THE TECHNICAL EFFICIENCY OF COUNTRY RESPONSES TO THE COVID-19 PANDEMIC

A Research Report submitted in partial fulfilment of the Degree of Master of Commerce (Health Economics) in the School of Economics and Finance, University of the Witwatersrand

Written by**:** Armand Desjardins **(**708113**)**

Supervised by: Professor Frederik Booysen

Word Count: 9981

28 February 2023

Abstract

The COVID-19 pandemic saw the most comprehensive implementation in history of pharmaceutical and non-pharmaceutical interventions by governments around the world. A better understanding of how efficient these interventions are, especially given their high economic costs to society is pivotal in designing similar policy for future outbreaks of this nature. This research estimates the technical efficiency of country responses to the COVID-19 pandemic using a Stochastic Frontier Analysis model and panel dataset combining indicators from the Our World in Data (OWID), Google Movement Index and World Bank Economic Indicators databases. The SFA regression also corrects for endogeneity in the input variables, as policy stringency tends to be influenced by the number of cases in each country. To do so, the OWID stringency index is used as an instrumental variable for each of the policy inputs to correct for the endogenous relationship between the input and output variables. The study found that the correction of technical efficiency estimates for endogeneity using the *xtsfkk* routine in Stata was an appropriate estimation strategy. The exogenously and endogenously estimated efficiency scores furthermore differed statistically significantly, with exogenously estimated scores representing overestimates of technical efficiency. Vaccination policies proved to be the most technically efficient channel of prevention followed by stay-at-home restrictions. A second stage OLS regression was used to estimate the effects of selected drivers on the technical efficiency scores computed in the Stage 1 SFA model. Health expenditure as a % of GDP, hospital beds per 1,000 population, and physicians per 1,000 population were all positively related to technical efficiency. Countries with reactive and flexible policy approaches, able to increase policy intensity levels based on COVID-19 waves, were found to obtain higher technical efficiency scores than wealthy countries. Improvements in the level of detail on the drivers of technical efficiency and the adoption of machine learning techniques to estimate efficiency scores are recommended to extend research in this area past its current scope.

Contents

1. Introduction 4
2. Literature Review 6
2.1 SFA Studies 6
2.2 DEA Studies 7
2.3 Instrumental Variable Studies 8
2.4 Machine Learning and Technical Efficiency Studies 9
3. Data 9
4. Methodology 10
4.2 Standard SFA Framework: 11
4.3 Endogeneity Controlled SFA Framework: (Karakaplan M. , 2022) 13
5. Analysis 15
5.1 Stage 1 Analysis – SFA 15
5.2 Tests for Normality in the Error Term 17
5.3 Stage 2 Analysis – TE Drivers 18
6. Results 18
6.1 Stage 1: SFA Results 19
6.1.1 Exogenous vs. Endogenous Results 19
6.1.2 NPI Channel Results 25
6.1.3 Income Level Results 26
6.1.4 Country Level Results 27
6.2 Tests for Normality in the Error Term 30
6.3 Stage 2: TE Drivers Results 30
7. Discussion 31
8. Conclusion 34
References 35

List of Tables

Table 2 - Summary statistics of exogenous variables (drivers of TE) 10
Table 3 – T test of Transport Difference Variable 23
Table 4 – T test of Stay-at-Home Difference Variable 23
Table 5 - T test of Vaccine Difference Variable 24
Table 6 - T test of Workplace Closure Difference Variable 24
Table 7 - T test of Mean Model Difference Variable 24
Table 8 - Stage 2 Regression Results: Drivers of Technical Efficiency 30

List of Figures

Figure 1 - Average New Cases per Million vs Average Stringency Index Trends 19
Figure 2 - Mean Technical Efficiency Score per Country - Endogenous vs. Exogenous Models 23
Figure 3 - Average Technical Efficiency Score per Policy Model 25
Figure 4 - Mean TE Score per NPI Channel vs. Mean Cases per Million Trend for 2021 25
Figure 5 - Mean TE by Income Level: Mean Model + Individual Models 26
Figure 6 - Mean Technical Efficiency Scores by Income Level and Date 27
Figure 7 - Mean Technical Efficiency Score per Country 27
Figure 8 - Mean New Cases per Million vs Mean TE Trend: USA 28
Figure 9 - Mean New Cases per Million vs Mean TE Trend: South Africa 29
Figure 10 - Mean New Cases per Million vs Mean TE Trend: Serbia 29
Figure 11 - Mean New Cases per Million vs Mean TE Trend: Bangladesh 30
Figure 12 - Normal Distribution: Residual Density Plot (transport, stay at home, vaccine, work closure) 30

1. Introduction

The first case of COVID-19 was detected in Wuhan China in late December 2019, and within months had spread to almost every corner of the globe, plunging the world into one of the costliest pandemics to occur in the past few decades. The precipitous spread of the virus forced governments to respond rapidly in an attempt to reduce the loss of life and negative impact on the economy. Facing different domestic conditions, and with limited information on global best practices, countries adopted COVID-19 policy interventions in an attempt to hinder transmission of the disease. This resulted in different forms of national lockdown in most countries, with governments imposing limitations on movement, closing down schools, businesses and transport hubs and enforcing new laws on the wearing of face-masks and social distancing etc. These new policies, targeted at controlling the spread of COVID-19 became known and Non-pharmaceutical Interventions (NPIs), and until the widespread distribution of the vaccine (a pharmaceutical intervention), were governments only weapons against the pandemic.

This research will use the well-established concept of Technical Efficiency (TE) to estimate and compare the efficiency with which different countries responded to the pandemic using targeted COVID-19 NPIs and pharmaceutical vaccination interventions. This will be done by making use of Google Mobility Reports dataset which captures changes in movement across different categories of places such as retail and recreation, groceries and pharmacies, parks, transit stations, workplaces, and residential. These movement indexes will be used as proxies to measure NPIs such as stay at home restrictions, closure of transport hubs and closure of workplaces. The number of vaccinations per 100,000 population will also be used as a proxy for Pharmaceutical Interventions (PIs). Regressing the number of new cases per million population (output variable) on these NPIs and PI (input variables) using a Stochastic Frontier Analysis (SFA) model, the TE score of each country in the sample will be estimated and compared. One of the factors that makes this research unique is that it is the first to control for endogeneity in policy stringency (due to simultaneity) using the SFA methodology. After controlling for endogeneity in the inputs, a second stage regression will be conducted to investigate the drivers of technical efficiency, where the TE scores estimated using the SFA model are regressed on a set of drivers of TE such as health expenditure as a share of GDP, the number of beds per 1,000 population and the number of specialists per 100,000 population. In doing so, the study will make three major contributions to the existing body of literature on country response efficiency to the COVID-19 pandemic.

