both sides, we obtain

$$lnD_I - lnx_1 = \beta_0 + \sum_{j=2}^{J} \beta_j ln\, \tilde{x}_j + \sum_{jm=1}^{M} \gamma_m ln\, y_m +$$

$$1/2[\sum j \sum_k \beta_{jk} ln\tilde{x}_j ln\tilde{x}_k + \sum m \sum_l \gamma_{ml} lny_m lny_l + \sum j \sum_m \delta_{jm} ln\tilde{x}_j lny_m], \quad (6)$$

Where $\tilde{x} = \left(\frac{x_2}{x_1}, \ldots, \frac{x_j}{x_1}\right)$ $and\ \beta_{jk} = \beta_{kj}\ and\ \ \gamma_{ml} = \gamma_{lm}$.

Denoting $lnD_I = u \geq 0$ and taking it to the right-hand side of the equation, we get an estimable equation in which the error term is v-u. Thus, one can use the standard translog TRE production function approach to estimate this model.

$$-lnx_{i,1,t} = \alpha_i + \sum_{j=2}^{J} \beta_j ln\, \tilde{x}_{i,j,t} + \gamma_m lny$$

$$1/2[\sum j \sum_k \beta_{jk} ln\tilde{x}_{i,j,t} ln\tilde{x}_{i,k,t} + \sum j\, \delta_j ln\tilde{x}_{i,j,t} lny + \mathrm{v}_{it} - \mathrm{u}_{it},$$

$$TE_{it} = 1/d_I = \exp(-\mathrm{u}_{it}),$$

$$TIE_{it} = 1 - TE_{it} \quad (7)$$

Where $x_{i,1,t}$ is the first food group; $\tilde{x}_{i,j,t}$ are the second and third food group after divided by $x_{i,1,t}$; y is the household daily total energy expenditure.

2.3 Determinants of Inefficiency in IDF model

We further expanded the IDF model to examine the impact of exogenous determinants on technical inefficiency by employing Caudill et al. (1995) model under the assumption of inefficiency's variance,

$$u_{it} \sim_{iid} \mathcal{N}^+(0, \sigma_{uit}^2),$$

$$\sigma_{uit}^2 = \exp\left(z'_{it}\sigma\right) \tag{8}$$

3. Data and Variables

3.1 Data description and definition of variables

3.1.1 Data

Our study employs data sourced from the China Health and Nutrition Survey (CHNS), a longitudinal household-based survey established in 1989. The CHNS is jointly conducted by the National Institute of Nutrition and Food Safety at the Chinese Center for Disease Control and Prevention, along with the Carolina Population Center at the University of North Carolina at Chapel Hill. The survey encompasses nine provinces: Heilongjiang, Liaoning, Jiangsu, Shandong, Henan, Hubei, Hunan, Guangxi, and Guizhou. Our analysis specifically utilized data from three survey rounds (2004, 2006, and 2009). Following rigorous data-cleaning procedures, we retained a consistent cohort of 808 households. The variables of interest included household total Basal Metabolic Rate (BMR), levels of physical activity, average daily food purchases, and comprehensive demographic details for each household. Table 1 shows the summary statistics of all variables.

3.1.2 Variables Description

(A) Output variables.

Our dependent variable is household Total energy expenditure (TEE) is calculated as the sum of each household member's BMR (Kcal/cap/day) multiplied by their physical activity level (PAL).

People's BMR is approximately equal to the individual's daily energy expenditure (Kcal) during sleep. Table 3.1 shows the equations for the prediction of BMR incorporating age and sex-specific data on height (m) and weight (kg) from NCD Risk Factor Collaboration (NCD-RisC).

Table 3.1 Equations for the prediction of basal metabolic rate in adults

	Age range	**Estimated BMR in kcal/day**
Male	10-18	16.6W + 77H + 572
	18-30	15.4W - 27H + 717
	30-60	11.3W + 16H + 901
	> 60	8.8W + 1 128H - 1 071
Female	10-18	7.4W + 482H + 217
	18-30	13.3W + 334H + 35
	30-60	8.7W - 25H + 865
	> 60	9.2W + 637H – 302
Boy	1	1200
	2	1410
	3	1560
	4	1690
	5	1810
	6	1900
	7	1990
	8	2070
	9	2150
Girl	1	1140
	2	1310
	3	1440
	4	1540
	5	1630
	6	1700
	7	1770
	8	1830
	9	1880

Source: FAO/WHO/UNU (1985). Basal metabolic rate (BMR) data is computed from age and sex-specific data on height (H) and weight (W) form NCD Risk Factor Collaboration (NCD-RisC). Parameters in table 3.1 are to estimate the BMR for individuals at a given age, gender, weight and height.

People's physical activity level (PAL) is categorized into four levels, light PA (1.55), moderate PA (1.76), and heavy PA (2.25) (FAO/WHO/UNU, 1985). For no working ability, we defined it as 1.

(B) Input variables.

The interviewers measured the household consumption (in grams) of 21 diverse food categories across three consecutive days and recorded them in the survey. We categorized them into three major food groups according to their nutritional characteristics. Subsequently, the quantities were converted into caloric content using the 'Chinese Food Composition Table,' which was published in 2002 and employed in the CHNS surveys of 2004, 2006, and 2009. These groups served as our independent variables.

The first food group, Carbohydrates, serves as the primary energy source and includes cereals, legumes, vegetables, and fruits. The second group encompasses proteins and fats, featuring meats, poultry, dairy products, seafood, and fats and oils. The third group consists of miscellaneous items such as snacks, beverages, infant foods, condiments, and medicinal edibles.

(C) Inefficiency determinant variables

We incorporated nine demographic variables to capture the diverse determinants of technical inefficiency across households. Region differentiates Northern and Southern provinces, delineated by the Qinling Mountains and the Huaihe River, with the North including Liaoning, Heilongjiang, Shandong, and Henan, and the South comprising Jiangsu, Hubei, Hunan, Guangxi, and Guizhou. Rural-urban classification identifies areas as rural or urban. Household Size measures the number of members per household. Storage Facilities indicates the presence (value 1) or absence (value 0) of a refrigerator. The number of Children and Number of Elderly Individuals count those under 18 and over 60, respectively. Household total gross income is adjusted for 2015 inflation rates. Education Level of Household Head and Education Level of Main Female Household Member record educational attainment, with levels of no education, grad from primary, lower middle school degree, upper middle school degree, technical or vocational degree, university or college degree, and master's degree or higher.

Table 3.2 Descriptive statistics of the variables used in the analysis.

The statistical description of the data in Year 2004				
Variable	Mean	Standard deviation	Minimum value	Maximum value
Output				
SumTEE (Kcal/day)	6334.27	2891.68	1229.20	12565.06
Input				
FoodGroup1(Kcal/day)	4092.83	4052.17	10.37	79436.96
FoodGroup2 (Kcal/day)	1481.20	1283.70	29.90	19850
FoodGroup3 (Kcal/day)	224.93	435.58	0.37	4805.60
Demographic Determinants				
North(0)-South(1)	0.48	0.50	0	1
Rural(1)-Urban(2)	1.82	0.38	1	2
Total Members	2.63	1.10	1	7
Refrigerator	0.32	0.47	0	1
Kid	0.49	0.68	0	4
Elder	0.39	0.68	0	2
Total Gross Income (Yuan)	25646.61	28032.86	0	549085.7
Head Education	1.65	1.20	0	5
Main Female Education	1.48	1.21	0	5
The statistical description of the data in Year 2006				
Variable	Mean	Standard deviation	Minimum value	Maximum value
Output				
SumTEE (Kcal/day)	6269.77	2761.27	1519.47	22777.29
Input				
FoodGroup1(Kcal/day)	3438.60	2597.81	257.36	53832.39
FoodGroup2 (Kcal/day)	1468.04	1102.65	44.95	22327.20
FoodGroup3 (Kcal/day)	228.65	434.94	1.05	3813.83
Demographic Determinants				

North-South	0.49	0.50	0	1
Rural-Urban	1.82	0.38	1	2
Total Members	2.41	1.05	1	8
Refrigerator	0.38	0.49	0	1
Kid	0.35	0.59	0	3
Elder	0.43	0.71	0	2
Total Gross Income (Yuan)	29066.33	35445.58	0	370338.5
Head Education	1.66	1.29	0	5
Main Female Education	1.45	1.29	0	5

The statistical description of the data in Year 2009

Variable	Mean	Standard deviation	Minimum value	Maximum value
Output				
SumTEE (Kcal/day)	6178.205	2950.40	1063.68	24117.01
Input				
FoodGroup1(Kcal/day)	3443.244	3220.38	78.82	51539.76
FoodGroup2 (Kcal/day)	1752.211	2542.73	27.92	47004.92
FoodGroup3 (Kcal/day	252.889	480.93	1.05	5411.75
Demographic Determinants				
North-South	0.49	0.50	0	1
Rural-Urban	1.82	0.38	1	2
Total Members	2.43	1.16	1	10
Refrigerator	0.55	0.50	0	1
Kid	0.29	0.58	0	1
Elder	0.54	0.77	0	3
Total Gross Income (Yuan)	44817.13	62403.44	0	822602.7
Head Education	1.55	1.29	0	5
Main Female Education	1.45	1.30	0	5

Table 3.2 presents a statistical summary of the output variable, input variables, and household demographic information included in our models for each year. It details characteristics of the sampled households, providing each variable's mean, standard deviation, minimum, and

maximum values. Notably, the trend in total household energy expenditure (SumTEE) shows a decrease from 2004 to 2009. During the same period, food consumption shifted, with carbohydrate intake (FoodGroup 1) decreasing and protein consumption (FoodGroup 2) increasing. Consumption from FoodGroup 3 remained consistent. Our sample indicates that households were equally selected from both North and South China. The ratio of rural to urban households was 4:1, reflecting the population distribution in China during the survey period. Additionally, the data reveals a continuous increase in household total gross income, which correlates with the rising prevalence of refrigerator ownership. The average number of children per household consistently decreased, while the number of elders increased. There were no significant changes in the educational levels of the household head and the main cooking female.

