TOTAL FACTOR PRODUCTIVITY IN KAZAKHSTAN'S SMALL AND MEDIUM ENTERPRISES (SME)

BY

MIZHGONA NURALIYEVA

THESIS

Submitted in partial fulfillment of the requirements

for the degree of Master of Science in Finance

in the Graduate School of Business

Nazarbayev University, 2024

Astana, Kazakhstan

Advisor: Dr. David Robert De Remer

This work was supported by the Nazarbayev University Faculty Development Competitive Research Grant Program (Funder Project Reference: 201223FD8834).

Special thanks to Professor Aigerim Yergabulova for providing data.

Table of Contents:

Abstract

This study investigates the key drivers of Total Factor Productivity (TFP) in small and medium enterprises (SMEs) in Kazakhstan, focusing on a combination of firm-level data (2011–2023) and advanced econometric methods. TFP, which is a measure of efficiency and innovation, is analyzed using Olley-Pakes (OP), Levinsohn-Petrin (LP), Ackerberg-Caves-Frazer (ACF), and kernel-based learning techniques (KLT) in this paper.

Exchange rate shocks serve as instrumental variables to address endogeneity concerns and explore the impact of materials, capital wage bill, energy costs, firm size, and other inputs on productivity. The results reveal that firm size and wage bills positively influence TFP under specific conditions, whereas inappropriate use of energy and capital can have negative influence on productivity.

The use of each method helps to analyze the estimation approaches properly, providing a clear understanding of TFP dynamics. To be able to proceed with practical policy recommendations, counterfactual simulations of wage bill reductions were created.

This research contributes to the existing literature of Zarina Adilkhanova by focusing on advanced methodologies to estimate TFP and providing practical policy recommendations to enhance SME productivity in Kazakhstan.

I. Literature review

Total Factor Productivity (TFP) is a critical measure of efficiency and innovation that reflects how well businesses combine labor, capital, and technology to produce goods and services. It is a key measure in Kazakhstan's economy as it captures economic potential and productivity growth. Even though small and medium enterprises constitute a significant portion of Kazakhstan's economy, there are some obstacles that limit their full capacity. Therefore, this paper explores literature, theoretical frameworks, and the challenges in analyzing TFP in Kazakhstan's SME.

The Cobb-Douglas production function provides analysis on the basic relationship between the output and its inputs. However, the U.S. Bureau of Statistics (2024) notes that the Cobb-Douglas approach alone does not always capture economic complexities, whereas TFP deals with that, highlighting the industries' operational efficiency. Research of Kim and Park's (2017) for the Asian Development Bank demonstrates TFP's relevance in analyzing middle-income economies, making it highly applicable to Kazakhstan.

Zarina Adilkhanova's (2022) paper provides firm-level evidence from Kazakhstan (2009–2017) and mentions decreasing TFP in several sectors because of the monopolization and lack of competition. Large firms constitute a large proportion in the key industries such as agriculture and manufacturing, which limits efficiency and hinders the growth of SMEs. In addition, large firms always benefit from government support, while SMEs usually struggle. It shows an issue of unequal distribution of subsidies.

Moreover, there are regional disparities in productivity that are significant in Kazakhstan. For example, Atyrau and Mangistau contribute to the higher TFP thanks to the resource extraction activities. Kostanay and Kyzylorda have declining TFP, driven by reduced labor productivity in services and manufacturing (Iootty De Paiva Dias et al., 2023). Regions like Almaty and Astana always have comparatively higher productivity due to better infrastructure and market access, whereas rural areas lag behind due to limited industrial diversification and access to capital (Lall et al., 2023).

Besides, studies such as Garicano et al. (2016) highlight the inefficiency of the policies in France that hindered SMEs from their full potential development. Kazakhstan also has an issue of ineffective rule such as high regulatory costs and limited access to finance, preventing SMEs from achieving higher TFP. The fact that only 19% of SMEs in Kazakhstan have access to credit, and high collateral requirements also emphasizes the reason for slow SME development.

Moreover, SMEs suffer from low levels of innovation, where only 8.1% engage in R&D activities (OECD, 2018). The lack of innovation are caused by the limited financial support for SMEs and a focus on short-term development rather than long-run perspective.

Kazakhstan has implemented programs such as the Business Road Map 2020 and the Damu Entrepreneurship Development Fund to support SMEs. However, while focusing on financial aid, these programs are useless in terms of advisory services or management training, which are also crucial for SME development (OECD, 2018). Although some programs such as "SME Top Management Training and Business Connections" showed significant contributions, their scope of influence remained limited, requiring more comprehensive approaches to foster TFP through innovation and industrialization (El Refae, Eletter & Kaba, 2022).

The World Bank Group in the article about Kazakhstan's stagnation (2019) discusses that there are some macroeconomic influences as well. Economic crises, such as the 2014 oil price shock, have disrupted Kazakhstan's productivity growth. Between 2010 and 2016, potential output declined due to fluctuating oil prices and inconsistent macroeconomic policies.

Besides, Kazakhstan's FDI is heavily concentrated in natural resource sectors, limiting productivity spillovers to SMEs. Weak integration into global value chains and low R&D investments further reduce opportunities for innovation, hence contributions to the increase in productivity (Adilkhanova, 2022).

While several studies provide valuable insights into TFP trends in Kazakhstan, significant gaps remain. For example, most research focuses on large firms or sector-wide performance, with limited analysis of SME-specific productivity drivers. Few studies address regional differences in productivity and the factors contributing to these disparities. Existing research often highlights correlations between inputs (capital, labor, subsidies) and TFP without addressing reverse causality (Adilkhanova, 2022).

This study aims to fill these gaps by analyzing how the causal factors impact SME productivity, identifying factors contributing to higher or lower TFP across Kazakhstan's regions and provide policy recommendations with practical solutions for reducing regulatory burdens, and supporting innovation.

I. Methodology

This study examines the relationship between Total Factor Productivity (TFP) and import exposure in small and medium firms in Kazakhstan, combining two datasets. The first one is a monthly trade dataset (2011-2023), while the second one is an annual firm-level dataset. The goal is to use exchange rate shocks as instrumental variables to address the endogeneity and observe its influence on the firm productivity.

The first one describes trade of the firms in Kazakhstan from 2011 to 2023 and has the following variables: MES-month, GOD-year, TE(TERRITORY CODE), hs_code, tr_c (trading country), shipc_c (shipping country), org_c (country of origin),dest_c (destination country),mode_tr (transpontation mode), mode_intr,incoterms,weight, other_unit (other measuring units according to codes), quantity, invoice_value_kzt, stat_value_kzt, stat_value_usd, classifier_goods,imex (director of trade),re_imex,pr, NID,TE_import, carrier_country . "IMEX==2" command was used to derive the firms that only imported goods and from that data, import exposure was derived by being equalized to stat_value_kzt. To sort the firms that were highly exposed to the import shocks, ranking was done: 20% of the firms with the highest exposure were put on top. Import exposure serves as a key variable in exploring the broader effect of imports on firm-level productivity.

The second dataset is descriptive, and provides information about irm id, year, sales, employment, materials, fuel_costs, energy_costs,wagebill, total_costs, gross_profits, capital, own, sector, region code, economic activity code, ownership, legal_form of the firms in Kazakhstan. The exchange rate shock was defined as the ratio between *stat_value_kzt* and the *stat_value_usd*. This newly created variable can help to observe the fluctuations in exchange rate shocks, which in turn affect import exposure, hence TFP. Exchange rate shocks are then categorized into substantial shocks, defined as changes greater than 5% relative to the previous month. These exchange rate shocks were used as instrumental variables because they have correlation with import exposure, but act as an exogenous variable to firm-specific factors, which makes it a valid instrument in estimating TFP.

The monthly and annual datasets were merged using common variables, which are "id" and "year". First, monthly import exposure for each firm was annualized. Then, the second dataset considering the number of unique products exported by each firm was created as another influencer of the TFP. Then these two datasets, later the firm-level dataset were all combined. TFP is estimated using a Cobb-Douglas production function, the natural logarithm of sales was regressed against natural logarithm of capital and materials. Taking log forms ensures linearity. Unobserved productivity shocks were predicted using 1) Levinsohn and Petrin (LP), which employs material inputs as a proxy for unobserved productivity shocks; 2) Olley and Pakes (OP), which models productivity shocks as a function of capital investment; 3)Ackerberg, Caves, and Frazer (ACF), which resolves collinearity issues in traditional production function estimation; and 4) Kernel-Locally Weighted Transformation (KLT), which utilizes kernel regression to predict productivity shocks and improve flexibility in estimation. OP addresses simultaneity and selection bias, LP addresses unobserved productivity shocks, KLT captures entry and exit, while the ACF addresses timing limitations. Therefore, these approaches were used in the paper (Beveren, 2012). Then IV regressions were implemented fixing sector-year effects to control for sector-specific and temporal heterogeneity and clustered by firm identifiers to correct for heteroscedasticity and autocorrelation. Moreover, a hypothetical policy scenario, where the wage bill was reduced by a 10% reduction in wage bills was conducted to analyze its effects on TFP distributions. Diagnostic tests, including the Kleibergen-Paap LM and Cragg-Donald Wald F statistics were used to check if the instruments are strong enough. Observing different metrics using F-stat helps to see the key drivers of TFP for SME in Kazakhstan. In sum, the paper provides robust analysis, which addresses endogeneity, structural heterogeneity that were missing in previous studies.