Firstly, this research will produce one of the first comprehensive applications of SFA as a parametric approach to estimating country-level NPI response efficiency to the COVID-19 pandemic. As this

approach has never been adopted before in conducting research on this particular topic, this is the primary contribution of the paper. While there are a few approaches to modelling technical efficiency in econometrics, the two most popular will be considered in this paper, one that deals with parametric models, - Stochastic Frontier Analysis (SFA)- and one that deals with non-parametric models - Data Envelopment Analysis (DEA). While parametric models generally produce more robust estimates, the majority of existing research has made use of non-parametric DEA techniques in estimating the efficiency of policy responses to COVID-19. In addition to this, SFA produces TE Scores as an absolute percentage, while DEA produces results scoring efficiency as a percentage relative to the most efficient firm, which can be more difficult/complicated to interpret.

Secondly, this study will make use of a panel dataset with daily data over a two-year period. Thus far, all of the country level COVID-19 response studies have used cross-sectional datasets in their analyses, limiting their ability to assess trends in TE over time, a particularly important perspective to have when considering interventions to contain the spread of viruses. By using a panel dataset, this study will be the first to explore the temporal dimensions of TE in the context of country responses to the pandemic. This use of panel data to establish temporal shifts in TE is therefore the second most important contribution of this paper to the existing body of literature. By looking at the fluctuations in TE over time in relation to COVID-19 waves in various countries, the study is able to assess how reactive policy was to spikes and drops in the number of cases over time.

Lastly, this will be the first study to correct for endogeneity in papers that estimate technical efficiency related to COVID-19 country responses. Thus far, none of the existing SFA or DEA studies consider the problem of endogeneity in their analyses. The problem of endogeneity in this case is due to simultaneity and is often encountered when dealing with matters of policy efficiency due to the reactive nature of policy stringency to the problem it is trying to solve. In other words, policy stringency and the problem it is trying to solve are jointly determined. This means that policy stringency has a causal effect on the problem and the problem in turn has a causal effect of the stringency of policy. The study will control for endogeneity from within the SFA framework by using the new *xtsfkk* package in Stata, built by (Karakaplan M. , 2022). Being a fairly new package, the study will be one of the first few to make use of the *xtsfkk* routine to control for endogeneity in a stochastic frontier model, especially in health economics literature.

The three above mentioned contributions of this research are important as they remedy either gaps in the existing body of research (i.e., using SFA model to estimate the TE of COVID-19 NPI stringency), or deficiencies in existing literature. With respect to the remedy of deficiencies in the exiting literature,

the only other study that uses SFA to estimate policy efficiency (Ahmad, Zainon, Nawi, & Ghazali, 2020) lacks a detailed description of the methodology and does not correct for the endogeneity caused by reverse-causality.

An interesting finding of the research was that the *xtsfkk* package was a statistically appropriate tool to estimate the TE scores of country responses to the COVID-19 pandemic for both pharmaceutical and non-pharmaceutical interventions. This was in large part due to the fact that the routine was able to handle endogeneity in the input variables using instrumental variable techniques. Following on from this, it was found that each of the four input variables were endogenous and required correction for endogeneity to obtain unbiased estimates. It was also found that the difference between the results for inputs corrected for endogeneity, and those not, was statistically significant proving that correction for endogeneity was indeed necessary. Importantly, the model that does not correct for endogeneity consistently overestimated TE scores in all cases by around 4% on average. Other key insights were that vaccination policies proved to be the most technically efficient channel of prevention followed by stay-at-home restrictions. The drivers of TE analysis returned satisfactory results, but limitations in data mean that there is room for improvement in assessing the drivers of TE.

The research is laid out as follows: Section 2 provides a brief review of the existing literature on technical efficiency and country responses to the COVID 19 pandemic; Sections 2 and 3 describe the data used in the analysis and the methodology applied; and Sections 5 and 6 present an analysis of the model specifications and the final set of results. Sections 7 and 8 provide some points of discussion and concluding remarks.

2. Literature Review

Having originated in production economics, TE analysis has since been applied to a variety of different fields in economics, especially in healthcare, and more recently in COVID-19-specific studies. Being a fairly recent phenomenon, the body of literature on the application of TE analysis to COVID-19 country responses is still limited, however, one SFA study and a few DEA studies have estimated country responses to the COVID-19 pandemic.

2.1 SFA Studies

The only published application of SFA to COVID-19 country responses is a study by (Ahmad, Zainon, Nawi, & Ghazali, 2020) which compares the technical efficiency of Malaysia's response to the pandemic with other South-east Asian countries. The study used data from the Worldometers database between the 15th of February 2020 and the 13th of June 2020. Countries with no cases were excluded

from the analysis leaving the sample size of countries at seven (n = 7). Although not a global comparison, the study applies the SFA methodology set out by (Battese & Coelli, 1992) to determine how the total deaths and recoveries (inputs) impact the cumulative number of positive cases (output).

An extract from the paper shows the stochastic frontier model expressed as:

$$\ln Y_{it} = \beta_0 + \beta_1 \ln x_{1it} + \beta_2 \ln x_{2it} + (V_{it} - U_{it})$$

where:

Y_{it} = Total positive cases of COVID -19
β = Vector of unknown parameters to be estimate
x_{1it}= Total death cases of COVID -19
x_{2it}= Total recoveries cases of COVID -19
U_{it} = Non-negative random variables, associated with technical inefficiency of total cases of COVID -19 of the country
V_{it} = Assumed to be independent and identically $_{it}$distributed (i.i.d) N (0, σ2 v) and captures statistical noise, measurement error, and other random.