---

4. Empirical Results

4.1 Output-oriented inefficiency estimates.

Table 2 shows the parameter estimates of the four distinct production models. Most all the estimated coefficients are positive and significant at the 1% significance level across the four models. The result can be explained that a household increases food group 1's consumption by one percent, the TEE will increase by 0.372% (FE), 0.279% (RE), 0.355% (TVD) and 0.346% (TRE); if a household increases food group 2's input by one percent, the TEE will increase by 0.106% (FE), 0.0889% (RE), 0.0986 (TVD) and 0.0998% (TRE); and if a household increases food group 3's inputs by one percent, the TEE will increase by 0.0210% (FE), 0.0113% (RE), 0.0188% (TVD), and 0.0180% (TRE). The elasticities with respect to food groups 1 and 2 are considerably larger, while those for food group 3 are relatively small across all four models. This indicates that carbohydrates and proteins significantly contribute to household energy expenditure, whereas other food groups do not.

Table 4.1 Estimate parameters of the stochastic frontier production in different models.

	RE	FE	TVD	TRE
Parameters	SumTER	SumTER	SumTER	SumTER
FOODGROUP1	0.372***	0.279***	0.355***	0.346***
	(31.72)	(20.56)	(29.81)	(28.41)
FOODGROUP2	0.106***	0.0889***	0.0986***	0.0998***
	(10.67)	(7.76)	(10.20)	(10.47)
FOODGROUP3	0.0210***	0.0113*	0.0188***	0.0180***
	(4.97)	(2.32)	(4.52)	(4.38)
_cons	4.843***	5.748***	5.327***	5.384***
	(51.32)	(51.45)	(51.40)	(48.96)
Bt				
b			-2.101***	
			(-3.74)	
c			0.269	
			(1.45)	
Usigma			0.153***	-2.115***
			(7.58)	(-18.16)
Vsigma			0.0744***	-3.479***
			(28.34)	(-22.47)
Theta				0.221***
				(23.75)
N	2424	2424	2424	2424

$*: \rho < 0.1$ $**: \rho < 0.05$ $***: \rho < 0.01$

Note: The parameters in this table are the result from each SFA model after log transformation of each variable with observation of 808 in each year and 2424 totally.

As discussed by Kumbhakar (1990), when we consider the time trend of technological change. Model TVD demonstrates that inefficiency is associated with time t but not with time t-

square, indicating the impact of technological changes over time. Table 3 showcases the estimated output-oriented inefficiency levels in household food usage. It's important to note that the Fixed Effects (FE) and Random Effects (RE) models assume efficiency to remain constant over time, with the most efficient household considered as 100% efficient. Conversely, in the Time-Varying (TVD) and True Random Effects (TRE) models, efficiency is presumed to fluctuate over time.

The relatively elevated inefficiency levels observed in the FE and RE models could be partially attributed to their misclassification of inefficiency as time-invariant. Additionally, these models capture unobserved firm-specific time-invariant effects, which may not necessarily correlate with inefficiency. Consequently, the inefficiency estimates derived from the FE and RE models are likely overstated with an inefficiency rate of about 25% and 10% more in FE and RE than TVD and TRE. Only the TRE model effectively distinguishes technical inefficiency from unobserved individual heterogeneity, resulting in the lowest technical inefficiency estimate.

Table 4.2 Output-Oriented Inefficiency (%) from Different Models by Year

	RE	FE	TVD	TRE
2004	35.14	53.04	26.70	25.94
2006	35.14	53.04	29.99	27.15
2009	35.14	53.04	30.68	29.58

4.2 Input-oriented inefficiency estimates.

Table 4 presents the estimators of the IDF within the TRE model. This model computes the input-oriented inefficiency based on the premise that a household’s total required energy dictates food purchases, rather than food purchases determining total energy expenditure. This approach is particularly relevant in China, a country rich in agricultural production. Additionally, the TRE model effectively distinguishes between time-varying inefficiency and household-specific heterogeneity, thereby providing a deeper insight into the dynamics of household behavior.

Since the dependent variable is negative in food group 1. The results show a positive but statistically insignificant relationship between household total energy expenditure (TEE) and food group 1 spending, indicated by a coefficient of 0.4994 with a standard error of 0.4455. Conversely, expenditure on food group 2 (FG2) displays a negative and statistically significant relationship with food group 1 spending, as evidenced by a coefficient of 0.8399 and marked statistical significance ($p < 0.01$), suggesting that increases in spending on food group 2 are associated with decreases in spending on food group 1. This observation aligns with practical expectations that at the same energy expenditure level, increased protein consumption typically corresponds with reduced consumption of cereals and starches. Meanwhile, expenditure on food group 3 (FG3) has a negative but statistically insignificant coefficient of 0.1002, indicating a weaker and uncertain association with food group 1 spending.

Table 4.3 Estimated parameters of the stochastic frontier production function at full sample

Parameters	TRE
β_{TEE}	-0.4994
	(0.4455)
β_{FG2}	0.8399***
	(0.1958)
β_{FG3}	0.1002
	(0.0915)
β_{FG2_2}	0.1588***
	(0.0156)
β_{FG3_2}	-0.0182***
	(0.0062)

β_{FG3_3}	0.0314***
	(0.0046)
β_{TEE_2}	-0.0551**
	(0.0231)
β_{TEE_3}	0.0069
	(0.0108)
β_{TEE_TEE}	-0.0398
	(0.0526)
σ_u	0.3051***
σ_v	0.3200***
λ	0.3961***
Observations	2424
Number of Firms	808

$*: \rho < 0.1$ $**: \rho < 0.05$ $***: \rho < 0.01$

The numbers in parentheses are the standard deviation of each parameter.

The parameters in this table are from the IDF function with TRE model. β_{TEE} is the parameter of household total energy relative to the dependent variable FoodGroup 1. β_{FG2} *and* β_{FG3} are the parameter of FoodGroup 2 and FoodGroup 3. β_{FGi_j} are the parameters of interactions between each pair independent variable.

In Table 4.4, we estimated the household food waste rate in China in 2004, 2006, and 2009, which are 21.39%, 20.51%, and 21.55%.

Table 4.4 Household Average Food Waste Rate (%) by Year	
2004	21.39
	(0.0678)
2006	20.51
	(0.0572)

2009	21.55 (0.0691)

Note: The numbers in parentheses are the standard deviation of each parameter.

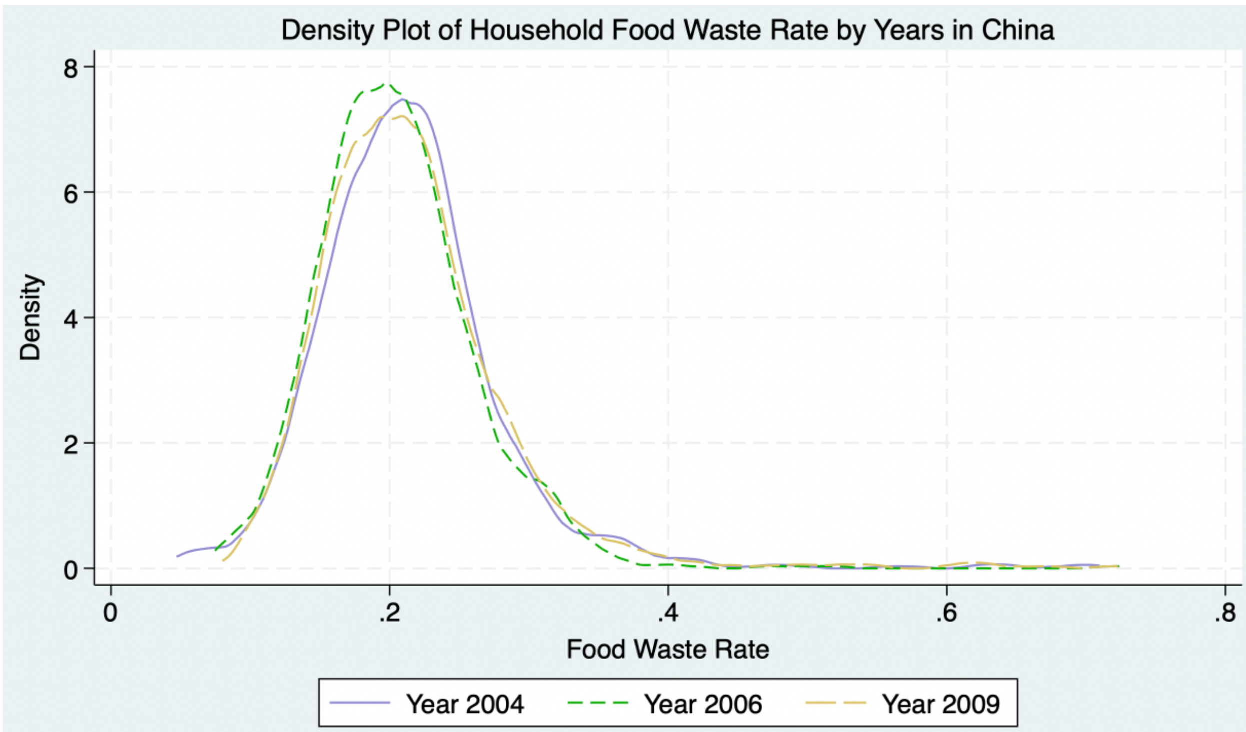


Figure 4.1

Note: These density plots show the estimated household food waste rate from IDF with TRE by year with mean of 21.39% in 2004, 20.51% in 2006, and 21.55% in 2009 with significant difference between 2004 and 2006, 2006 and 2009. But the food waste rate difference are small with fluctuation.