III.Results

Number of Clusters(id)	781		#of obs	19029

Total centered SS	36278.17268		F(1,780)	72.86
Total uncentered SS	35537.14546		Prob>F	0.000
Residual SS	35537.14546		Centered R2	0.0204
			Uncentered R2	0.0204
			Root MSE	1.37

tfp_lp	coefficient	Robust st.	t	P>[t]	conf.intv	
firm_size	1.712808	0.200663	8.54	0.00	1.318904	2.106711

Table 1

This regression analyzes the relationship between TFP estimated through the LP method and firm size, using exchange rate shocks as an instrumental variable to address potential endogeneity. Coefficient for firm_size is 1.712808, which is statistically significant. It indicates that the firm size positively impacts productivity, specifically, a one-unit increase in firm size increases TFP for 1.71 units. This large and positive coefficient captures that large firms have a better access to resources and advanced their economies of scale.The confidence interval that does not cross zero also implies the robustness between TFP and firm_size. The underidentification test in this analysis strongly rejects the null hypothesis that the exchange rate shock is irrelevant. It means that the variables serve as a valid instrument for firm size. The weak identification test exceeds all critical values, which also shows that the instrument is not weak and the estimates reliable. The R^2 equals 0.0204. Even though the number is low, it is common for the productivity to have low Rsq because there are many unobservable factors that influence TFP. The overall F-stat is statistically significant meaning that the changes in TFP can be explained by the mentioned variables.

Number of Clusters(id)	722		#of obs	17792
Total centered SS	32952.29288		F(6,721)	39.32
Total uncentered SS	32952.29288		Prob>F	0.000
Residual SS	28438.93212		Centered R2	0.1370
			Uncentered R2	0.1370
			Root MSE	1.268

tfp_lp	coefficient	Robust st.	t	P>[t]	95%	conf.int
firm_size	1.6439	0.2329	7.06	0.00	1.18	2.101
fuel_costs	-4.26e-08	2.18e-08	-1.96	0.051	-8.54e-08	1.13e-10
wagebill	-6.19e-09	1.67e-08	-0.37	0.712	-3.91e-08	2.67e-08
ownership	0.0389	0.0097	3.99	0.00	0.0198	0.058

total_costs	1.29e-09	1.34e-09	0.96	0.336	-1.35e-09	3.94e-09
Gross profit	-4.53e-11	6.97e-10	-0.07	0.0948	-1.41e-09	1.32e-09

Table 2

The regression is also about the relationship between TFP and firm size including other control variables such as fuel costs, wage bill, ownership, total costs, and gross profits. The coefficient for firm size is 1.643899, indicating that a one-unit increase in firm size is associated with a 1.64-unit increase in TFP. This relationship is statistically significant at the 1% level. Fuel costs have a negative coefficient, which is significant (0.051). This result means that higher fuel costs are slightly associated with lower TFP. It can be associated with the rise of operational costs and the inefficiencies in energy use. Due to the rise in fuel costs, energy-intensive firms can overcome substantial and disproportionate productivity losses. The coefficient for wage bill is negative and not statistically significant (p = 0.712), meaning that the influence on TFP is not consistent. It also highlights the possible need for the firms to optimize their use of labor as the increase in wages does not guarantee the rise in efficiency. This indicates that labor costs do not have a clear or consistent impact on TFP in this specification. Ownership has a positive coefficient of 0.0380915 and is statistically significant at the 1% level. It indicates the importance of corporate structure as the firms with better regulations may achieve higher TFP results thanks to their management. Total costs have a positive and not statistically significant coefficient, meaning that the variable does not affect TFP in the set model. One explanation is because it includes both productive and unproductive expenditures. Gross profits have a negative and not statistically significant (p = 0.948) coefficient, which means that profitability is not enough to estimate the TFP. The underidentification test (p = 0.000) confirms the validity of the exchange rate shock, because the null hypothesis of underidentification is strongly rejected.

Both regressions confirm a strong, positive, and statistically significant relationship between firm size and TFP. Even though the instrument is valid and the relationships are robust in both cases , the coefficients in the second regression were slightly lower because additional control variables absorbed the explanatory power of firm size.The residual sum of squares (SS) is lower in the second regression, which means a better explanation of TFP variance.

Number of Clusters(id)	781	#of obs	19092
Total centered SS	3658.27624	F(1,780)	13.48
Total uncentered SS	36578.27624	Prob>F	0.000
Residual SS	223966.7653	Centered R2	-5.1229
		Uncentered R2	-5.1229
		Root MSE	3.434

tfp_lp	coefficient	Robust st.	t	P>[t]	conf.intv	
Wage bill	2.84e-07	7.73e-08	3.67	0.00	1.33e-07	4.35e-07

Table 3

This analysis explores the relationship between TFP and the wage bill using the LP method. The coefficient for wage bill is 2.84e−07 and statistically significant. This suggests that higher labor costs are associated with higher productivity, potentially reflecting improved worker quality or increased employee satisfaction, which is counterfactual to the previous case. The underidentification test rejects the null hypothesis, which confirms that the instrument is strongly correlated with wage bill. Furthermore, the weak identification test, with a Cragg-Donald Wald F-statistic of 13.48 is higher than the threshold meaning that the instruments are not weak.

Number of Clusters(id)	667	#of obs	16546
Total centered SS	31092.6188	F(7,666)	3.44
Total uncentered SS	310926188	Prob>F	0.0013
Residual SS	196793.8301	Centered R2	-5.3293
		Uncentered R2	-5.3293
		Root MSE	3.46

tfp_lp	coefficient	Robust st.	t	P>[t]	95%	conf.int
wagebill	8.34e-07	4.27e-07	1.95	0.0052	-5.78e-09	1.67e-06
Firm size	0.6139	0.19	3.14	0.002	0.2298	0.998

Energy costs	-3.85e-07	2.94e-07	-1.31	0.19	-9.62e-07	1.92e-07
Fuel costs	-4.62e-07	2.63e-07	-1.75	0.08	-9.79e-07	5.55e-08
ownership	0.039	0.0223	1.73	0.084	-0.00528	0.0836
Total costs	-4.41e-08	2.33e-08	-1.89	0.059	-8.99e-08	1.66e-09

Table 4

The results of this regression also analyzes the relationship between TFP and wage bill including other control variables. The coefficient for wage bill is positive, but not statistically significant, meaning that wage bill does not influence TFP. Firm size has a coefficient of 0.6139676 and is statistically significant. This result mentioned the previous conclusions that larger firms can lead to higher TFP due to the economies of scales, though the wide confidence interval highlights the variability of these aspects. The rest of the variables such as ownership, gross profits, total costs, energy, and fuel costs have insignificant negative relationships with TFP in this model, suggesting that they do not explain the variance in TFP. Overall, the F-stat is very low, questioning the strength of the instrument. However, it is still valid as Hansen J-test fails to reject the null hypothesis.

Number of Clusters(id)	697	#of obs	17411
Total centered SS	33403.7675	F(1,696)	14.21
Total uncentered SS	33403.7675	Prob>F	0.000

Residual SS	136694.6835	Centered R2	-3.092
		Uncentered R2	-3.092
		Root MSE	2.81

tfp_lp	coefficient	Robust st.	t	P>[t]	conf.intv	
Energy costs	1.04e-06	2.75e-07	3.77	0.00	4.98e-07	1.58e-06

Table 5

Number of Clusters(id)	667	#of obs	16546
Total centered SS	31092.6188	F(7,666)	0.00
Total uncentered SS	31092.6188	Prob>F	1.00
Residual SS	1242232647	Centered R2	-4.0e+04
		Uncentered R2	-4.0e+04
		Root MSE	274.9

tfp_lp	coefficient	Robust st	t	P>[t]	95%conf	interv
Energy cost	0.001	0.004	0.03	0.976	-0.008	0.008

Firm size	-75.1258	2536.641	-0.03	0.976	-5055.902	4905.65
Fuel cost	2.77e-06	0.00	0.03	0.977	-0.00187	0.00019
wagebill	-0.00	0.00	-0.03	0.976	-0.0013	0.0013
ownership	-2.56	87.207	-0.03	0.977	-173.7988	168.6714
Total cost	-3.54e-07	0.00	-0.03	0.976	-0.00	0.00
Gross profit	9.71e-07	0.00	0.03	0.976	-0.00	0.00

Table 6

The tables compare two instrumental variable regressions exploring the relationship between energy costs and TFP, estimated through the Levinsohn-Petrin LP method. The first regression focuses solely on energy costs as the explanatory variable, while the second includes other additional controls.

In the first regression, the coefficient for energy costs is 1.04e-06, highlighting a positive impact of energy costs on TFP. The relationship with a p-value of 0.000 is statistically significant. It means that an increase in energy costs can be responsible for the higher TFP, potentially due to energy-intensive operations driving productivity. However, it can also be the case that the energy costs were just counted in the production without giving efficient outcomes. The underidentification test with the p = 0.0015 confirms the validity of the instrument, and the Wald F statistic equals 51.311, which exceeds all critical thresholds, ensuring a strong and reliable instrument.

In the second regression, the coefficient for energy costs becomes statistically insignificant. Even the 95% confidence interval that ranges from -0.00008066 to 0.0008255, indicates a lack of robustness, which in turn, suggests that the observation from the first regression about positive impact of energy on TFP was absorbed by other factors such as firm size, labor costs, or operational practices. Notably, all the mentioned variables are statistically insignificant and both underidentification (p-value is 0.9760) and Wald F-statistic (0.001) confirm the weakness of the instrument being used.

This comparison proposes that energy costs alone can not independently influence productivity because their impact can be intertwined with other firm-level variables. In conclusion, while the first regression identifies a significant positive relationship between energy costs and TFP, the second regression fails to confirm this relationship, emphasizing the importance of considering additional firm-level variables and addressing potential weaknesses in the instrument. These findings indicate the importance of the appropriate modeling to avoid causal effects due to the complex nature of TFP determinants.