From the specification, it is not immediately clear how to interpret x_{1it} (total death cases) and x_{2it} (total recovery cases) as inputs in the production of output Y_{it} *(*total positive cases). Aside from this, the authors have not considered the problem of endogeneity in the input variables, as the number of cases will directly affect the number of deaths and recoveries. The results of the study show that Thailand was most efficient with a score of around 88.34% followed by Malaysia at 88.33 and Indonesia at 83.3%.

2.2 DEA Studies

Aside from the Malaysia study, there are three country level papers that use DEA to estimate country response efficiency. The first, and most notable of these is a study by (Breitenbach, Ngobeni, & Aye, 2021), which estimated the technical efficiency of country resource management in the fight against COVID-19. The study used a sample of 36 countries, together accounting for over 90% of all COVID-19 cases worldwide. Data was taken from World Bank and Worldometers databases and used to estimate the efficiency with which countries utilised resources such as national budgets, healthcare workers and hospital beds (inputs) to minimise the rate of infection and death (outputs). The paper found that on average, countries were not very efficient at managing the pandemic given health resource constraints with only six of thirty-six countries being efficient. The authors suspect that the root of this inefficiency lies in the way wealthy countries “threw” resources at fighting the pandemic leading to wasted resource use and lower level of efficiency.

Unlike the technical efficiency estimates produced by the SFA which give absolute measures of TE as a percentage, the DEA estimates of technical efficiency offer relative TE scores based on DMUs that are found to be perfectly efficient (TE score of 1).

In an earlier DEA study by the same authors, the technical efficiency of country responses to the first wave of the pandemic was estimated. While very similar in design, the most interesting findings from this study were that some of the wealthiest nations in the world were among the most inefficient in dealing with the pandemic. Some of these countries such as Germany, Canada and the US were less efficient than those countries hardest hit by the pandemic, despite their relatively larger health budget and access to health resources amidst lower infection rates. This is indicative that stringent public policy and high levels of resource allocation did not always result in meaningful reductions in the number of cases.

A third study by (Shirouyehzad & Jouzdani, 2020) makes use of a two-stage DEA model in which the first stage interrogates contagion control measures, and the second investigates the efficiency of clinical interventions. Data from the National Evaluation of Rural Primary Health Care Programs was used in conjunction with International Health Regulations Core Capacity Scores data as a measure of country health levels and contact frequency data from UN databases in the form of population density statistics. An improvement to this configuration may be to replace population density with Google movement indexes as this gives a more detailed view of movement in crowded places. In the first stage, the number of cases was regressed on population density and health scores, and in the second, the number of cases is regressed on deaths and recoveries, all using DEA models.

2.3 Instrumental Variable Studies

In one of the few papers that make use of instrumental variables to correct for endogeneity in COVID-19 mobility restriction policy, (Fakir & Bharati, 2021) make use of Google movement indexes as well as the stringency index in their analysis. In a first stage OLS regression, the mean stringency index in all countries except for the country in question is used as an IV for that country's stringency index (which is endogenous). In the second stage, the mobility index and the growth rates in cases are regressed on stringency to estimate how stringency impacts the mobility index or number of cases outputs.

This study is particularly interesting as it is one of the few studies that considers temporal effects on NPI implementation, although it does not report results from a DEA or SFA model and thus does not estimate TE levels. Rather, a 2SLS model using Instrumental variables is chosen. It should be noted that an SFA model was run by the authors to compare with the 2SLS model, but it was not reported on

in the paper. Overall, the study found that "while stricter contemporaneous measures affected mobility, stringency in seven to fourteen days mattered most for containing the contagion" (Fakir & Bharati, 2021, p. 1).

2.4 Machine Learning and Technical Efficiency Studies

There is one study that makes use of Machine Learning (ML) techniques in estimating the TE of country responses to the COVID-19 pandemic. (Taherinezhad & Alinezhad, 2022) first use a 2-Stage DEA model that allows for both desirable and undesirable variables is used to estimate efficiency scores for 50 countries on 5 December 2020. A Multi-layer-perception (MLP) network with a Limited memory BFGS (L-BFGS) optimisation algorithm is then applied to the static December 5th results to predict TE scores for each country at any given time during the pandemic.

3. Data

The data for the analysis was sourced from two different databases. The COVID related data fields came from Our World in Data (OWID) database, and the endogenous variables (health as a % of GDP, beds per 1,000 and health workers per 1,000 etc.) fields from the World Bank Development Indicators database. Note that the Google Movement Index data formed part of the OWID dataset already, and so did not need to be included separately.

The data sample consists of 45 countries (n = 45), and stretches between May 1st, 2020, and April 30th, 2022 (t = 2 years). These 45 countries account for over 90% of all new daily cases in the dataset. While the COVID data (number of cases and policy variables) are reported at the daily level, the exogenous variables (beds, specialists, health as a % of GDP etc.) are reported at the annual level. For these exogenous variables, where data was missing, an average between the years 2015 and 2021 was taken. This was done as the data was not reported consecutively in most cases and did not change considerably between reporting periods where consecutive years were available. For this reason, there is little to no intertemporal variation in these driver variables, and as such a panel dataset is not possible to obtain for the second stage. Rather, a standard OLS regression on annualised data will be used. This remain a data constraint and a component of the study that can potentially be improved upon in future.

The dependent or output variable in the study is the number of new cases per day per 1 million population (daily case rate). Denmark (728.61), Austria (638.61), Slovakia (636.47) and the Netherlands (627.64) had the highest average daily case rates per million population, while Bangladesh (15.73), Indonesia (30.21), India (41.89), the Philippines (44.36), Mexico (61.85) and Iraq (73.10) had the lowest rates.

Another key variable in the study is the stringency index, a composite measure of government stringency regarding various NPIs. It gives and overall measure of how stringent a country was with its COVID-19 policy as a percentage. Interestingly, of the 6 countries with the lowest average case rate, the Philippines (69.32), India (69.81), and Iraq (69.44) also appeared in the list of countries with the highest average stringency index score over the period. This indicates that some stringent countries also displayed low average levels of new cases per day and that an investigation of causality into the specific NPIs that may have driven this outcome is needed.