Pairwise t-tests were conducted to assess the average household food waste across three survey years: 2004, 2006, and 2009. The results indicate statistically significant differences in the mean levels of food waste between 2004 and 2006 ($p = 0.0050$) and between 2006 and 2009 ($p =$

0.0010). However, the annual food waste rates were relatively stable: 21.39% in 2004, 20.51% in 2006, and 21.55% in 2009. These findings suggest only minor fluctuations, indicating no significant change in household-level food waste in China over the five-year period.

4.3 Inefficiency Determinants Estimation

To investigate the effects of exogenous factors on technical inefficiency, utilized the approach proposed by Caudill et al. (1995), which proposed specifying the variance of the inefficiency distribution in the model (8). As shown in Appendix 1.5, geographic living location (south or north) in Figure 4.2, possession of a refrigerator in Figure 4.3, and number of children in the household in Figure 4.4—significantly impact food waste. Specifically, households in the south, those with a refrigerator, and those with more children tend to produce more food waste.

Households possessing refrigerators overall waste more food typically but exhibit a higher minimum and a lower maximum food waste threshold due to improved storage conditions. The increase in the lower limit of food waste can be attributed to several factors. First, the presence of a refrigerator often signifies a household's financial capability to purchase food in bulk, which, while economically advantageous, may lead to over-purchasing. Consequently, the convenience of having a variety of food options readily available can alter eating habits, resulting in food being overlooked or stored beyond its peak freshness, despite the preservation benefits offered by refrigeration. However, a refrigerator can improve food storage conditions and extend shelf life for large food purchases, which lowers the food waste maximum threshold.

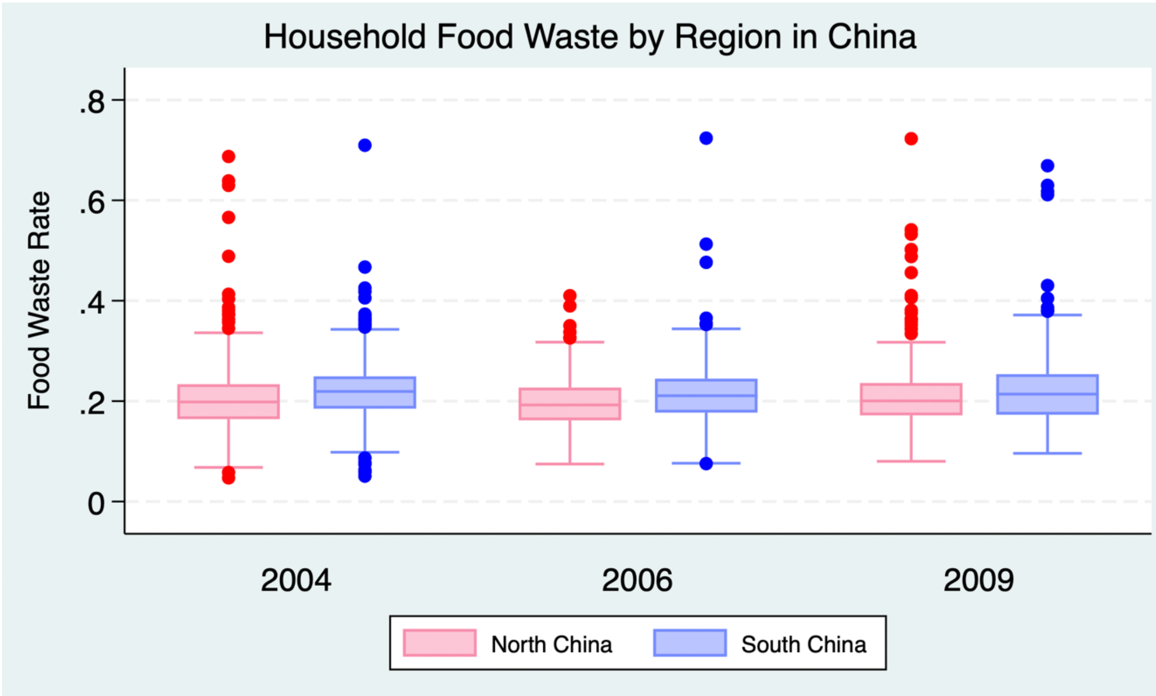


Figure 4.2 These box plots show household food waste distribution in North and South China by year. The average rates are significantly different with 20.40% in North China and 21.22% in South China.

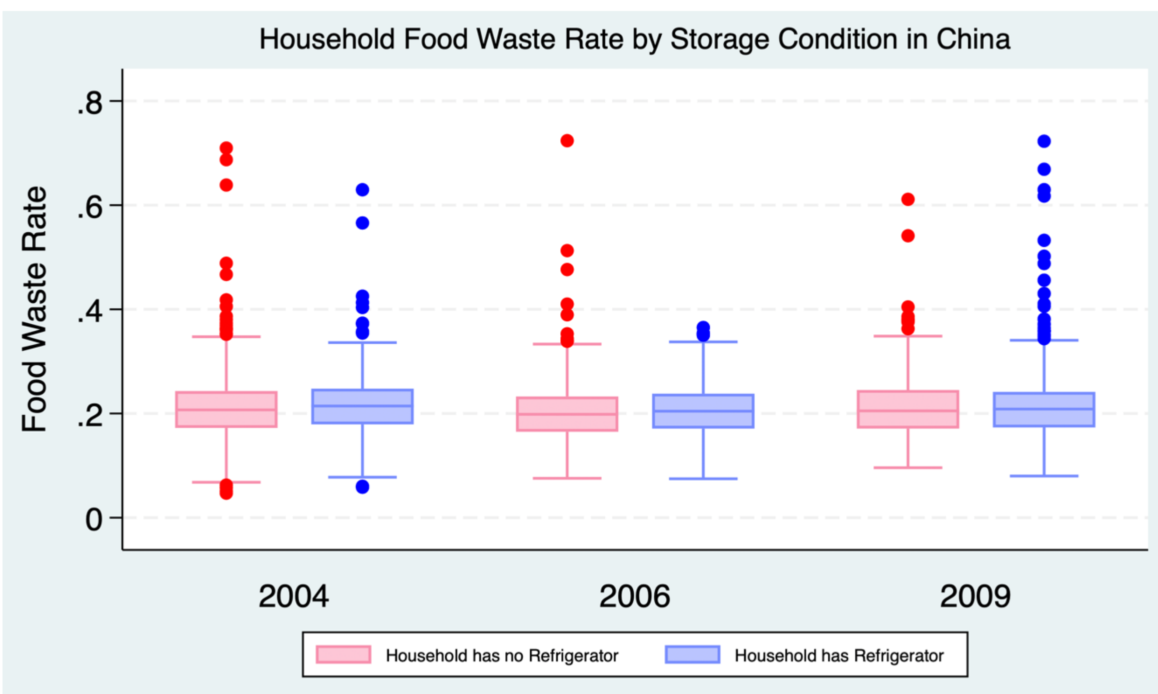


Figure 4.3 These box plots show food waste distribution of households with different storage conditions. By using having refrigerators as an indicator, the average food waste rates are significantly different with 20.9% for households without refrigerator and 21.50% for households having refrigerators.