Source	SS	df	MS
Model	36,628.4603	3	12,209.4868
Residual	546.999794	15 002	0.036461791
Total	37,175.4601	15 005	2.47753816

#of obs	15006
F(3,15002)	9999

Prob>F	0.000
Rsq	0.9853
Adjusted Rsq	0.9853
Root MSE	0,191

tfp_lp	coefficient	Robust st	t	P>[t]	95%conf	interv
logsales	coefficient	st.error	t	P>[t]	95%conf	interval
logmaterials	0.3383	0.001	278.86	0.00	0.3359	0.34069
Log capital	0.3635	0.001	316.68	0.00	0.3612	0.3657
omega	0.996	0.02	495.18	0.00	0.992	1.00
cons	5.49	0.012	446.99	0.00	5.47	5.51

Table 7

This regression evaluates the relationship between firm sales and capital, materials , productivity shocks (omega_predicted) using the Ackerberg, Caves, and Frazer (ACF) method. Due to the high R-squared of 0.9853, the model demonstrates that 98.53% of the variation in firm sales is explained by the included variables.

The coefficient for “logmaterials” is 0.3383205, which is positive and statistically significant. It means that a 1% increase in materials is associated with an approximately 0.34% increase in sales. However, it suggests diminishing returns to material inputs, which could arise from inefficiencies in material utilization or production processes. The coefficient for “logcapital” is 0.3635323 and statistically significant, meaning that a 1% increase in capital is associated with a 0.36% increase in sales. It is also associated with a diminishing returns to capital, possibly due to underutilization of machinery or inefficient investment in productive assets. The comparison of these two coefficients shows that the capital is slightly more productive than materials in generating output. The diminishing results in both cases emphasize the need for optimum resource allocation to achieve higher output results. The coefficient for omega_predicted is 0.9966304 and statistically significant. This result mentions the huge role of the variable as it acts as almost a perfect predictor for the TFP after accounting for materials and capital. It reflects the fact that there are a lot of unobserved factors except materials and capital that have an impact on output. It can include the influence of firm-specific managerial practices, technology adoption, or innovation capabilities and suggest the development of technologies with a focus on operational efficiency. The high R-squared with significant coefficients highlights that this method effectively captures the relationship between inputs and sales.

Number of Clusters(id)	781	#of obs	19026
Total centered SS	8965	F(1,780)	16.57
Total uncentered SS	8965	Prob>F	0.000
Residual SS	8991	Centered R2	0.0204
		Uncentered R2	0.0204
		Root MSE	0.069

tfp_acf	coefficient	Robust st.	t	P>[t]	conf.intv	
firm_size	0.39	0.0958	4.07	0.00	0.2019	0.578

Table 8

Number of Clusters(id)	722	#of obs	17789
Total centered SS	8039.066	F(6,721)	5.27
Total uncentered SS	8039.066	Prob>F	0.000
Residual SS	6848.61	Centered R2	0.1481
		Uncentered R2	0.1481
		Root MSE	0.6224

tfp_acf	coefficient	Robust st	t	P>[t]	95%conf	interv
Firm size	0.3369	0.11	3.04	0.02	0.1194	0.5544
Fuel cost	-1.53e-08	7.42e-09	-2.06	0.04	-2.99e-08	-7.34e-10
wagebill	4.88e-09	5.6e-09	0.87	0.384	-6.11e-09	1.59e-08
ownership	0.017	0.0043	4.09	0.00	0.0092	0.2633
Total cost	-3.64e-11	4.38e-10	-0.08	0.934	-8.96e-10	8.24e-10

Gross profit	5.53e-10	2.3e-10	2.4	0.017	1.01e-10	1.01e-09

Table 9

The tables demonstrate the relationship between Total Factor Productivity and firm size, including other explanatory variables with the ACF method. The use of exchange_rate_shock as an instrumental variable ensures the mitigation of endogeneity concerns in the analysis.

The coefficient for firm size is consistently positive and statistically significant in both regressions, though slightly lower in the second regression (0.3369592 and 0.3900732) because of the absorption of the explanatory power that was discussed with the methods before. The observation with more control variables provides deeper insights: for example,fuel costs are negative and significant, highlighting inefficiencies in energy usage. Ownership is positive and significant, emphasizing the role of governance and strategic decision-making, whereas gross profits are positive and significant, suggesting reinvestment into productivity-enhancing measures. The wage bill and total costs are insignificant, describing no or limited direct impact of the production. The diagnostic tests confirm the validity and strength of the instrument in the first regression, being not sufficient enough in the second.

In this ACF method, both wage bill and energy costs are presenting very low F-stats (9.40 and 8.71) even in their reduced forms, while it was not the case in LP method. There can be several reasons for this difference between the methods. For example, the ACF method dilutes the observable variation to deal with the potential simultaneity bias.It also models the inputs as the functions of unobserved shocks, which allows for the less variation in wage bill and energy costs compared to the LP method, hence additional variations become even harder to interpret. Also, labor inputs can be connected with the wage bill in this method, which can already be captured by the shocks, while energy costs might be more volatile to the external market shocks, which have less connection with the exchange rate shocks. In contrast, the LP method uses intermediate inputs such as materials to proxy for productivity shocks that allows for more variation in wage bills and energy costs. Therefore, it might be the case that the F-statistic for the ACF approach is lower making it less effective in capturing exogenous variation of the control variables such as wage bill and energy costs.

Number of Clusters(id)	787	#of obs	19067
Total centered SS	36349	F(1,786)	72.86
Total uncentered SS	36349	Prob>F	0.000
Residual SS	36033	Centered R2	0.0087
		Uncentered R2	0.0087
		Root MSE	1.378

tfp_op	coefficient	Robust st.	t	P>[t]	conf.intv	
firm_size	1.72	0.2022	8.54	0.00	1.329	2.12

Table 10

Number of Clusters(id)	726	#of obs	17813
Total centered SS	33001.442	F(6,725)	38.79

Total uncentered SS	33001.442	Prob>F	0.00
Residual SS	28794.4836	Centered R2	0.1275
		Uncentered R2	0.1275
		Root MSE	1.275

tfp_lp	coefficient	Robust st	t	P>[t]	95%conf	interv
Firm size	1.6556	0.233	7.08	0.00	1.197	2.1147
Fuel cost	-4.31e-08	2.20e-08	-1.96	0.05	-8.62e-08	4.91e-11
wagebill	-6.47e-09	1.68e-08	-0.38	0.701	-3.95e-08	-2.66e-08
ownership	0.039	0.0098	3.98	0.00	0.0198	0.058
Total cost	1.29e-09	1.35e-09	0.96	0.339	-1.36e-09	3.94e-09
Gross profit	-3.68e-11	6.96e-10	-0.05	0.958	-1.40e-09	1.33e-09

Table 11

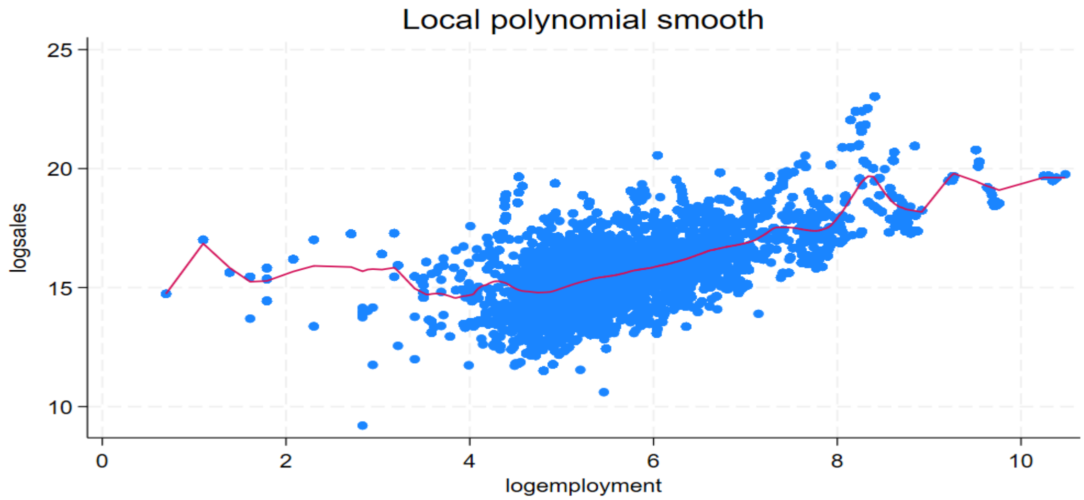
The tables present the relationship between firm size and total factor productivity (TFP) using the Olley-Pakes methodology, exchange rate shocks acting as instrumental variables. The first regression (reduced form) examines firm size as the sole variable, while the second regression includes additional control variables mentioned in the previous approaches.

In the reduced form regression, the coefficient for firm size is 1.726056, meaning that 1 unit increase in the firm size results in the 1.73 unit increase of sales. The p-value is zero, highlighting the importance of the observation. The underidentification test confirms the validity of the instrument with having statistically significant p-values. Besides, the F statistic of 38.79 is far above the critical value mentioning the validity and the reliability of the instrument.

In the full form regression, the coefficient for firm size decreases to 1.655687, but the relationship is statistically significant. The confidence interval narrows slightly, suggesting that the additional controls do not fundamentally alter the positive impact of firm size on TFP but reduce its magnitude slightly. The Wald F statistic of 72.86 again confirms the instrument's validity and strength.

The coefficient for firm size decreases slightly in the full form regression compared to the reduced form. This suggests that some of the explanatory power attributed to firm size in the reduced model is absorbed by the control variables in the full model.Firm size remains statistically significant in both models, indicating that its impact on TFP is robust to the inclusion of additional control variables. Only the ownership has a statistically significant p-value with the coefficient of 0.038 indicating modest impact on output, whereas other variables are insignificant. This indicates that even though these variables can have effect on the TFP separately, in this analysis, their compounded effect is not strong enough. Both regressions show strong underidentification and weak identification test results, indicating that the exchange rate shock is a valid and powerful instrument for firm size.

The results for wage bill and energy costs have similar conclusions as the LP method. They both present high F-stat in their reduced forms (13.52 and 14.21), being ineffective in the full regression analysis (3.42 and 0.00). The strong relationship between wage bill and TFP in the reduced form underscores the importance of labor cost efficiency. Higher labor costs being associated with the importance of investing in the labor and energy costs being crucial determinants of TFP in energy-reliant sectors.