Table 1 shows the summary statistics of the drivers of TE, the variables that will be investigated in the second stage regression. The health expenditure as a % of GDP variable had a mean of just under 8% indicating that the sample of countries is likely skewed towards larger wealthier economies with large health budgets and national health institutions that are able to respond effectively to the pandemic.

Table 1 - Summary statistics of exogenous variables (drivers of TE)

Variable	Obs	Mean	Std. dev.	Min	Max
Health as a % of GDP	120	7.98	2.997	2.507	16.726
Beds per 1000	105	3.845	2.649	0.530	13.078
Nurses per 1000	114	3.000	0.000	3.000	3.000
Physicians per 1000	117	3.274	1.890	0.412	7.937
Specialists per 100 000	75	72.186	40.146	2.910	163.540

4. Methodology

As mentioned, when estimating technical efficiency, there are two major approaches to modelling the frontier, the linear programming DEA method and the econometric SFA method. This study has chosen to use SFA, as it has some notable advantages over DEA results and meets the necessary assumptions regarding data size and structure. “Its biggest advantage lies in the fact that it introduces a disturbance term representing noise, measurement error, and exogenous shocks beyond the control of the production unit. This, in turn, permits the decomposition of deviations from the efficient frontier into two components, inefficiency and noise.” (Worthington, 2003, p. 157). Additionally, SFA estimates absolute efficiency scores as a percentage while DEA estimates efficiency scores relative to a technically efficiency country (TE = 1).

The methodology[1] adopted in this study uses a Panel Stochastic Frontier Analysis model with instrumental variables to handle endogeneity in the model inputs. The model was developed by (Karakaplan M. , 2022) and is published in the Stata Journal along with the Stata Manual for the *xtsfkk*

[1] *Note: The econometric models described in this section was extracted directly from the Stata Journals* (Belotti et al., 2013) and (Karakaplan M. , 2022).

routine which defines the package. Before presenting this adapted model however, the standard SFA model will be presented.

4.2 Standard SFA Framework:

Beginning with the standard SFA model, in the 1970s, production function theory in economics started to move away from the deterministic production function, a non-stochastic model that did not account for inefficiency, toward the frontier production function that could estimate inefficiency (Aigner et al., 1977). The production function with no stochastic disturbance term can be written as:

$$y_i = f(x_i ; \beta) \quad \text{.....(1)}$$

Where:

y_i *is* the maximum output obtainable,

x_i is a vector of non-stochastic inputs, and

β is the unknown parameter variable to be estimated

This paradigm shift was pivotal as it moved the focus towards production efficiency/inefficiency by assuming that not all production points are efficient, and as such, a frontier must exist between efficient and inefficient production points. This was necessary to account for firms with identical inputs mixes but different outputs. Further, the inefficiency could be estimated as part of the stochastic disturbance term using econometric techniques (Aigner et al., 1977).

The result was the stochastic production function which can be written as:

$$y_i = f(x_i ; \beta) + \varepsilon_i \quad \text{.....(2)}$$

Where:

$y_i = output$ of firm i

$x_i = the\ input\ vector$ of firm i

β = technology vector of firm i

ε_i = the composite error term comprising:

u_i as the stochastic noise term; and

v_i as the estimate of inefficiency.

Next, output is assumed to be subject to random shocks, and as such:

$$y_i = f(x_i , \beta)\ \varepsilon_i \exp(v_i) \quad \text{.....(3)}$$

Taking the natural log of both sides:

$\ln(y_i) = \ln\{f(x_i, \beta)\} + \ln(\varepsilon_i) + v_i$ .....(4)

Assuming *k* inputs and that the production function is linear in logs:

$\ln(y_i) = \beta_0 + \sum_{j=1}^{k} \beta_j \ln(x_{ji}) + v_i + u_i$ .....(5)

Where:

ε_i *is* the composite error term comprising:

$u_i = -\ln(\varepsilon_i)$ as the stochastic noise term; and

v_i as the estimate of inefficiency.

In order to estimate the model, we first need to list the econometric assumptions. It is assumed that u_i and v_i are independent from one another and are independent and identically distributed (IID) across observations. A half normal distribution is assumed by (Aigner et al., 1977) yielding $u_i \sim N^+(0;\sigma^2)$.

According to (Belotti et al., 2013, p. 722), "The distributional assumption required for the identification of the inefficiency term implies that this model is usually fit by maximum likelihood (ML), even if modified ordinary least-squares or generalized method of moments estimators are possible (often inefficient) alternatives."

The estimation of the model entails two steps:

Step 1: estimates of the model parameters $\hat{\theta}$ are obtained by maximizing the log-likelihood function $l(\theta)$, where $\theta = (\alpha, \beta', \sigma_u^2, \sigma_v^2)'$

Step 2: point estimates of inefficiency can be obtained through the mean (or the mode) of the conditional distribution $f(u_i \mid \widehat{\varepsilon_i})$ where $\widehat{\varepsilon_i} = y_i - \hat{\alpha} - x_i' \hat{\beta}$.

The derivation of the likelihood function is based on the independence assumption between u_i and v_i. Because the composite model error ε_i is defined as $\varepsilon_i = v_i - u_i$, its probability density function is the convolution of the two component densities as

$f_\varepsilon(\varepsilon_i) = \int_0^{+\infty} f_u(u)\, f_v(\varepsilon_i + u_i)\, d\, u_i$ .....(6)

Hence, the log-likelihood function for a sample of n productive units is

$l(\theta) = \sum_{i=1}^{n} \log f_\varepsilon\ (\varepsilon i \mid \theta)$ .....(7)

The second estimation step is necessary because the estimates of the model parameters allow for the computation of residuals $\widehat{\varepsilon_i}$ but not for the inefficiency estimates. Therefore, a technique to separate

the technical efficiency component from the stochastic noise component in the compound error term is required.

As mentioned by (Belotti et al., 2013)," the most well-known solutions to this problem, proposed by (Jondrow et al., 1982) and Battese and Coelli (1988), exploit the conditional distribution of u given ε. Thus, a point estimate of the inefficiencies can be obtained using the mean E(u | $\widehat{\varepsilon_i}$) - (or the mode M[u | $\widehat{\varepsilon_i}$]) of this conditional distribution.