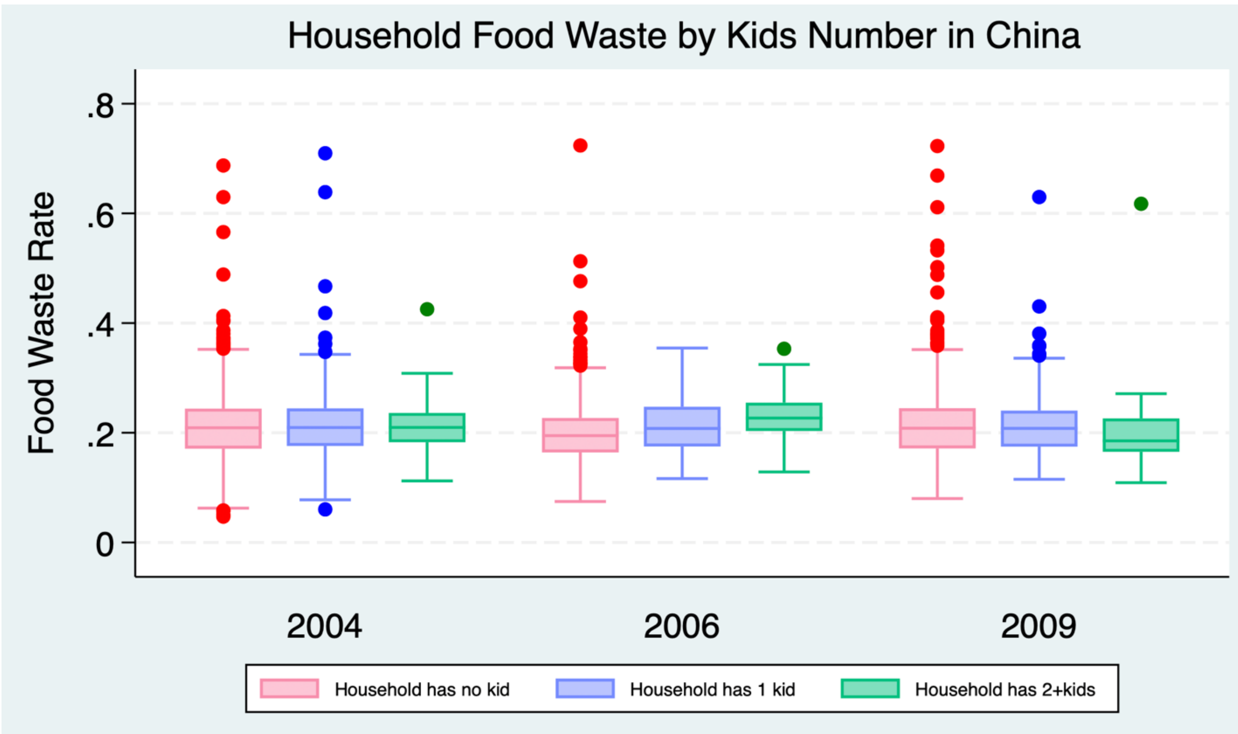


Figure 4.4 These box plots show the food waste distribution for households with different number of kids by year. The differences are significant among no kid (20.98%), one kid (21.49%), and two and two more kids (21.51%).

The number of members in a household as shown in Figure 4.5 suggests that an increase in household size tends to lead to more food waste, although this effect is not always pronounced.

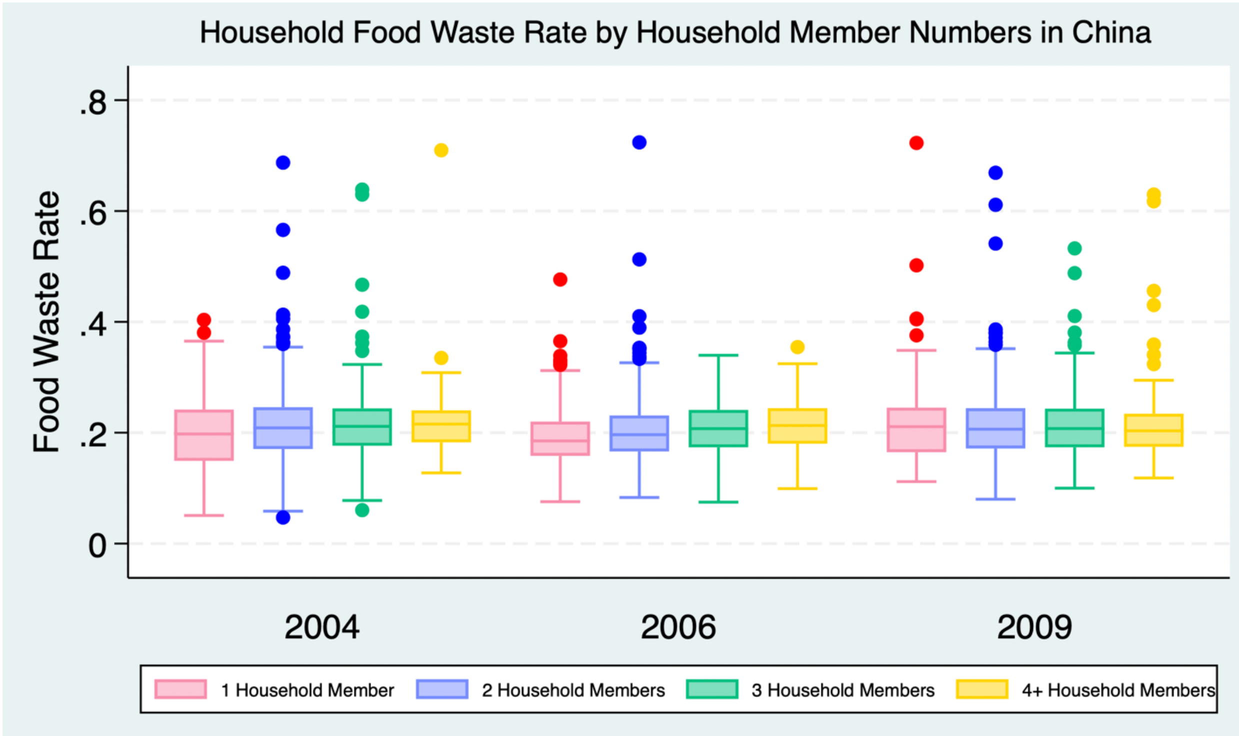


Figure 4.5 These box plots show the household food waste rate distribution by year. With more members, the rate is significant larger.

Other examined factors including education (Figures 4.6 and 4.7), income (Figure 4.8), living in rural or urban areas (Figure 4.9), and elder people numbers (Figure 4.10) did not show significant results in T-tests, indicating no discernible impact on food waste from these variables.

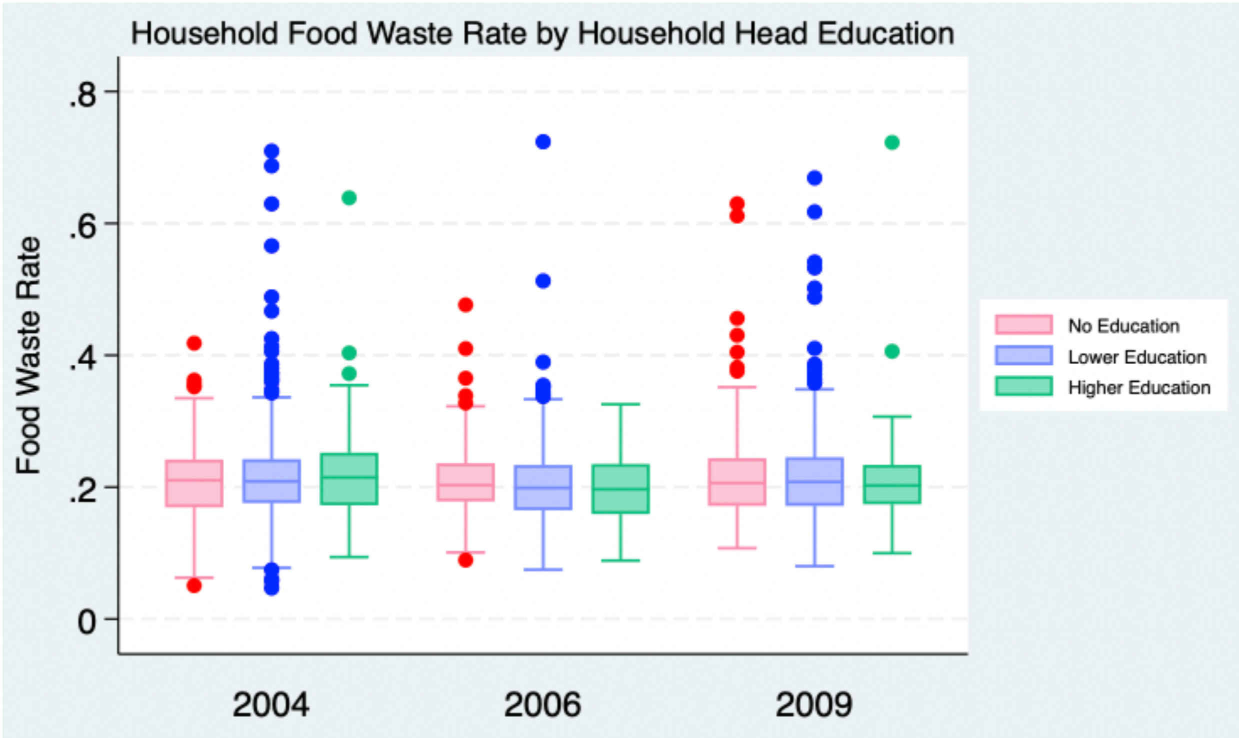


Figure 4.6 These box plots show the food waste distribution for household with different educated household head by year. The differences are not significant among the households with head without receiving education, receiving high school and lower education, and higher education than high school.