Graph 1

This local polynomial smooth plot visualizes the relationship between the logarithm of employment and the logarithm of sales. The red line represents the fitted local polynomial curve, while the blue dots correspond to individual observations. This plot offers insights into the nature of the relationship between employment and sales at different levels of employment.

At very low levels of "logemployment" the fitted line indicates a slightly declining trend, which proposes that employment increase will not lead to proportionate increase of the output when the level of the employment is low. This conclusion supports the argument that labor resources are used inefficiently, while other inputs (for example, "capital" and "labor") are utilized better. In the middle of the graph, a constant upward trend is captured, which means possible expansion of the workforce, hence increase in sales. It also confirms the expectation about employment being a crucial factor in the output estimation. At higher levels of the variable, the line dips slightly again meaning less contribution to the growth. It can be associated with restrictions from other inputs, decreasing marginal efficiency because of overstaffing (high level of labor) and fixed capacity constraints.

Blue dots spread around the red line indicate the heterogeneity among firms. Different firms may exhibit distinct efficiencies in converting labor into sales. The fitted curve provides a smoothed estimate of the average relationship, but the wide dispersion of individual observations highlights potential firm-level factors such as industry, technology, or management practices that have impact on the relationship of “logemployment” and “logsales”.

Source	SS	df	MS
Model	31049.5606	3	10349.8535
Residual	6125.89949	15 002	0.408338854
Total	37,175.4601	15 005	2.47753816

#of obs	15006
F(3,15002)	25346.24
Prob>F	0.000
Rsq	0.8352
Adjusted Rsq	0.8352
Root MSE	0,63901

logsales	coefficient	st.error	t	P>[t]	95%conf	interval
logcapital	0.322971	0.00387	83.34	0.00	0.3153	0.330567
Logmaterials	0.2179	0.00424	51.33	0.00	0.2096	0.2262633
omega	0.4976	0.005	90.73	0.00	0.486	0.5084239
cons	7.903	0.04861	162.60	0.00	7.808563	7.999126

Table 12

This regression re-estimates the production function using the KLT method, with the logarithm of sales as the dependent variable and the logarithm of capital, the logarithm of materials, and the predicted productivity shocks derived from the KLT estimation as explanatory variables.

The coefficient for “logcapital” is -0.322971, which indicates that a 1% increase in capital is associated with a 0.323% decrease in sales; this result is also statistically significant. In contrast to theoretical expectations, this result presents an anomaly as capital usually positively contributes to TFP. It can be explained by inefficient use of capital, for example, having old machinery, or by over-capitalization. The coefficient for materials is 0.2179404, which means that a 1% increase in material inputs results in a 0.218% increase in sales. Even though this result is also statistically significant, low elasticity means that the use of materials is not proceeded efficiently, requiring better policies for the rational exploitation of this input. The coefficient for predicted productivity shocks is 0.4976725, which means that a 1-unit increase in productivity shocks leads to a 0.498% increase in sales. Similar to the ACF method, the role of the unobserved shocks is statistically significant, underscoring the significance of addressing internal inefficiencies of the firms. Overall, the R-squared value of 0.8352 highlights that most of the changes in “logsales” are explained by the included variables. The F-statistic of 25346.24 also proves that the strong variables were used in the regression.

Number of Clusters(id)	781	#of obs	19026
Total centered SS	11481.54602	F(1,780)	37.13
Total uncentered SS	11481.54602	Prob>F	0.000
Residual SS	11611.27056	Centered R2	-0.0113
		Uncentered R2	-0.0113
		Root MSE	0.7834

tfp_klt	coefficient	Robust st.	t	P>[t]	conf.intv	
firm_size	0.6930364	0.1137	6.09	0.00	0.046977	0.9162

Table 13

Number of Clusters(id)	722	#of obs	17789
Total centered SS	10348.77845	F(6,721)	8.18
Total uncentered SS	10348.77845	Prob>F	0.00

Residual SS	8840.6087	Centered R2	0.1457
		Uncentered R2	0.1457
		Root MSE	0.7071

tfp_klt	coefficient	Robust st	t	P>[t]	95%conf	interv
Firm size	0.6367	0.13	4.9	0.00	0.381	0.89196
Fuel cost	-2.10	1.03e-08	-2.03	0.043	-4.12e-08	-7.04e-10
wagebill	1.45e-09	7.60e-09	0.19	0.849	-1.35e-08	1.64e-08
ownership	0.0230645	0.005162	4.47	0.00	0.0129	0.0331
Total cost	3.18e-10	6.35e-10	0.5	0.617	-9.29e-10	1.56e-09
Gross profit	3.68e-10	3.37e-10	1.09	0.276	-2.94e-10	1.03e-09

Table 14

These tables analyze the results taken from the relationship of firm size and TFP including other control variables with the KLT method. In the first reduced-form regression, the coefficient for firm size is 0.693064 and statistically significant. The underidentification test rejects the null hypothesis that the instrument is irrelevant and the strong weak identification test confirms that the instrument is strong.

In the second full regression, the coefficient for firm size decreases slightly being equal to 0.6366712, but remaining significant, highlighting the robust relationship between the firm size and the output. Fuel costs have a negative, but statistically significant coefficient, which means that the possible inefficiencies in energy utilization cause decrease in the TFP. Ownership has a positive and statistically significant coefficient (0.02), whereas “wagebill”, “total costs”, “gross profits” do not present significant coefficients, which implies that these variables do not have effect on TFP in this particular analysis. However, it should be noted that the F-stat does not ensure the reliability of the estimates, as the instruments are not strong. Overall, these findings mention that larger firms tend to be more productive due to economies of scale. Also, ownership responsible for governance has a positive influence on TFP, whereas inefficient use of energy affects negatively.

Number of Clusters(id)	781	#of obs	19089
Total centered SS	11546.98	F(1,780)	12.30
Total uncentered SS	11546.98	Prob>F	0.0005
Residual SS	38722.59	Centered R2	-2.3535
		Uncentered R2	-2.3535
		Root MSE	1.428

tfp_klt	coefficient	Robust st.	t	P>[t]	conf.intv	
wagebill	1.15e-07	3.28e-08	3.51	0.00	5.07e-08	1.8e-07

Table 15

Number of Clusters(id)	667	#of obs	16543
Total centered SS	9525	F(7,666)	3.49
Total uncentered SS	9525	Prob>F	0.001
Residual SS	34725	Centered R2	-2.6454
		Uncentered R2	-2.6454
		Root MSE	1.454

tfp_klt	coefficient	Robust st	t	P>[t]	95%conf	interv
wagebill	3.37e-07	1.71e-07	1.97	0.049	1.02e-09	6.73e-07
Firm size	0.223	0.07863	2.97	0.003	0.079e051	0.3879
Energy costs	-1.69e-07	1.17e-07	-1.45	0.148	-3.99e-07	6.04e-08
Fuel costs	-1.87e-07	1.05e-07	1.78	0.076	-3.94e-07	1.98e-08
ownership	0.022	0.0094	2.38	0.017	0.0039332	0.041
Total cost	-1.76e-08	9.35e-09	-1.88	0.061	-3.59e-08	7.93e-10
Gross profit	3.79e-09	2.03e-09	1.86	0.063	-2.02e-10	7.78e-09

Table 16

These tables present both reduced and full-form regression analysis, where the relationship of the wage bill with the TFP using KLT method is discussed. In the reduced form regression, "wagebill" has a profound, positive and statistically significant impact on the output. Wald F-statistic test also confirms that the instrument is strong providing the number equal to 12.30. In the full regression specification, the conclusion for the "wage bill" remains the same, meaning that the influence on TFP is retained even after adding other control variables. Notably, the F-test decreased significantly underscoring the fact that some explanatory power of the wage bill was absorbed by other variables mentioned in the regression. Firm size (0.233449 and p = 0.003) and ownership (0.022354, p = 0.017) also have positive and statistically significant coefficients confirming the previous analysis with other methods. However, "energy", "fuel", and "total costs" "gross profits "are not significant, meaning that they do not jointly contribute to TFP.

Number of Clusters(id)	697	#of obs	17408
Total centered SS	10217.38092	F(1,696)	12.24
Total uncentered SS	10217.38092	Prob>F	0.0005
Residual SS	27532.30604	Centered R2	-1.6947
		Uncentered R2	-1.6947
		Root MSE	1.261

tfp_klt	coefficient	Robust st.	t	P>[t]	conf.intv	
Energy costs	4.15e-07	1.18e-07	3.50	0.00	1.82e-07	6.47e-07

Table 17

Number of Clusters(id)	667	#of obs	16543
Total centered SS	9525	F(7, 666)	0.02
Total uncentered SS	9525	Prob>F	1
Residual SS	65227577	Centered R2	-6.8e+03
		Uncentered R2	-6.8e+03
		Root MSE	63

tfp_klt	coefficient	Robust st	t	P>[t]	95%conf	interv
Energy costs	0.00	0.00548	0.05	0.959	-0.001	0.001
Firm size	-16.88	332.0825	-0.05	0.959	-668.93	635.1736
Fuel costs	6.21e-07	0.00	0.05	0.962	-0.00	0.00
wagebill	-4.61e-06	0.00	-0.05	0.959	-0.001	0.000

ownership	-0.57	11.571	-0.05	0.960	-23.2948	22.14802
Total cost	-8.11e-08	1.60e-06	-0.05	0.960	-3.23-06	3.06e-06
Gross profit	2.23e-07	4.32e-06	0.05	0.959	-8.27e-06	8.71e-06