Once point estimates of u are obtained, estimates of the technical (cost) efficiency can be derived as

$$\text{Eff} = \exp(-\widehat{u})$$

where $\widehat{u}$ is either E(u| $\hat{\varepsilon}$) or M(u|$\hat{\varepsilon}$)).

In the current case, we are investigating how COVID policies (input variables) affect the number of COVID-19 cases (output variable). COVID-19 policy is an endogenous variable because not only does policy affect the number of cases, but the number of cases directly affects policy also.

4.3 Endogeneity Controlled SFA Framework: (Karakaplan M. , 2022)

It has been forty years since the original development of the SFA model, and as an extension of this framework, a model able to handle endogeneity has been introduced, first by (Karakaplan & Kutlu, 2017) with a cross-sectional version, and later by (Karakaplan, 2022) who introduced the panel variation.

The following is a specification of the (Karakaplan M. , 2022) SFA model as recorded in the Stata Journal:

$$lnL_i = lnL_{i,y|x} + lnL_{i,x}$$

$$lnL_{i,y|x} = -\frac{1}{2}\left\{T_i \ln(2\pi\sigma_w^2) + \frac{e_{i.}'e_i}{\sigma_w^2} + \left(\frac{\mu^2}{\sigma_u^2} - \frac{\mu_{i*}^2}{\sigma_{i*}^2}\right)\right\} + \ln\{\frac{\sigma_{i*}\Phi\left(\frac{\mu_{i*}}{\sigma_{i*}}\right)}{\sigma_u\Phi\left(\frac{\mu}{\sigma_u}\right)}\}$$

$$lnL_{i,y|x} = -\frac{1}{2}\sum_{t=1}^{T_i}\{\ln(|2\pi\Omega|) + e_{it}'\Omega^{-1}\epsilon_{it}\}$$

$$\mu_{i*} = \frac{\sigma_w^2\mu - s\sigma_u^2 e_{i.}'h_{i.}}{\sigma_u^2 h_{i.}'h_{i.} + \sigma_w^2}$$

$$\sigma_{i*}^2 = \frac{\sigma_w^2\sigma_u^2}{\sigma_u^2 h_{i.}'h_{i.} + \sigma_w^2}$$

$$e_{it} = y_{it} - x_{1it}'\beta - \epsilon_{it}'\eta$$

$$\epsilon_{it} = x_{it} - Z_{it}\delta$$

where a vector of observations corresponding to the panel *i* is represented by a subscript *i.*; T_i is the number of time periods for panel *I*; s = 1 for cost functions or s = -1 for production functions; y_{it} is the logarithm of the production or cost of the *ith* productive unit at time t; x_{yit} is a vector of exogenous and endogenous variables; x_{it} is a vector of all endogenous explanatory variables; $Z_{it} = I_p \otimes z'_{it}$ where z_{it} is a vector of all exogenous variables; v_{it} and _$_{it}$ are two-sided error terms; $u_{it} \geq 0$ is a one-sided error term capturing inefficiency; $h_{it} = h(x'_{uit}\varphi_u) > 0$; x_{uit} is a vector of exogenous and endogenous variables excluding the constant; u_i^* is a producer-specific random component independent from v_{it} and ϵ_{it}; Ω is the variance-covariance matrix of ϵ_{it}; σ_v^2 is the variance of v_{it}; ρ is the vector representing correlation between $\widetilde{\epsilon_{it}}$it and v_{it}; $w_{it} = \sigma_v\sqrt{1 - \rho'\rho}\widetilde{w_{it}} = \sigma_v\widetilde{w_{it}}$; $\eta = \sigma_w\Omega^{-1/2}\rho / \sqrt{1 - \rho'\rho}$; e_{it} is conditionally independent from the regressors given x_{it} and z_{it}; Φ denote the standard normal cumulative distribution function; $u_i^* \sim N^+(\mu, \sigma_u^2)$ (N^+ is standard notation for half-normal distribution); and $h_{it}^2 = \exp(x'_{uit}\varphi_u)$. (Karakaplan & Kutlu, 2017) provide all the details about assumptions and how they derived the estimator.

Furthermore, to predict efficiency, $Eff_{it} = \exp(-u_{it})$, (Karakaplan & Kutlu, 2017) give the following formula:

$$\exp\{-E(u_{it}|e_i)\} = \exp\left[-h_{it}\left\{\mu_{i*} + \frac{\sigma_{i*}\emptyset\left(\frac{\mu_{i*}}{\sigma_{i*}}\right)}{\Phi\left(\frac{\mu_{i*}}{\sigma_{i*}}\right)}\right\}\right]$$

Where ∅ denotes the standard normal probability distribution function.

Finally, Karakaplan and Kutlu (2017b) offer a test for endogeneity based on a reasoning similar to that of the standard Durbin-Wu-Hausman test. The test is conducted by looking at the joint significance of the components of the error term. If the components of the error term are jointly significant, then that would tell us there is endogeneity in the model and a correction through (1) would be needed. If, on the other hand, the joint significance of the components is rejected, then correction for endogeneity would not be needed and the model can be fit by traditional exogenous frontier models.

While the SFA framework was initially exclusive to production economics, it is now used in various different economic fields to assess efficiency levels, including health economics. (Papadopoulos, 2021, p. 163) argues that the SFA model "is the clearest and most colourful example of how

frontier modelling can be applied to very diverse situations, economic and non-economic alike". While the initial SFA model referred to firms under the production context, the more universal Decision-Making Unit (DMU) has now been adopted to refer to the individual members that make up the panel. In the current context, the DMUs are countries.

A Note on Instrumental Variables

An instrumental variable is an additional variable that is introduced to a regression with endogenous variables (variables that are indirectly affected by other variables in the regression) to correct for this unintended relationship between variables. The instrument acts as a proxy for the endogenous variable as it mimics the relationship between the endogenous variable and the outcome variable but corrects for the unwanted relationships between variables.

According to (Labrecque & Swanson, 2018, p. 1), "IV analysis requires first and foremost an instrumental variable. That is, it requires a variable that meets three conditions: (1) it is associated with the exposure ("relevance"), (2) it only affects the outcome through the exposure ("exclusion restriction"), and (3) its effect on the outcome is unconfounded ("exchangeability")".