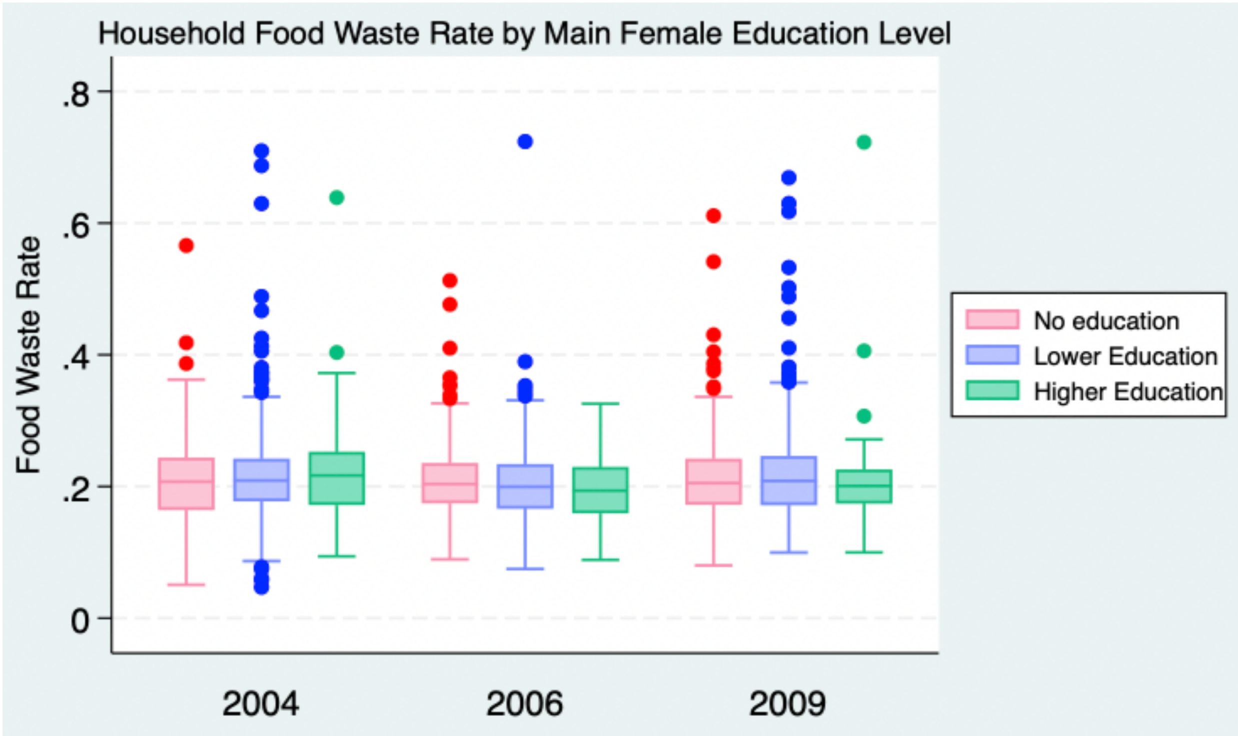


Figure 4.7 These box plots show the food waste distribution for household with different educated main cooker by year. The differences are not significant among households with the main cooker without receiving education, receiving high school and lower education, and higher education than high school.

Regarding the lack of significant influence of increasing income on food waste change, it suggests a complex interplay between financial resources and food consumption behaviors. As household income rises, there might be an initial increase in food waste due to greater purchasing power allowing for more abundant and varied food buying, often in larger quantities than needed. However, higher income levels also often correlate with high-quality food purchasing and increased food awareness, better access to food preservation technology and information, and more substantial engagement in sustainable practices. Thus, any initial increase in waste might be offset by more efficient food management as income continues to grow, leading to a plateau or even a decrease in food waste at higher income levels. This reflects a possible transition from quantity-focused to quality-focused food consumption behaviors as economic conditions improve. However, each year, the density plots consistently reveal that households in the middle-income bracket waste more food than those in both the lower and higher-income groups.

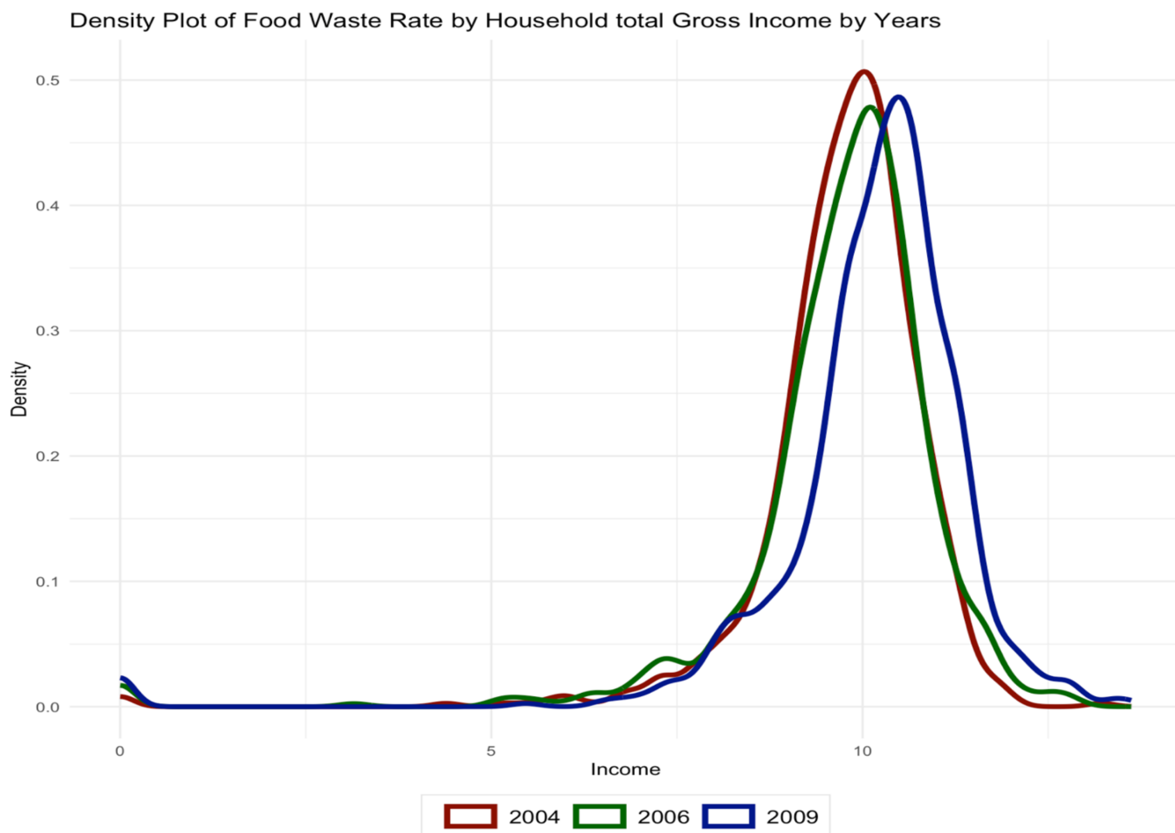


Figure 4.8 These density plots show the food waste rate for household with different total household gross income by year. Note, the income calculated by log. There is a trend that with income increases, the food waste will go up and then begin to drop. However, the TRE model didn't show income as a significant determinant influencing household food waste.

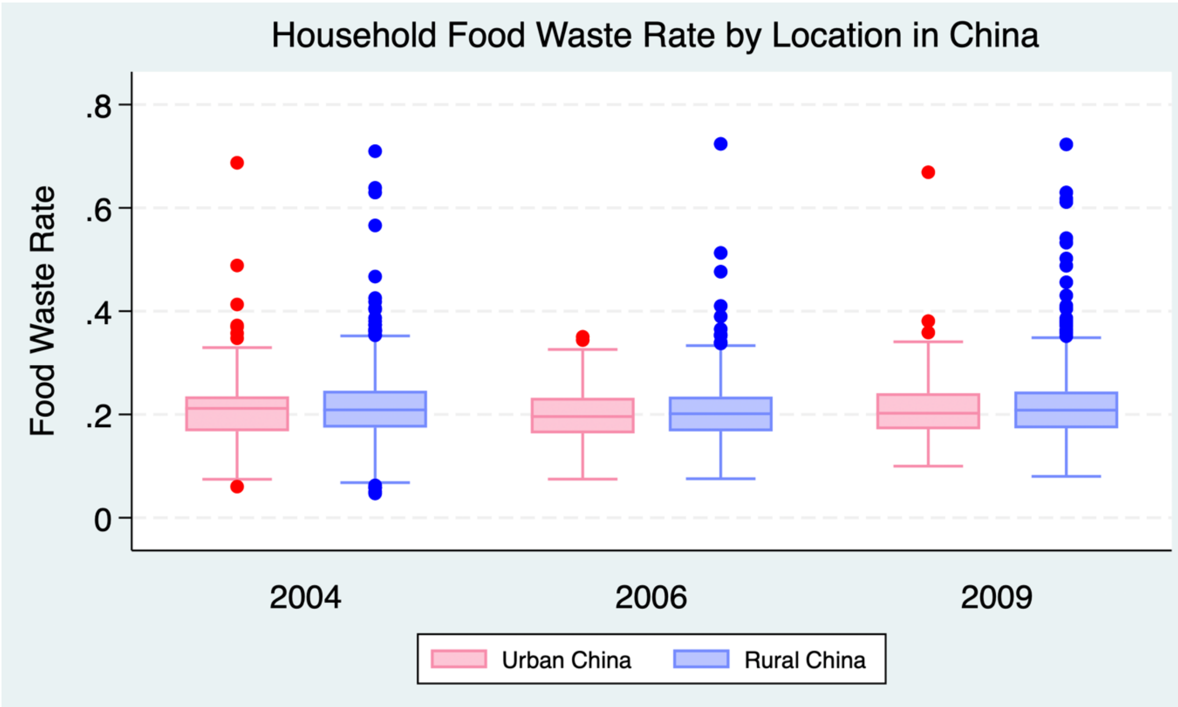


Figure 4.9 These box plots show household food waste distribution in Urban and Rural China by year. Average rate is 20.81% in Urban China and 21.94% in Rural China.