Table 18

These tables present both reduced and full-form regression analysis, where the relationship of the energy costs with the TFP using KLT method is discussed. In the reduced form regression, the coefficient for energy costs is high and statistically significant. The F-statistics for weak identification tests is 12.24, which confirms the validity of exchange rate shock as IV. In the full regression analysis, which includes additional control variables like firm size, fuel costs, wage bill, ownership, total costs, and gross profits, the coefficient for energy costs drops dramatically to 0.0000284, and the statistical significance disappears (p-value = 0.959). The F-statistic less than 10 also highlights the loss of the explanatory power.

```
-> firm_size_category = Large

      Source |       SS           df       MS      Number of obs   =    10,023
-------------+----------------------------------   F(2, 10020)     =   1553.51
       Model |  1630.61314         2  815.30657   Prob > F        =    0.0000
    Residual |  5258.66558    10,020  .524816925   R-squared       =    0.2367
-------------+----------------------------------   Adj R-squared   =    0.2365
       Total |  6889.27872    10,022  .687415558   Root MSE        =    .72444

------------------------------------------------------------------------------
     tfp_klt | Coefficient  Std. err.      t    P>|t|     [95% conf. interval]
-------------+----------------------------------------------------------------
  logcapital |   .0575389   .0055008    10.46   0.000     .0467563    .0683216
logmaterials |   .1871247   .0061809    30.27   0.000      .175009    .1992405
       _cons |   4.206783   .0699406    60.15   0.000     4.069685    4.343881
```

```
-> firm_size_category = Medium

      Source |       SS           df       MS      Number of obs   =     8,869
-------------+----------------------------------   F(2, 8866)      =    119.57
       Model |   165.88561         2  82.9428049   Prob > F        =    0.0000
    Residual |  6149.93572     8,866  .693653928   R-squared       =    0.0263
-------------+----------------------------------   Adj R-squared   =    0.0260
       Total |  6315.82133     8,868  .712203578   Root MSE        =    .83286

------------------------------------------------------------------------------
     tfp_klt | Coefficient  Std. err.      t    P>|t|     [95% conf. interval]
-------------+----------------------------------------------------------------
  logcapital |   -.069861   .0067926   -10.28   0.000     -.083176   -.0565459
logmaterials |   .0938232   .0060694    15.46   0.000     .0819258    .1057206
       _cons |   7.304528   .0756889    96.51   0.000      7.15616    7.452896
------------------------------------------------------------------------------
```

```
-> firm_size_category = Small

      Source |       SS           df       MS      Number of obs   =       207
-------------+----------------------------------   F(2, 204)       =      2.48
       Model |  3.73899028         2  1.86949514   Prob > F        =    0.0863
    Residual |  153.827806       204  .754057872   R-squared       =    0.0237
-------------+----------------------------------   Adj R-squared   =    0.0142
       Total |  157.566796       206   .76488736   Root MSE        =    .86837

------------------------------------------------------------------------------
     tfp_klt | Coefficient  Std. err.      t    P>|t|     [95% conf. interval]
-------------+----------------------------------------------------------------
  logcapital |   .0652094   .0443819     1.47   0.143    -.0222965    .1527154
logmaterials |  -.0736332   .0331293    -2.22   0.027    -.1389528   -.0083135
       _cons |   7.964495   .5076332    15.69   0.000     6.963614    8.965376
```

Table 19

The analysis presents the relationship of output with capital and labor using KLT method for firms with different sizes (small, medium, large), highlighting the importance of the inputs with various firm size categories.

Considering the large firms, “logcapital” coefficient (0.057) is positive and statistically significant, meaning that large firms use capital efficiently. It can be associated with the benefits from the economies of scale. The coefficient for materials (0.187) is also positive and highly significant, which shows that this input explains more variations in TFP than the capital for large firms. Turning to medium firms, the coefficient of capital is negative and statistically significant, suggesting that an increase in capital reduces TFP for medium firms. The result can be associated with several factors such as inefficient allocation of money and outdated technology. Notable, this effect is smaller than in large firms; it can be due to limited economies of scale.

Capital coefficient for small firms is positive but not statistically significant, which means that capital is not a determinant of growth for small firms. The reason can lie on smaller economies of scale and lower investments. The coefficient for materials (-0.073) is negative and significant, implying that the higher material inputs reduce TFP. The result can be associated with poor procurement practices. Importantly, the model explains only 1.4% of the variance in TFP for small firms, meaning that productivity for them is identified through other factors.

Overall, large firms benefit significantly from capital investments, whereas medium firms experience a negative impact, and small firms being unaffected by capital at all. This conclusion draws attention to optimizing capital use for medium and small firms. Materials contribute positively to TFP in large and medium firms but negatively for small firms, which highlights the importance of the policies targeted to better allocation of materials for small firms. The explanatory power of the model decreases with firm size, suggesting that TFP in smaller firms is influenced by unobserved factors, such as labor productivity, technology adoption, or access to finance.

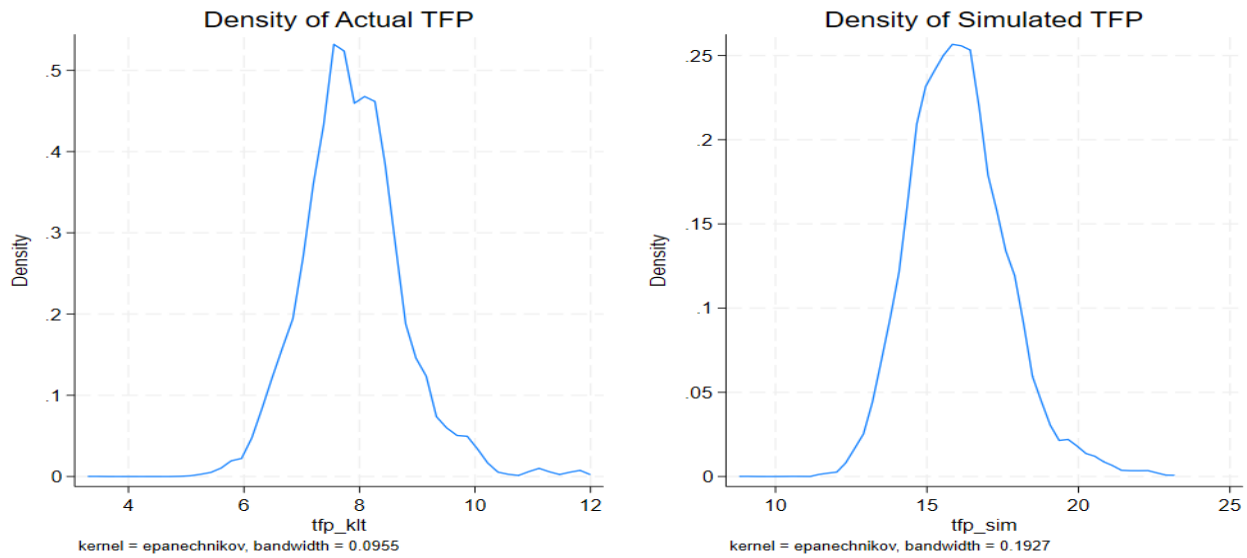
Variable	Obs	Mean	St.Dv	Min	Max

tfp_klt	19 099	7.88	0.8614	3.389	11.901
tfp_sim	19 099	16.04	1.59	8.988	22.99

Table 20

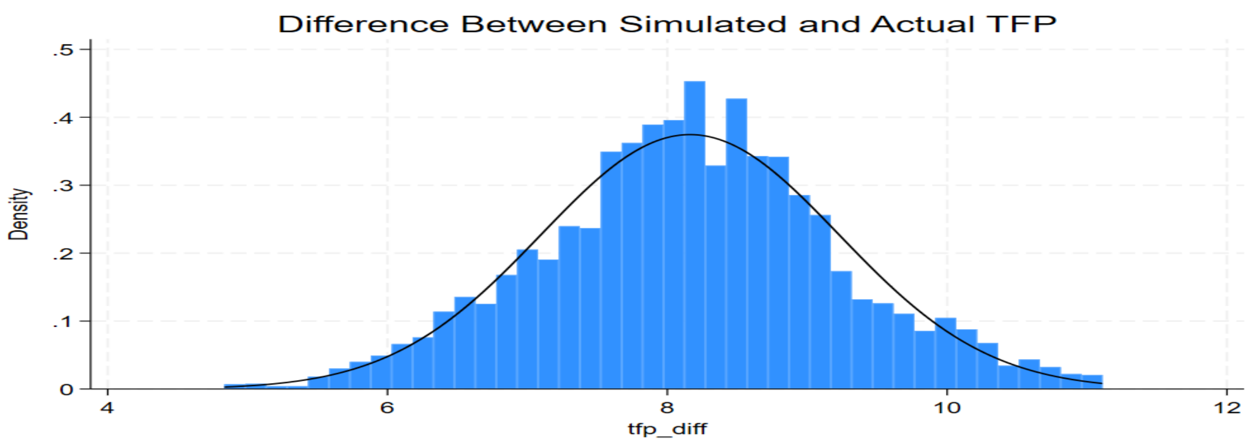
The table compares the actual TFP derived from the KLT approach and simulated one, which was derived by adjusting the capital for 0.1 logarithmic terms, which can indicate correction for biases or measurement errors. This change leads to significant increases in output. Additionally, the higher mean and maximum values mention that increasing capital enhances productivity across firms, especially for those already operating efficiently. High standard deviation means that this change is different for the firms, suggesting that ones with better resource allocation will take more advantage from this change.

The subsequent analysis was done using KLT method as it captures non-linear relationships, addresses simultaneity bias and scale dynamics in firms.