5. Analysis

The analysis consists of a set of models that make use of the (Karakaplan M. , 2022) SFA endogeneity model. As such, each model is to be specified according to the *xtsfkk* routine in Stata.

Testing the Significance of Correction for Endogeneity

As one of the major contributions of this study, the question of whether the correction for endogeneity on the input variables is in fact significant will be answered. To do this, both the exogenous and endogenous TE scores will be estimated. Having done this, the difference between the two ($TE_{EXOG} - TE_{ENDOG}$) will be calculated. This difference variable will then be tested for statistical significance (using a one-sample t-test) to establish if the difference between exogenous and endogenous TE scores are statistically significantly different from zero.

5.1 Stage 1 Analysis – SFA

Model 1: Closure of Transport Hubs

$$\ln(\#cases\ per\ million) = \beta_0 + \beta_1 \ln(google\_transport\_index) + \ln(\varepsilon_i) + v_i$$

In this scenario, we are assessing the TE with which countries used the closure of transportation hubs to contain the pandemic. As such we regress the natural logarithm of new cases per million population (output variable) on the Google movement index for transportation hubs (input variable). We note that the input variable is endogenous as the number of new cases output variable affects the movement in transportation hubs input variable. To control for this endogeneity in the input variable, an instrumental variable – the natural logarithm of the stringency index – is introduced to the model. The stringency index is a valid instrument for the Google transport movement index input, as it can only affect the number of cases output variable through the reduction in movement through transport hubs, being the input variable. It should be noted that the relationship between the endogenous Google movement index for transport and the stringency index instrumental variable is negative because as the government becomes more stringent, people move away from transport hubs due to closure of these hubs.

Model 2: Stay at Home Restrictions

$$\ln(\#cases\ per\ million) = \beta_0 + \beta_1 \ln(google\_residential\_index) + \ln(\varepsilon_i) + v_i$$

In this scenario, we are assessing the TE with which countries used stay at home restrictions to contain the pandemic. As such we regress the natural logarithm of new cases per million population (output variable) on the Google movement index for residential areas (input variable). We note that the input variable is endogenous as the number of new cases output variable affects the movement in residential areas input variable. To control for this endogeneity in the input variable, an instrumental variable – the natural logarithm of the stringency index – is introduced to the model. The stringency index is a valid instrument for the Google residential area movement index input, as it can only affect the number of cases output variable through the increase in movement to residential areas as they remain home, being the input variable. It should be noted that the relationship between the endogenous Google movement index for residential areas and the stringency index instrumental variable is positive as the more stringent government becomes, the more people concentrate in residential areas due to stay at home restrictions.

Model 3: Vaccinations

$$\ln(\#cases\ per\ million) = \beta_0 + \beta_1 \ln(\#people\ vaccinated\ per\ 100\ 000) + \ln(\varepsilon_i) + v_i$$

In this scenario, we are assessing the TE with which countries used vaccination policy to contain the pandemic. As such we regress the natural logarithm of new cases per million population (output variable) on the natural logarithm of people vaccinated per 100,000. We note that the input variable is

endogenous as the number of new cases output variable affects the number of people vaccinated (input variable). To control for this endogeneity in the input variable, an instrumental variable – the natural logarithm of the stringency index – is introduced to the model. The stringency index is a valid instrument for the number of people vaccinated input variable, as it can only affect the number of cases output variable through the increase in the number of people vaccinated, being the input variable. It should be noted that the relationship between the endogenous number of people vaccinated, and the stringency index instrumental variable is positive as the more stringent government becomes, the more people get vaccinated.

Model 4: Workplace Closures

$$\ln(\#cases\ per\ million) = \beta_0 + \beta_1 \ln(google\_workplace\_index) + \ln(\varepsilon_i) + v_i$$

In this scenario, we are assessing the TE with which countries used closure of workplaces policy to contain the pandemic. As such we regress the natural logarithm of new cases per million population (output variable) on the natural logarithm of the google workplaces movement index. We note that the input variable is endogenous as the number of new cases output variable affects the movement of people into workplaces (input variable) through workplace closure policy. To control for this endogeneity in the input variable, an instrumental variable – the natural logarithm of the stringency index – is introduced to the model. The stringency index is a valid instrument for the google workplaces movement index input variable, as it can only affect the number of cases output variable through the decrease in the number of people in workplaces, being the input variable. It should be noted that the relationship between the endogenous movement of people through workplaces and the stringency index instrumental variable is negative as the more stringent government becomes, the less people move through workplaces due to work closure restrictions. All coefficients and the eta endogeneity results are all highly statistically significant.

Model 5: Mean TE Model

In this fifth scenario, we construct a model that takes the mean of the technical efficiency scores estimated in models 1 – 4 for each country for each day.

5.2 Tests for Normality in the Error Term

The *xtsfkk* routine assumes a normal/truncated normal distribution in the error term. It is good practice to assess the normality of the error term post estimation either visually using a histogram or using tests of skewness and kurtosis.

5.3 Stage 2 Analysis – TE Drivers

The stage 2 analysis comprises the assessment of drivers of TE. For this analysis, standard OLS regressions were used to regress the TE scores estimated in the Stage 1 SFA on a set of drivers. Note that the *Income_Level and Nurses per 1,000 population* drivers variable were dropped in the Stage-2 regression due to collinearity with the other drivers as the health expenditure and human resources for health variables are correlated with the country's income level. Robust standard errors were specified for all OLS regressions.