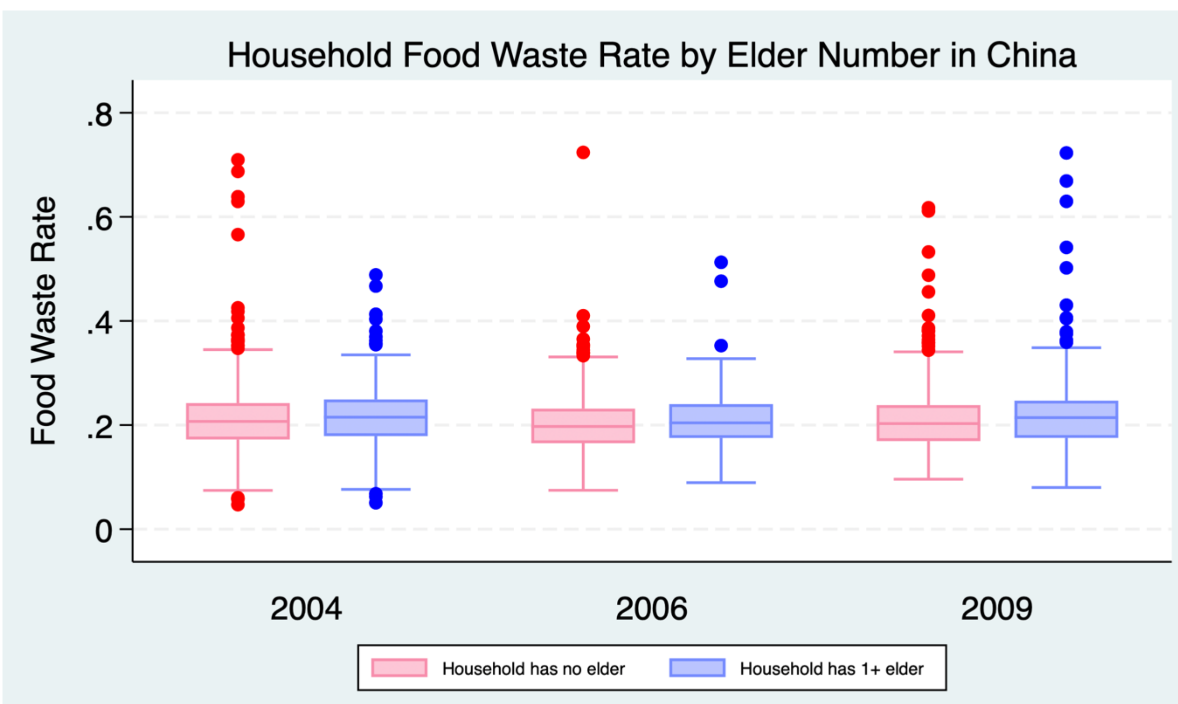


Figure 4.10 These box plots show the food waste distribution for household with different elders by year. The differences are not significant among groups.

5. Discussion and Limitation

The analysis presented in this study has highlighted significant insights into household food waste dynamics within China, underscoring the influence of household characteristics on waste levels. The application of the Input Distance Function (IDF) within the True Random-Effect (TRE) model has demonstrated that not all household characteristics impact food waste equally. Notably, factors such as geographic location, presence of refrigeration, and number of children play significant roles. Households in the south, those equipped with refrigerators, and larger families tend to exhibit higher levels of waste. This suggests that interventions aimed at reducing food waste in China should consider regional and household-specific strategies.

However, the finding that other demographic variables, such as income and education level, did not show significant effects on food waste suggests complexities that might require deeper investigation into cultural habits or more refined data collection methods. The study also reaffirms the utility of the stochastic frontier approach (SFA) in analyzing inefficiencies within household food consumption, aligning with previous research that suggests household food waste can be viewed through the lens of production inefficiency.

Our results should be interpreted in the context of several potential limitations. For instance, a notable limitation of the True Random Effects (TRE) model is its potential to conflate household-specific heterogeneity with time-invariant structural inefficiency. Consequently, the presence of a time-invariant component of inefficiency, alongside a time-varying element, may lead the TRE model to underestimate total inefficiency and, as a result, overestimate technical efficiency. Future research will aim to differentiate persistent inefficiency from time-invariant household heterogeneity. Furthermore, the model currently aggregates food waste across all food groups without specifying the waste attributable to each category. Future studies would benefit from using more recent, extended temporal data to examine household food waste dynamics in China more thoroughly.

6. Conclusion

This research enhances the existing literature on indirectly estimating household-level food waste. Previous studies have utilized Stochastic Frontier Analysis (SFA) within the literature to impute food waste as an inefficiency in household production functions using cross-sectional data. In this study, we extend the application of these methodologies by introducing SFA into panel data settings. Specifically, in this study we quantify the technical inefficiency of the China household food usage inefficiency using a panel dataset over the years of 2004, 2006, and 2009. We contrast four distinct stochastic frontier production models, specifically, conventional fixed-effects (FE), random-effects (RE), Time-variant random-effect (TVD), and "true" random-effect (TRE) models. The results indicate that inefficiency estimates are sensitive to model specifications of household unobserved heterogeneity. The conventional FE and RE models appear to overestimate the inefficiency since the inefficiency is time-variant and household-specific unobserved heterogeneity is confounded with the inefficiency term.

Considering the suitability of the TRE model for estimating the household output-oriented technical inefficiency, we integrate it with IDF to estimate the input-oriented efficiency, i.e., household food waste. Our results suggest modest changes in household-level food waste rate from

2004 to 2009 in China. This temporal analysis facilitates an examination of the socio-economic dynamics influencing varying degrees of household food waste across the region.

Our findings highlight key socioeconomic factors linked to higher levels food waste at the household level in China, offering crucial insights for policymakers and intervention strategies dedicated to mitigating food waste. Refrigerators' prevalence, kids number increase, and the eating habits difference between North and South China are the main factors associated with higher degrees of household food waste. The findings are particularly relevant for policymakers and stakeholders in designing targeted interventions that address the specific needs of diverse households. Moving forward, addressing food waste will contribute to broader environmental sustainability and food security goals in China and potentially other similar contexts globally.

References

Aigner, D., Lovell, C. K., & Schmidt, P. (1977). Formulation and estimation of stochastic frontier production function models. Journal of econometrics, 6(1), 21-37.

Barrera, E. L., & Hertel, T. (2021). Global food waste across the income spectrum: Implications for food prices, production and resource use. Food Policy, 98, 101874.

Battese, G. E., & Coelli, T. J. (1995). A model for technical inefficiency effects in a stochastic frontier production function for panel data. Empirical economics, 20, 325-332.

Chinese dietary reference intakes, 1st edn. Light Industry Press, Beijing, China

FAO, Ifad, UNICEF, WFP, & WHO. (2020). The state of food security and nutrition in the world 2020. transforming food systems for affordable healthy diets. *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*.

Food Losses and Waste in China and Their Implication for Water and Land, Junguo Liu, Jan Lundqvist, Josh Weinberg, and Josephine Gustafsson, Environmental Science & Technology 2013 47 (18), 10137-10144 DOI: 10.1021/es401426b

Ghose, B. (2014). Food security and food self-sufficiency in China: from past to 2050. *Food and Energy Security*, *3*(2), 86-95.

Gustavsson, J., Cederberg, C., Sonesson, U., Van Otterdijk, R., & Meybeck, A. (2011). Global food losses and food waste.

Lam, H. M., Remais, J., Fung, M. C., Xu, L., & Sun, S. S. M. (2013). Food supply and food safety issues in China. *The Lancet*, *381*(9882), 2044-2053.

Li, X., Jiang, Y., & Qing, P. (2023). Estimates of household food waste by categories and their determinants: evidence from China. Foods, 12(4), 776.

Li, Y., Wang, L. E., Liu, G., & Cheng, S. (2021). Rural household food waste characteristics and driving factors in China. *Resources, Conservation and Recycling*, *164*, 105209.

Liu, J., Lundqvist, J., Weinberg, J., & Gustafsson, J. (2013). Food losses and waste in China and their implication for water and land. Environmental science & technology, 47(18), 10137-10144.

Liu, Y., & Zhou, Y. (2021). Reflections on China's food security and land use policy under rapid urbanization. *Land Use Policy*, *109*, 105699.

Nations, U. (2015). Transforming our world: The 2030 agenda for sustainable development. *New York: United Nations, Department of Economic and Social Affairs*, *1*, 41.