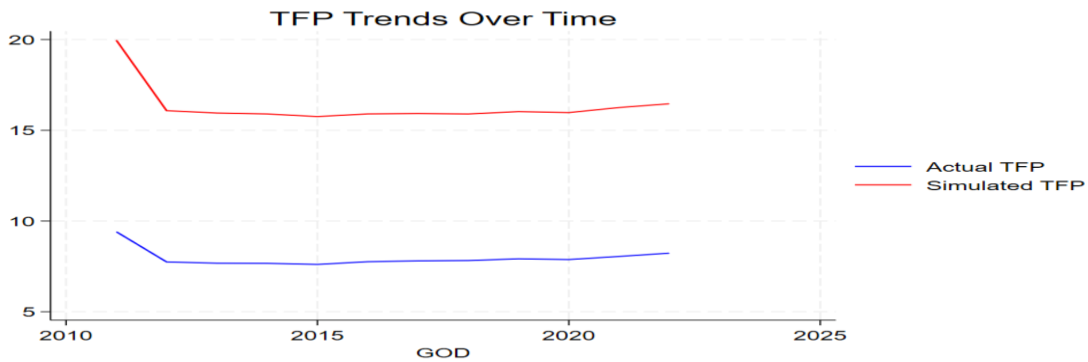
Graph 2

The two density plots illustrate the distribution of TFP under the two above mentioned scenarios. The actual TFP distribution shows that the productivity level is the same for most of the firms; the left tail is thin indicating lower TFP, hence showing that a limited number of firms have low output levels. These levels do not deviate from the mean as most of them are placed around it. The narrow distribution can be explained by the lack of differentiation, which in turn, might be caused by structural and operational constraints. The simulated TFP distribution is wider and shifted to the right, implying the fact that some of the firms experienced improvements in their productivity levels. The range of the values are broader and peaks are lower, indicating heterogeneity.



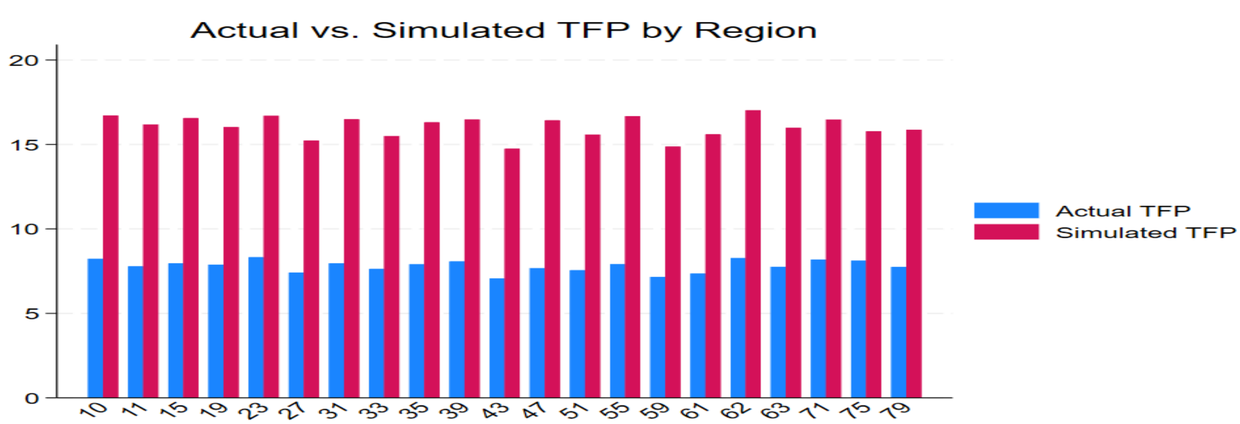
Graph 3

This density plot illustrates the distribution of the difference between simulated and actual TFP. The clustering of the bars around the peak, which is approximately 8, means that most of the firms are better off by improving their capital efficiency. The spread is wide meaning different responses of the firms to these changes in the capital. Firms closer to the production frontier may exhibit higher marginal returns, while those farther away might struggle to translate additional capital into productivity gains.



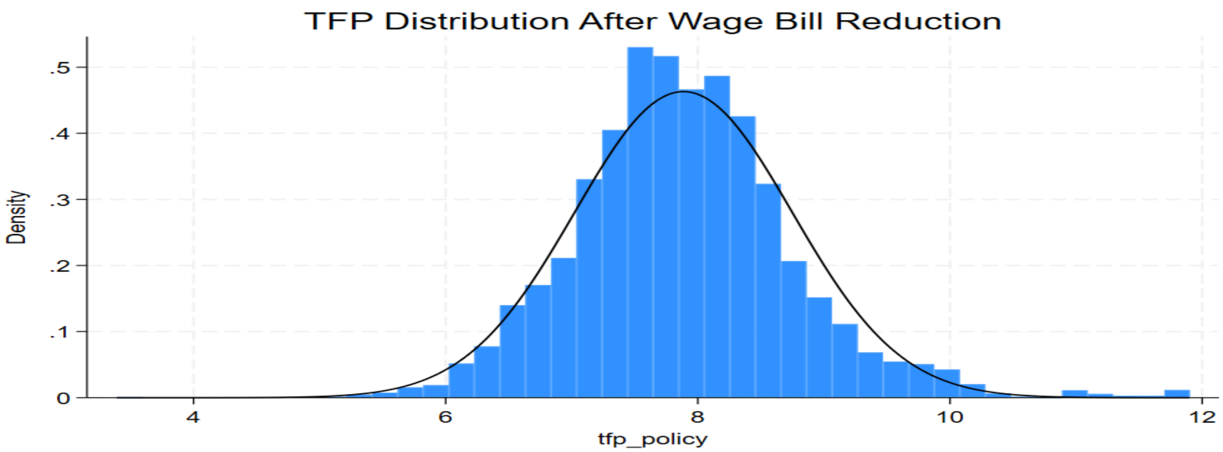
Graph 4

This time-series plot illustrates the trends of actual Total Factor Productivity (TFP) and simulated TFP over the years. The actual TFP (blue line) remains relatively stable over the period, highlighting the modest improvements in firm-level productivity. The simulated TFP (red line) demonstrates a more upward trajectory, mentioning the profits from better capital utilization. The gap between the two lines demonstrates the unrealized productivity gains that could be achieved through these changes in capital. Both trends display a flattening effect and downward effect in post-2015 that can be associated with a global oil price collapse in mid-2014 and transition to a floating exchange rate regime.



Graph 5

This bar chart compares the actual Total Factor Productivity (TFP) and simulated TFP across different regions. Simulated TFP values are consistently higher than actual TFP across all regions. Some regions in the actual TFP present higher actual TFP values, supporting the literature review about the regional differences in Kazakhstan in terms of the technology and better allocation of resources. Similar to the previous graphs, the regions with the largest gaps between actual and simulated TFP show unrealized potential.



Graph 6

The histogram illustrates the distribution of Total Factor Productivity (TFP) after a simulated 10% reduction in the wage bill. The simulated version is dispersed to the right meaning that these changes positively influence TFP. However, policymakers should be cautious about making the decision to reduce labor cost as it reduces employee motivation.

IV. Policy Recommendations

From the analysis in the above section, it can be concluded that SME in Kazakhstan often struggle to achieve economies of scale due to resource constraints and operational inefficiencies. Therefore, policymakers should provide subsidies and grants for SMEs to ensure their business expansion and development. Another solution will be the creation of shared working environments such as technology parks that will allow the business owners to exchange ideas and reduce the aggregate operational costs. Similarly, developing regional innovation hubs will enhance partnerships between different enterprises and contribute to the development of R&D.

The analysis also highlights the critical role of workforce productivity. The significant relationship between labor costs (wage bills) and TFP highlights the significance of the wise use of labor. Therefore, policymakers should analyze the skills required for the workers to improve the production and construct training programs accordingly. The KPI evaluation system is not used everywhere; so, this method should be integrated to raise the motivation of the workers, hence the output growth. Another small addition is the reduction of bureaucracy because it creates regulatory barriers, which in turn, hinders the full functioning of the firms.

Energy efficiency is a significant determinant of TFP in Kazakhstan's SMEs as it was mentioned by several approaches. However, not all firms are efficient in using energy resources, therefore, the development of renewable energy is important and some practical recommendations on reducing the waste of energy can be proposed to SMEs. Economies of scale were mentioned a lot in the analysis highlighting the need for the regular track of the inputs used by the government.

Regional disparities are evident in Kazakhstan. To erase the substantial differences between the regions, grants to each sector should be distributed. Then, economic activities must be decentralized as it was mentioned in the literature review, as most of them are concentrated in Almaty, Astana and East Kazakhstan. Industrial clusters that focus on the development of certain goods can be one approach to mitigate the issue.Infrastructure should be the main concern because roads and railways connect the regions internally, and Kazakhstan with other neighboring countries allowing for more trade opportunities. Digital infrastructure is also an important factor as the development of the internet helps the business owners to find answers to their questions regarding certain aspects of their business and connect with larger enterprises.

Government support for regional startups might also be useful as the residents of particular regions are more informed about the business type that can be developed in their places.

Finally, monitoring and adapting policies is essential for sustained productivity growth. Policymakers should periodically evaluate the efficiency of their solutions considering different economic situations in the country and identify the difficulties of landscapes.

V. Conclusion

This study set out to explore the research question: *What are the key drivers of Total Factor Productivity (TFP) in small and medium enterprises (SMEs) in Kazakhstan?* Using different robust methodologies, such as Olley-Pakes (OP), Levinsohn-Petrin (LP), Ackerberg-Caves-Frazer (ACF), and Kernel-based Learning Techniques (KLT), this paper analyzed the factors affecting TFP in small and medium enterprises of Kazakhstan. Exchange rate shock was used as an instrumental variable to address the endogeneity and better understand the relationship between the output and its inputs.

The findings reveal several critical drivers of TFP. For example, firm size and wage bills positively impact TFP under specific conditions, while inefficiencies in energy and capital utilization hinder productivity growth. The analysis of different methods helped to observe different sector-specific dynamics: the LP and OP had similar conclusions regarding labor and capital relationships on growth, while the ACF and KLT methods added robustness in addressing simultaneity and heterogeneity across firms.