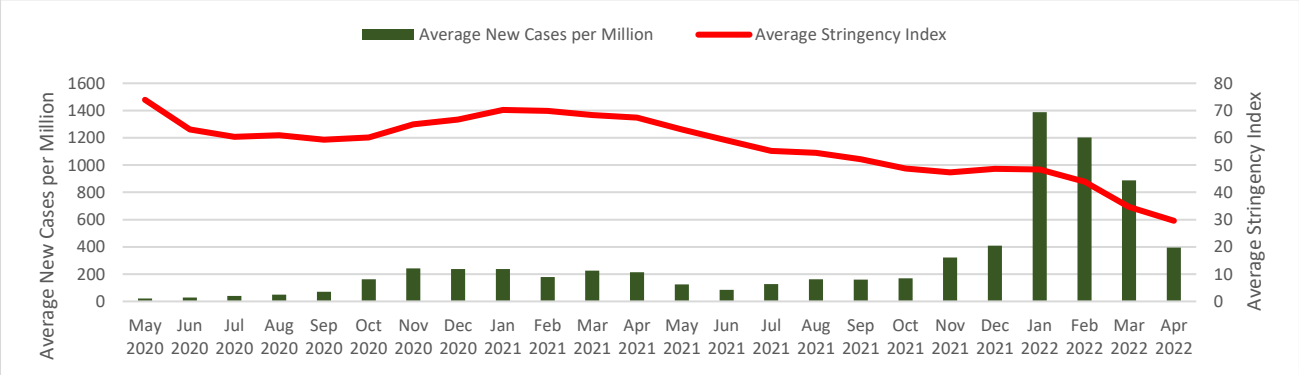
The drivers were selected a priori based specifically on health policy related inputs such as human resources, hospital beds and health budgets. These drivers also appeared in studies such as (Breitenbach, Ngobeni, & Aye, 2021), and are particularly appropriate because they provide a measure of resource mobilisation efficiency from a policy perspective, and how this differs between countries. *Beds per 1,000 population* offers a good measure of the current level of infrastructure in relation the size of the population, an important factor during the pandemic given the high demand for beds and ventilators. *Health Expenditure as a % of GDP* serves as a good proxy for the amount of funding a government has to throw at the pandemic and will affect factors such as access to vaccines etc. *Physicians per 1,000, Nurses per 1,000* and *Specialists per 100,000* together give a good idea of the availability of human resources for health in fighting COVID-19.

6. Results

Before looking at the econometric results, it is worth investigating some descriptive trends in the country data for some context. Figure 1 plots the average number of cases per million population and the average stringency index across time to investigate trends. Interestingly, it appears that stringency levels on average were fairly reactive to increases in the case rate up until the end of 2020 after which it steadily decreases. This is despite the drastic increase in the case rate around January 2022 where we actually see a dramatic decrease in the levels of stringency.

The first dip coincides with the release of the first vaccine, as governments likely reduced stringency as the vaccine was successfully distributed. The second dip around the January 2022 wave is harder to explain and may be as a result of countries coming to grips with the fact that strict national lockdowns for each COVID-19 wave may in fact be unaffordable in the long run given the economics costs associated. The spike in early 2022 new cases is around the time of the emergence of the omicron strain of the COVID-19 virus.

Figure 1 - Average New Cases per Million vs Average Stringency Index Trends



The econometric results will be presented in the two stages, the first presenting the results of the SFA regression models and final TE scores, and the second presenting the results of the exogenous driver regressions.

6.1 Stage 1: SFA Results

6.1.1 Exogenous vs. Endogenous Results

Model 1 SFA Regression Output

Endogenous stochastic production frontier model with normal/half-normal specification

Model EN log likelihood = -12738.749 Number of obs = 3,848

frontier_y	Coefficient	Std. err.	z	P>z	[95% conf.	interval]
log_Transport	.6125523	.1131702	5.41	0.000	.3907428	.8343618
_cons	5.421036	.3208632	16.90	0.000	4.792156	6.049917

eta Endogeneity Test

Ho: Correction for endogeneity is not necessary.

Ha: There is endogeneity in the model and correction is needed.

(1) [/]eta1_log_Transport = 0

chi2(1) = 59.43

Prob > chi2 = 0.0000

Result: Reject Ho at 0.1% level.

Exogenous stochastic production frontier model with normal/half-normal specification

Model EX log likelihood = -6278.6016 Number of obs = 3,848

frontier_y	Coefficient	Std. err.	z	P>z	[95% conf.	interval]

log_Transport	-.255139	.0202706	-12.59	0.000	-.2948686	-.2154093
_cons	6.750997	.2470942	27.32	0.000	6.266702	7.235293

From the output tables above, we see that correction for endogeneity is in fact required. In the eta endogeneity test, we therefore reject the null hypothesis that correction for endogeneity is not necessary in favour of the alternative hypothesis that there is endogeneity and correction is needed. We also see that the *log_transport* coefficient changes from -0.255 under the exogenous model to 0.613 after correcting for endogeneity, indicating a positive relationship between mobility around transport hubs and number of cases. The coefficients are highly statistically significant in both models.

Model 2 SFA Regression Output

Endogenous stochastic production frontier model with normal/half-normal specification

Model EN log likelihood = -83726.021 Number of obs = 26,318

frontier_y	**Coefficient**	**Std. err.**	**z**	**P>z**	**[95% conf.**	**interval]**
log_Residen	-.3327855	.0305024	-10.91	0.000	-.3925691	-.273002
_cons	6.721225	.1353928	49.64	0.000	6.45586	6.98659

eta Endogeneity Test

Ho: Correction for endogeneity is not necessary.

Ha: There is endogeneity in the model and correction is needed.

(1) [/]eta1_ log_Residen = 0

chi2(1) = 875.39

Prob > chi2 = 0.0000

Result: Reject Ho at 0.1% level.

Exogenous stochastic production frontier model with normal/half-normal specification

Model EX log likelihood = -53160.081 Number of obs = 26,318

frontier_y	**Coefficient**	**Std. err.**	**z**	**P>z**	**[95% conf.**	**interval]**
log_Residen	.4289754	.0148212	28.94	0.000	.3999264	.4580244
_cons	5.179303	.125966	41.12	0.000	4.932414	5.426191

In the case of stay-at-home restrictions, again we see a change in sign after correction for endogeneity as the *log_Residential* coefficient changes from 0.429 to -0.333. In the eta endogeneity test, we reject the null hypothesis that correction for endogeneity is not necessary in favour of the alternative hypothesis that there is endogeneity and correction is needed. This is indicative of the necessity for

endogeneity correction as we would expect increased movement into residential areas (stay home restrictions) to be negatively related to number of cases as families isolate at home and hence a negative coefficient. The coefficient of *log_Residential* is highly statistically significant as is the eta Endogeneity Test Prob > chi2 value.