Smith, T. A., & Landry, C. E. (2021). Household food waste and inefficiencies in food production. American Journal of Agricultural Economics, 103(1), 4-21.

Song, G., Li, M., Semakula, H. M., & Zhang, S. (2015). Food consumption and waste and the embedded carbon, water and ecological footprints of households in China. *Science of the Total Environment*, *529*, 191-197.

Song, G., Semakula, H. M., & Fullana-i-Palmer, P. (2018). Chinese household food waste and its' climatic burden driven by urbanization: A Bayesian Belief Network modelling for reduction possibilities in the context of global efforts. Journal of Cleaner Production, 202, 916-924.

Song, G., Semakula, H. M., & Fullana-i-Palmer, P. (2018). Chinese household food waste and its' climatic burden driven by urbanization: A Bayesian Belief Network modelling for reduction possibilities in the context of global efforts. Journal of Cleaner Production, 202, 916-924. https://doi.org/10.1016/j.jclepro.2018.08.233

United Nations Environment Programme. 2021. Food Waste Index Report 2021. Nairobi.

Wang, L. E., Liu, G., Liu, X., Liu, Y., Gao, J., Zhou, B., ... & Cheng, S. (2017). The weight of unfinished plate: A survey based characterization of restaurant food waste in Chinese cities. *Waste Management*, *66*, 3-12.

Yang, Y. X., Wang, G. Y., & Pan, X. C. (2009). China food composition. *Peking University Medical Press, Beijing*, *42*, 795-799.

Yu, Y., & Jaenicke, E. C. (2020). Estimating food waste as household production inefficiency. American Journal of Agricultural Economics, 102(2), 525-547.

Zhang, H., Duan, H., Andric, J. M., Song, M., & Yang, B. (2018). Characterization of household food waste and strategies for its reduction: A Shenzhen City case study. Waste Management, 78, 426-433.

Aigner, D., Lovell, C. K., & Schmidt, P. (1977). Formulation and estimation of stochastic frontier production function models. Journal of econometrics, 6(1), 21-37.

Jondrow, J., Lovell, C. K., Materov, I. S., & Schmidt, P. (1982). On the estimation of technical inefficiency in the stochastic frontier production function model. Journal of econometrics, 19(2-3), 233-238.

Schmidt, P., & Sickles, R. C. (1984). Production frontiers and panel data. Journal of Business & Economic Statistics, 2(4), 367-374.

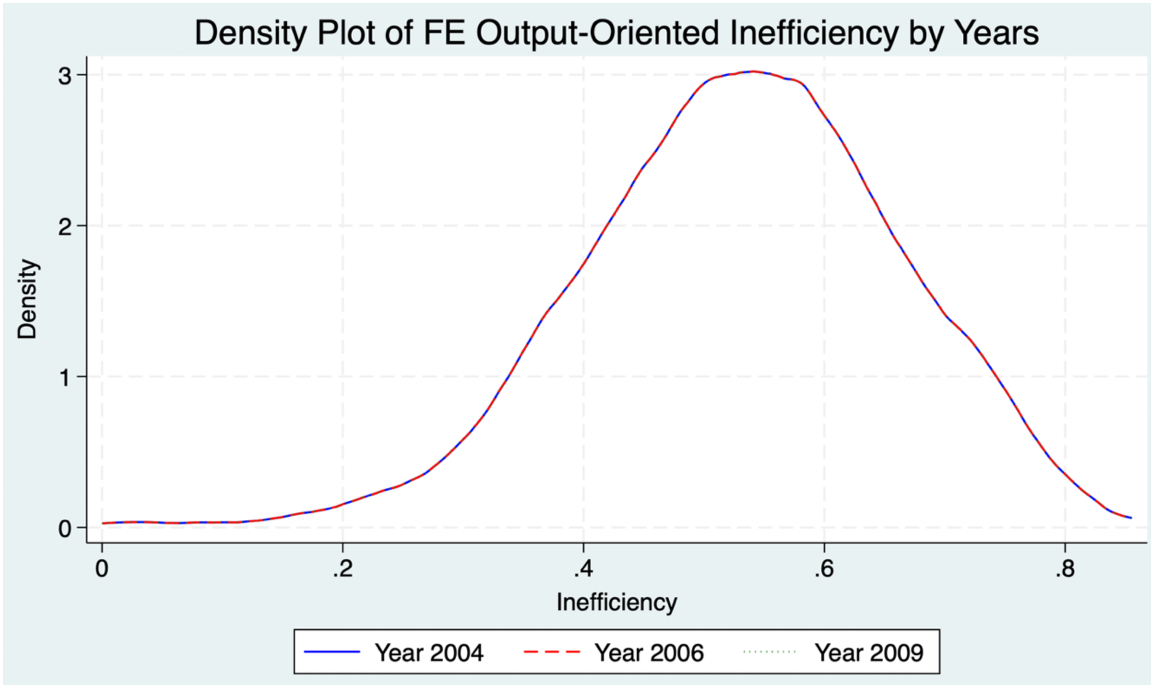
Kumbhakar, S. C. (1990). Production frontiers, panel data, and time-varying technical inefficiency. Journal of econometrics, 46(1-2), 201-211.

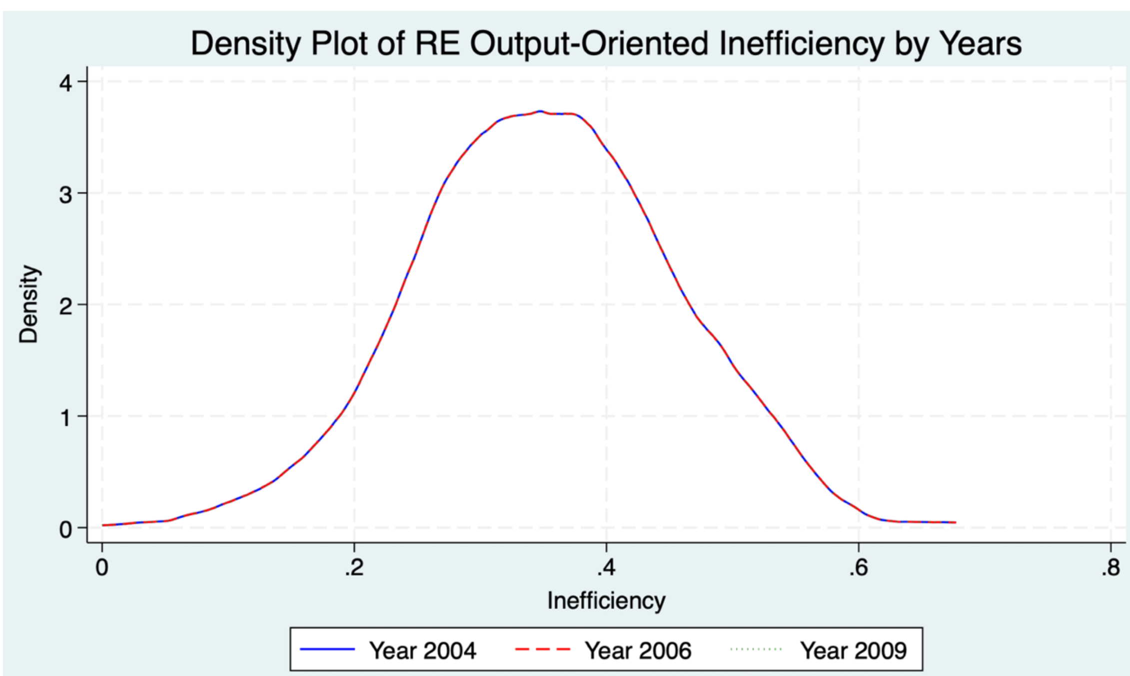
Greene, W. (2005). Fixed and random effects in stochastic frontier models. Journal of productivity analysis, 23, 7-32.

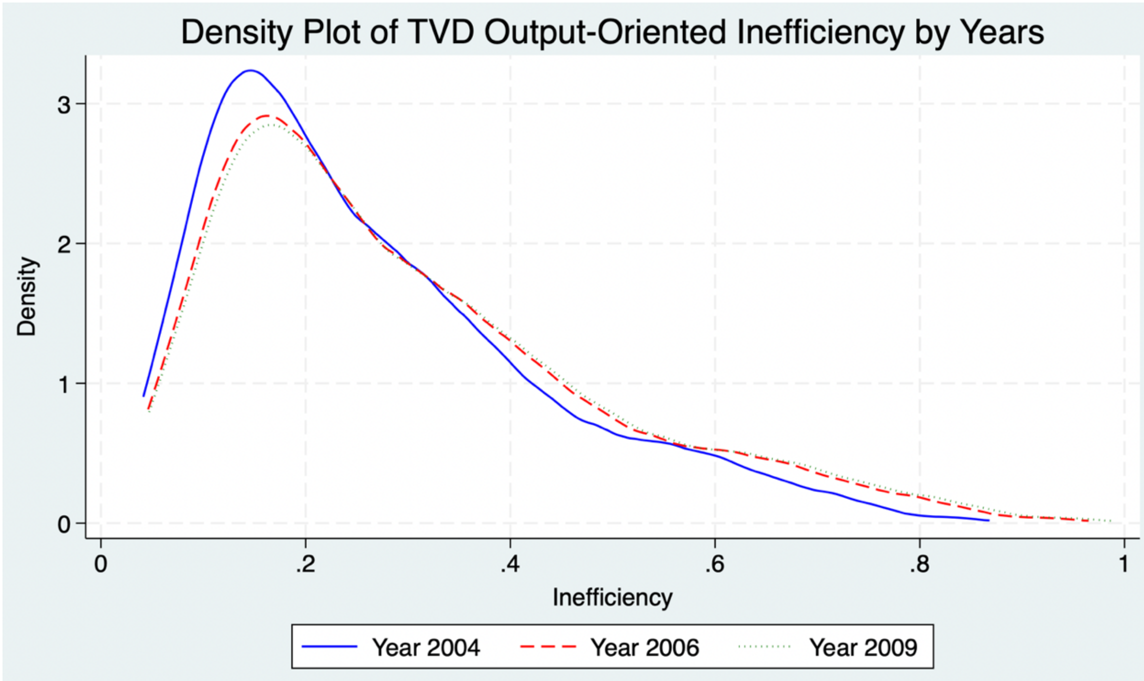
Jondrow, J., Lovell, C. K., Materov, I. S., & Schmidt, P. (1982). On the estimation of technical inefficiency in the stochastic frontier production function model. Journal of econometrics, 19(2-3), 233-238.