This paper makes a significant contribution to the literature by providing comparative methodological analysis that allows us to see the drivers of TFP in each step. Then, using the KLT method, more specific analysis on sectors and firm sizes were made that drew a better conclusion regarding the inputs allocation across regions and SME in Kazakhstan. The inclusion of simulated and counterfactual cases are useful in providing policy recommendations on governmental level.

The research question has been answered comprehensively, offering both theoretical and practical implications. The study confirms that TFP in SMEs is influenced by a combination of internal factors, such as labor and capital , and external shocks, such as exchange rate fluctuations. The paper can contribute on the academic level to those who want to analyze the total factor productivity in Kazakhstan on a firm level and also, practically, by providing policy recommendations that can serve as a roadmap for a better allocation of resources and economies of scale.

VI. Limitations

This research provides valuable insights into the drivers of Total Factor Productivity (TFP) in small and medium enterprises (SMEs) in Kazakhstan. However, there are several limitations that need to be addressed in the future analysis.

First, firm-level datasets do not fully reflect all the determinants of TFP. For instance, unobservable factors such as firm-specific management, organization, and networks are also important for the TFP, but they are not reflected in the current paper. Future studies could include survey data or qualitative insights to better capture these dimensions.Similarly, even though labor, capital, and materials are crucial for the TFP estimation, technological advancements and R&D were not mentioned because of the data limitations.

The use of exchange rate shocks as an instrumental variable (IV) introduces its own limitations. While exchange rate shocks are exogenous to firm-level decisions and provide a valid instrument for addressing endogeneity, they might not capture the full range of external shocks affecting SMEs. Besides, IV captures local average treatment effect, which is based on the time deviations of exogenous shocks.

Regional disparities in productivity were acknowledged but not thoroughly analyzed in this research. Only KLT under certain assumptions was provided, therefore, further research can include the analysis based on all the approaches and focus deeply on the sectors instead of just the regions and firm sizes. In addition to the previous limitation, the assumptions regarding wage bill reduction and increasing capital were simplified, which is not a usual case in the real economy. The use of dynamic models can be a solution that can capture long-term effects of the policy interventions.

References:

Adilkhanova, Z. (2022). *Sectoral total factor productivity and its determinants: Firm-level evidence from Kazakhstan* (NAC Analytica Working Paper 26). NAC Analytica, Nazarbayev University. https://ideas.repec.org/p/ajx/wpaper/26.html

Garicano, L., Lelarge, C., & Van Reenen, J. (2016). Firm size distortions and the productivity distribution: Evidence from France. *American Economic Review, 106*(11), 3439–3479. https://doi.org/10.1257/aer.20131517

Ilke Van Beveren (2012,February). "Total Factor Productivity Estimation: A Practical Review," *Journal of Economic Surveys, vol. 26(1)*, pages 98-128.https://doi/10.1111/j.1467-6419.2010.00631.x

Iootty De Paiva Dias, M., Bizhan, A., & Correa, P. G. (2023, January 10). *Boosting productivity in Kazakhstan with micro-level tools: Analysis and policy lessons* (English). World Bank. https://documents1.worldbank.org/curated/en/607591468303061568/pdf/E42490P146021000Box377382B00PUBLIC0.pdf

Kim, J., & Park, J. (2017, November). The role of total factor productivity growth in middle-income countries.

Lall, S. V., Izvorski, I. V., Correa, P. G., Seitz, W. H., Rahardja, S., Lootty, M., Sivaev, D., Bogdan, O., & Davies, E. A. R. (2023, February 8). *Kazakhstan Country Economic Memorandum: Dependence, distance, dispersion - Options for upgrading Kazakhstan's economy* (English). World Bank. https://documents.worldbank.org/en/publication/documents-reports

OECD. (2018). *SME and entrepreneurship policy in Kazakhstan 2018* (OECD Studies on SMEs and Entrepreneurship). OECD Publishing. https://doi.org/10.1787/9789264301450-en

U.S. Bureau of Labor Statistics. (2024, March 21). Total factor productivity major industry contributions to output. U.S. Bureau of Labor Statistics. https://www.bls.gov/productivity/highlights/contributions-of-total-factor-productivity-major-industry-to-output.htm

Washington, D.C. : World Bank Group. (2019, February 18). *Kazakhstan: Reversing productivity stagnation - Country economic memorandum*. World Bank Group. http://documents.worldbank.org/curated/en/615051550479498194/Kazakhstan-Reversing-Productivity-Stagnation-Country-Economic-Memorandum

Yasmin, T., El Refae, G. A., Eletter, S., & Kaba, A. (2022). Examining the total factor productivity changing patterns in Kazakhstan: An input-output analysis. *Journal of Eastern European and Central Asian Research (JEECAR)*, *9*(6), 938–950. https://doi.org/10.15549/jeecar.v9i6.958

Appendix

Number of clusters (id) =	781	Number of obs =	19029
		F(1, 780) =	72.86
		Prob > F =	0.0000
Total (centered) SS =	36278.17268	Centered R2 =	0.0204
Total (uncentered) SS =	36278.17268	Uncentered R2 =	0.0204
Residual SS =	35537.14546	Root MSE =	1.37

tfp_lp	Coefficient	Robust std. err.	t	P>|t|	[95% conf.	interval]
firm_size	1.712808	.200663	8.54	0.000	1.318904	2.106711

Table 1

Number of clusters (id) =	722	Number of obs =	17792
		F(6, 721) =	39.32
		Prob > F =	0.0000
Total (centered) SS =	32952.29288	Centered R2 =	0.1370
Total (uncentered) SS =	32952.29288	Uncentered R2 =	0.1370
Residual SS =	28438.93212	Root MSE =	1.268

tfp_lp	Coefficient	Robust std. err.	t	P>|t|	[95% conf.	interval]
firm_size	1.643899	.2329026	7.06	0.000	1.18665	2.101147
fuel_costs	-4.26e-08	2.18e-08	-1.96	0.051	-8.54e-08	1.13e-10
wagebill	-6.19e-09	1.67e-08	-0.37	0.712	-3.91e-08	2.67e-08
ownership	.0389015	.0097412	3.99	0.000	.019777	.0580261
total_costs	1.29e-09	1.34e-09	0.96	0.336	-1.35e-09	3.94e-09
gross_profits	-4.53e-11	6.97e-10	-0.07	0.948	-1.41e-09	1.32e-09

Table 2

Number of clusters (id) =	781	Number of obs =	19092
		F(1, 780) =	13.48
		Prob > F =	0.0003
Total (centered) SS =	36578.27624	Centered R2 =	-5.1229
Total (uncentered) SS =	36578.27624	Uncentered R2 =	-5.1229
Residual SS =	223966.7653	Root MSE =	3.434

tfp_lp	Coefficient	Robust std. err.	t	P>|t|	[95% conf.	interval]
wagebill	2.84e-07	7.73e-08	3.67	0.000	1.32e-07	4.35e-07

Table 3

Number of clusters (id) =	667		Number of obs =	16546
			F(7, 666) =	3.44
			Prob > F =	0.0013
Total (centered) SS =	31092.6188		Centered R2 =	-5.3293
Total (uncentered) SS =	31092.6188		Uncentered R2 =	-5.3293
Residual SS =	196793.8301		Root MSE =	3.46

tfp_lp	Coefficient	Robust std. err.	t	P>|t|	[95% conf.	interval]
wagebill	8.34e-07	4.27e-07	1.95	0.052	-5.78e-09	1.67e-06
firm_size	.6139676	.1956233	3.14	0.002	.229855	.9980803
energy_costs	-3.85e-07	2.94e-07	-1.31	0.190	-9.62e-07	1.92e-07
fuel_costs	-4.62e-07	2.63e-07	-1.75	0.080	-9.79e-07	5.55e-08
ownership	.0391862	.0226486	1.73	0.084	-.005285	.0836573
total_costs	-4.41e-08	2.33e-08	-1.89	0.059	-8.99e-08	1.66e-09
gross_profits	9.00e-09	4.96e-09	1.81	0.070	-7.46e-10	1.88e-08

Table 4

Number of clusters (id) =	697		Number of obs =	17411
			F(1, 696) =	14.21
			Prob > F =	0.0002
Total (centered) SS =	33403.7675		Centered R2 =	-3.0922
Total (uncentered) SS =	33403.7675		Uncentered R2 =	-3.0922
Residual SS =	136694.6835		Root MSE =	2.81

tfp_lp	Coefficient	Robust std. err.	t	P>|t|	[95% conf.	interval]
energy_costs	1.04e-06	2.75e-07	3.77	0.000	4.98e-07	1.58e-06

Table 5

Number of clusters (id) =	667		Number of obs =	16546
			F(7, 666) =	0.00
			Prob > F =	1.0000
Total (centered) SS =	31092.6188		Centered R2 =	-4.0e+04
Total (uncentered) SS =	31092.6188		Uncentered R2 =	-4.0e+04
Residual SS =	1242232647		Root MSE =	274.9

tfp_lp	Coefficient	Robust std. err.	t	P>|t|	[95% conf.	interval]
energy_costs	.0001242	.0041409	0.03	0.976	-.0080066	.008255
firm_size	-75.12586	2536.641	-0.03	0.976	-5055.902	4905.65
fuel_costs	2.77e-06	.0000967	0.03	0.977	-.000187	.0001926
wagebill	-.0000202	.0006745	-0.03	0.976	-.0013445	.0013042
ownership	-2.563685	87.20767	-0.03	0.977	-173.7988	168.6714
total_costs	-3.54e-07	.000012	-0.03	0.976	-.0000239	.0000232
gross_profits	9.71e-07	.0000325	0.03	0.976	-.0000629	.0000649

Table 6

. regress logsales logmaterials logcapital omega_predicted

Source	SS	df	MS
Model	36628.4603	3	12209.4868
Residual	546.999794	15,002	.036461791
Total	37175.4601	15,005	2.47753816

Number of obs	=	15,006
F(3, 15002)	>	99999.00
Prob > F	=	0.0000
R-squared	=	0.9853
Adj R-squared	=	0.9853
Root MSE	=	.19095

logsales	Coefficient	Std. err.	t	P>|t|	[95% conf.	interval]
logmaterials	.3383205	.0012132	278.86	0.000	.3359424	.3406985
logcapital	.3635323	.001148	316.68	0.000	.3612821	.3657824
omega_predicted	.9966304	.0020127	495.18	0.000	.9926853	1.000575
_cons	5.490365	.0122828	446.99	0.000	5.466289	5.514441

Table 7

Number of clusters (id)	=	781
Total (centered) SS	=	8964.952116
Total (uncentered) SS	=	8964.952116
Residual SS	=	8991.128126

Number of obs	=	19026
F(1, 780)	=	16.57
Prob > F	=	0.0001
Centered R2	=	-0.0029
Uncentered R2	=	-0.0029
Root MSE	=	.6893

tfp_acf	Coefficient	Robust std. err.	t	P>|t|	[95% conf.	interval]
firm_size	.3900732	.0958303	4.07	0.000	.2019573	.5781891

Table 8

Number of clusters (id)	=	722
Total (centered) SS	=	8039.066104
Total (uncentered) SS	=	8039.066104
Residual SS	=	6848.617032

Number of obs	=	17789
F(6, 721)	=	5.27
Prob > F	=	0.0000
Centered R2	=	0.1481
Uncentered R2	=	0.1481
Root MSE	=	.6224

tfp_acf	Coefficient	Robust std. err.	t	P>|t|	[95% conf.	interval]
firm_size	.3369592	.1108043	3.04	0.002	.1194216	.5544969
fuel_costs	-1.53e-08	7.42e-09	-2.06	0.040	-2.99e-08	-7.34e-10
wagebill	4.88e-09	5.60e-09	0.87	0.384	-6.11e-09	1.59e-08
ownership	.0177919	.0043526	4.09	0.000	.0092467	.0263372
total_costs	-3.64e-11	4.38e-10	-0.08	0.934	-8.96e-10	8.24e-10
gross_profits	5.53e-10	2.30e-10	2.40	0.017	1.01e-10	1.01e-09

Table 9

Number of clusters (id) =	787		Number of obs =	19067
			F(1, 786) =	72.86
			Prob > F =	0.0000
Total (centered) SS =	36349.96471		Centered R2 =	0.0087
Total (uncentered) SS =	36349.96471		Uncentered R2 =	0.0087
Residual SS =	36033.05046		Root MSE =	1.378

tfp_op	Coefficient	Robust std. err.	t	P>|t|	[95% conf.	interval]
firm_size	1.726056	.2022198	8.54	0.000	1.329101	2.123011

Table 10

Number of clusters (id) =	726		Number of obs =	17813
			F(6, 725) =	38.79
			Prob > F =	0.0000
Total (centered) SS =	33001.44287		Centered R2 =	0.1275
Total (uncentered) SS =	33001.44287		Uncentered R2 =	0.1275
Residual SS =	28794.48366		Root MSE =	1.275

tfp_op	Coefficient	Robust std. err.	t	P>|t|	[95% conf.	interval]
firm_size	1.655687	.2338079	7.08	0.000	1.196666	2.114708
fuel_costs	-4.31e-08	2.20e-08	-1.96	0.050	-8.62e-08	4.91e-11
wagebill	-6.47e-09	1.68e-08	-0.38	0.701	-3.95e-08	2.66e-08
ownership	.0389994	.0097932	3.98	0.000	.0197729	.058226
total_costs	1.29e-09	1.35e-09	0.96	0.339	-1.36e-09	3.94e-09
gross_profits	-3.68e-11	6.96e-10	-0.05	0.958	-1.40e-09	1.33e-09

Table 11

```
. *Re-estimate production function
. regress logsales logcapital logmaterials omega_predicted
```

Source	SS	df	MS
Model	31049.5606	3	10349.8535
Residual	6125.89949	15,002	.408338854
Total	37175.4601	15,005	2.47753816

Number of obs	=	15,006
F(3, 15002)	=	25346.24
Prob > F	=	0.0000
R-squared	=	0.8352
Adj R-squared	=	0.8352
Root MSE	=	.63901

logsales	Coefficient	Std. err.	t	P>|t|	[95% conf.	interval]
logcapital	.322971	.0038753	83.34	0.000	.3153749	.330567
logmaterials	.2179404	.0042461	51.33	0.000	.2096175	.2262633
omega_predicted	.4976725	.0054851	90.73	0.000	.486921	.5084239
_cons	7.903845	.04861	162.60	0.000	7.808563	7.999126

Table 12

Number of clusters (id)	=	781
Total (centered) SS	=	11481.54602
Total (uncentered) SS	=	11481.54602
Residual SS	=	11611.27056

Number of obs	=	19026
F(1, 780)	=	37.13
Prob > F	=	0.0000
Centered R2	=	-0.0113
Uncentered R2	=	-0.0113
Root MSE	=	.7834

tfp_klt	Coefficient	Robust std. err.	t	P>|t|	[95% conf.	interval]
firm_size	.6930364	.1137322	6.09	0.000	.4697789	.9162938

Table 13

Number of clusters (id) =	722		Number of obs	=	17789
			F(6, 721)	=	8.18
			Prob > F	=	0.0000
Total (centered) SS	=	10348.77845	Centered R2	=	0.1457
Total (uncentered) SS	=	10348.77845	Uncentered R2	=	0.1457
Residual SS	=	8840.6087	Root MSE	=	.7071

tfp_klt	Coefficient	Robust std. err.	t	P>|t|	[95% conf.	interval]
firm_size	.6366671	.1300335	4.90	0.000	.3813776	.8919566
fuel_costs	-2.10e-08	1.03e-08	-2.03	0.043	-4.12e-08	-7.04e-10
wagebill	1.45e-09	7.60e-09	0.19	0.849	-1.35e-08	1.64e-08
ownership	.0230645	.005162	4.47	0.000	.0129302	.0331989
total_costs	3.18e-10	6.35e-10	0.50	0.617	-9.29e-10	1.56e-09
gross_profits	3.68e-10	3.37e-10	1.09	0.276	-2.94e-10	1.03e-09

Table 14

Number of clusters (id) =	781		Number of obs	=	19089
			F(1, 780)	=	12.30
			Prob > F	=	0.0005
Total (centered) SS	=	11546.98505	Centered R2	=	-2.3535
Total (uncentered) SS	=	11546.98505	Uncentered R2	=	-2.3535
Residual SS	=	38722.58808	Root MSE	=	1.428

tfp_klt	Coefficient	Robust std. err.	t	P>|t|	[95% conf.	interval]
wagebill	1.15e-07	3.28e-08	3.51	0.000	5.07e-08	1.80e-07

Table 15

Number of clusters (id) =	667		Number of obs	=	16543
			F(7, 666)	=	3.49
			Prob > F	=	0.0011
Total (centered) SS	=	9525.890439	Centered R2	=	-2.6454
Total (uncentered) SS	=	9525.890439	Uncentered R2	=	-2.6454
Residual SS	=	34725.70801	Root MSE	=	1.454

tfp_klt	Coefficient	Robust std. err.	t	P>|t|	[95% conf.	interval]
wagebill	3.37e-07	1.71e-07	1.97	0.049	1.02e-09	6.73e-07
firm_size	.2334493	.0786325	2.97	0.003	.0790518	.3878469
energy_costs	-1.69e-07	1.17e-07	-1.45	0.148	-3.99e-07	6.04e-08
fuel_costs	-1.87e-07	1.05e-07	-1.78	0.076	-3.94e-07	1.98e-08
ownership	.022354	.0093815	2.38	0.017	.0039332	.0407748
total_costs	-1.76e-08	9.35e-09	-1.88	0.061	-3.59e-08	7.93e-10
gross_profits	3.79e-09	2.03e-09	1.86	0.063	-2.02e-10	7.78e-09

Table 16

Number of clusters (id) =	697	Number of obs =	17408
		F(1, 696) =	12.24
		Prob > F =	0.0005
Total (centered) SS =	10217.38092	Centered R2 =	-1.6947
Total (uncentered) SS =	10217.38092	Uncentered R2 =	-1.6947
Residual SS =	27532.30604	Root MSE =	1.261

tfp_klt	Coefficient	Robust std. err.	t	P>|t|	[95% conf.	interval]
energy_costs	4.15e-07	1.18e-07	3.50	0.000	1.82e-07	6.47e-07

Table 17

Number of clusters (id) =	667	Number of obs =	16543
		F(7, 666) =	0.02
		Prob > F =	1.0000
Total (centered) SS =	9525.890439	Centered R2 =	-6.8e+03
Total (uncentered) SS =	9525.890439	Uncentered R2 =	-6.8e+03
Residual SS =	65227576.74	Root MSE =	63

tfp_klt	Coefficient	Robust std. err.	t	P>|t|	[95% conf.	interval]
energy_costs	.0000284	.000548	0.05	0.959	-.0010475	.0011044
firm_size	-16.88104	332.0825	-0.05	0.959	-668.9357	635.1736
fuel_costs	6.21e-07	.0000131	0.05	0.962	-.0000251	.0000264
wagebill	-4.61e-06	.0000895	-0.05	0.959	-.0001804	.0001712
ownership	-.573429	11.57172	-0.05	0.960	-23.29488	22.14802
total_costs	-8.11e-08	1.60e-06	-0.05	0.960	-3.23e-06	3.06e-06
gross_profits	2.23e-07	4.32e-06	0.05	0.959	-8.27e-06	8.71e-06

Table 18

Variable	Obs	Mean	Std. dev.	Min	Max
tfp_klt	19,099	7.885903	.8614321	3.389479	11.90186
tfp_sim	19,099	16.04816	1.591018	8.988248	22.98854

Table 20