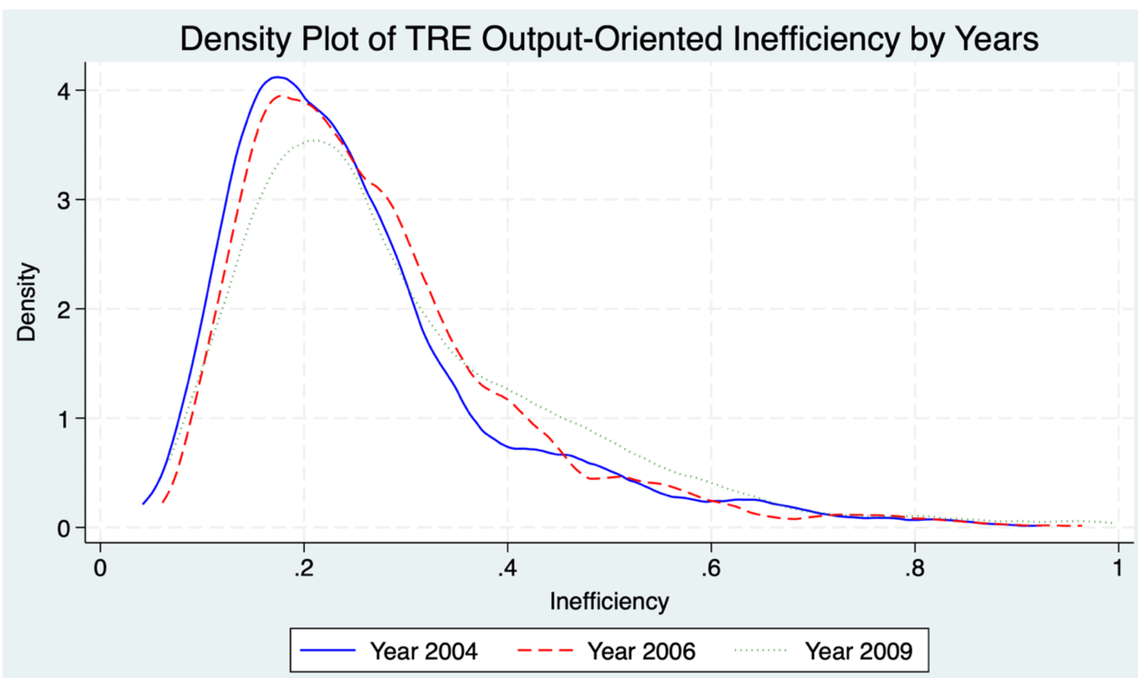
Caudill, S. B., Ford, J. M., & Gropper, D. M. (1995). Frontier estimation and firm-specific inefficiency measures in the presence of heteroscedasticity. Journal of Business & Economic Statistics, 13(1), 105-111.

Appendix 1.1 Output-oriented inefficiency density plots from different models






Density Plot of TVD Output-Oriented Inefficiency by Years
Density
0
1
2
3
0
.2
.4
.6
.8
1
Inefficiency
Year 2004
Year 2006
Year 2009


Density Plot of TRE Output-Oriented Inefficiency by Years
Density
0
1
2
3
4
0
.2
.4
.6
.8
1
Inefficiency
Year 2004
Year 2006
Year 2009

Appendix 1.2 Two-sample t test with equal variances (2004 and 2006)

Group	Obs	Mean	Std. err.	Std. dev.	[95% conf. interval]	
2004	808	.2138843	.0023868	.0678456	.2091993	.2185694
2006	808	.2050981	.0020137	.0572401	.2011454	.2090508
Combined	1616	.2094912	.0015647	.0629015	.2064221	.2125603
diff		.0087862	.0031228		.0026611	.0149114

diff = mean(2004) - mean(2006) t = 2.8136

H0: diff = 0 Degrees of freedom = 1614

Ha: diff < 0	Ha: diff != 0	Ha: diff > 0
Pr(T < t) = 0.9975	Pr(|T| > |t|) = 0.0050	Pr(T > t) = 0.0025

Appendix 1.3 Two-sample t test with equal variances (2004 and 2009)

Group	Obs	Mean	Std. err.	Std. dev.	[95% conf. interval]	
2004	808	.2138843	.0023868	.0678456	.2091993	.2185694
2009	808	.2154825	.0024327	.0691499	.2107074	.2202577
Combined	1616	.2146834	.0017036	.0684843	.2113419	.218025
diff		-.0015982	.003408		-.0082829	.0050865

diff = mean(2004) - mean(2009)		t = -0.4690
H0: diff = 0		Degrees of freedom = 1614
Ha: diff < 0	Ha: diff != 0	Ha: diff > 0
Pr(T < t) = 0.3196	Pr(|T| > |t|) = 0.6392	Pr(T > t) = 0.6804

Appendix 1.4 Two-sample t test with equal variances (2006 and 2009)

Group	Obs	Mean	Std. err.	Std. dev.	[95% conf. interval]	
2006	808	.2050981	.0020137	.0572401	.2011454	.2090508
2009	808	.2154825	.0024327	.0691499	.2107074	.2202577
Combined	1616	.2102903	.0015838	.0636675	.2071838	.2133968
diff		-.0103844	.003158		-.0165787	-.0041902

diff = mean(2006) - mean(2009)		t = -3.2883
H0: diff = 0		Degrees of freedom = 1614
Ha: diff < 0	Ha: diff != 0	Ha: diff > 0
Pr(T < t) = 0.0005	Pr(|T| > |t|) = 0.0010	Pr(T > t) = 0.9995

Appendix 1.5 Inefficiency Determinants

Table 7. Marginal Effect of determinants (MODEL EXPANDED)

	Z1	Z2	Z3	Z4	Z5	Z6	Z7	Z8	Z9
Z1	+***								
Z1,Z3	+*		+**						
Z1,Z4	+***			+***					
Z1,Z5	+				+**				
Z1,Z6	+***					+***			
Z1,Z7	-						-		
Z1,Z8	+***							-	
Z1,Z9	+***								+*
Z1,Z2,Z3	+	+	+**						
Z1,Z2,Z4	+***	+		+***					
Z1,Z2,Z6	+***	+				+***			
Z1,Z2,Z8	+***	+						+	
Z1,Z2,Z9	+***	+							+*
Z1,Z3,Z4	+***		+	+***					
Z1,Z3,Z7	-		+				-		
Z1,Z3,Z8	+		+**					-	
Z1,Z3,Z9	+***		+						+
Z1,Z4,Z5	+***			+***	+				
Z1,Z4,Z6	+***			+***		+***			
Z1,Z4,Z7	+***			+***			-		
Z1,Z4,Z8	+***			+***				-	
Z1,Z4,Z9	+***			+**					+
Z1,Z5,Z6	+*				+*	+***			
Z1,Z5,Z7	-				-		-		
Z1,Z5,Z9	+***				+				+
Z1,Z6,Z7	-					+	-		
Z1,Z6,Z8	+***					+***		+*	

Z1,Z6,Z9	+***					+***			+
Z1,Z7,Z8	+***						+	+	
Z1,Z7,Z9	+***						+		+
Z1,Z8,Z9	+***							+	+
Z1,Z2,Z3,Z4	+	+	+	+					
Z1,Z2,Z3,Z6	+*	+	+			+**			
Z1,Z2,Z3,Z7	+	+	+**				+		
Z1,Z2,Z4,Z8	+***	+		+***				+	
Z1,Z2,Z4,Z9	+***	+		+***					+
Z1,Z2,Z5,Z6	+	+			+***	+			
Z1,Z2,Z6,Z8	+***	+				+***		+**	
Z1,Z2,Z6,Z9	+***	+*				+***			+***
Z1,Z2,Z7,Z8	-	+					-	-	
Z1,Z2,Z7,Z9	-	+					-		-*
Z1,Z2,Z8,Z9									

$* : \rho < 0.1$ $** : \rho < 0.05$ $*** : \rho < 0.01$