Model 3 SFA Regression Output

Endogenous stochastic production frontier model with normal/half-normal specification

Model EN log likelihood = -55693.103 Number of obs = 15,498

frontier_y	Coefficient	Std. err.	z	P>z	[95% conf.	interval]
log_PeopleVacPer100	-.0237979	.0191112	-1.25	0.213	-.0612552	.0136595
_cons	6.138249	.1023969	59.95	0.000	5.937555	6.338943

eta Endogeneity Test

Ho: Correction for endogeneity is not necessary.

Ha: There is endogeneity in the model and correction is needed.

(1) [/]eta1_log_PeopleVacPer100 = 0

chi2(1) = 101.07

Prob > chi2 = 0.0000

Result: Reject Ho at 0.1% level.

Exogenous stochastic production frontier model with normal/half-normal specification

Model EX log likelihood = -28604.433 Number of obs = 15,498

frontier_y	Coefficient	Std. err.	z	P>z	[95% conf.	interval]
log_PeopleVacPer100	.1487133	.008178	18.18	0.000	.1326846	.164742
_cons	5.496784	.0754584	72.85	0.000	5.348888	5.644679

Looking at the *log-PeopleVacPer100* coefficient, again we see a change in sign after correcting for endogeneity from 0.149 to -0.02. The corrected coefficient indicates that we would expect an increase in the number of people vaccinated to be associated with a decrease in the number of cases. Again, we see that correction for endogeneity is necessary to obtain an unbiased coefficient. In the eta endogeneity test, we reject the null hypothesis that correction for endogeneity is not necessary in favour of the alternative hypothesis that there is endogeneity and correction is needed.

Model 4 SFA Regression Output

Endogenous stochastic production frontier model with normal/half-normal specification

Model EN log likelihood = -10397.67 Number of obs = 3,191

frontier_y	Coefficient	Std. err.	z	P>z	[95% conf.	interval]
log_WorkPlace	-3.199328	.6784276	-4.72	0.000	-4.529021	-1.869634
_cons	13.08852	1.480127	8.84	0.000	10.18752	15.98951

eta Endogeneity Test

Ho: Correction for endogeneity is not necessary.

Ha: There is endogeneity in the model and correction is needed.

(1) [/]eta1_log_WorkPlace = 0

chi2(1) = 19.27

Prob > chi2 = 0.0000

Result: Reject Ho at 0.1% level.

Exogenous stochastic production frontier model with normal/half-normal specification

Model EX log likelihood = -5190.8335 Number of obs = 3,191

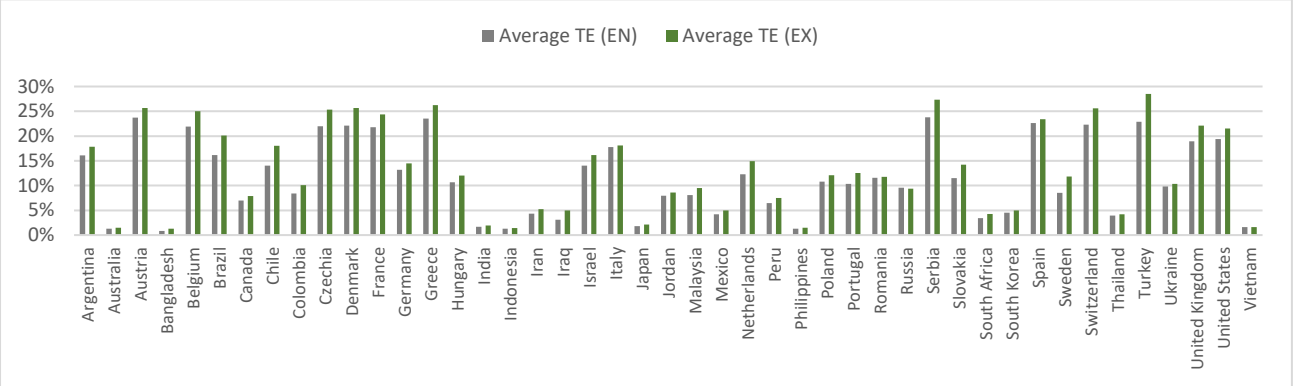
frontier_y	Coefficient	Std. err.	z	P>z	[95% conf.	interval]
log_WorkPlace	-.3133919	.0200465	-15.63	0.000	-.3526824	-.2741014
_cons	6.421525	.1523144	42.16	0.000	6.122995	6.720056

Considering the *log_WorkPlace* coefficient, which describes the closure of workplaces policy, we see that there is no change in sign, but after correcting for endogeneity, the coefficient becomes much larger (by a factor of around 10). This indicates that the model that does not control for endogeneity would have overestimated the effect of workplace closure on number of daily new cases. As expected, we see a negative relationship between *log_WorkPlace* and the output variable number of cases as avoid crowded offices and new daily cases fall as a result.

The Significance of Correction for Endogeneity

Figure 2 plots the Model 5 Mean TE scores for the exogenous and endogenous results. It is clear that the exogenous results are in all cases higher than the endogenous results indicating that the exogenous model overestimates TE on average.

Figure 2 - Mean Technical Efficiency Score per Country - Endogenous vs. Exogenous Models



To ensure that the difference between the exogenous and endogenous TE scores is statistically significant, a one-sample t test interval test was run on the difference between the two scores.

Table 2 – T test of Transport Difference Variable

```
One-sample t test
```

Variable	Obs	Mean	Std. err.	Std. dev.	[95% conf.	interval]
te_1_d~f	**3,848**	**.0826944**	**.0013189**	**.0818162**	**.0801085**	**.0852802**

```
    mean = mean(te_1_diff)                                    t =  62.6981
H0: mean = 0                                     Degrees of freedom =     3847

    Ha: mean < 0                 Ha: mean != 0                 Ha: mean > 0
 Pr(T < t) = 1.0000         Pr(|T| > |t|) = 0.0000          Pr(T > t) = 0.0000
(6,532 missing values generated)
```

Table 3 – T test of Stay-at-Home Difference Variable

```
One-sample t test
```

Variable	Obs	Mean	Std. err.	Std. dev.	[95% conf.	interval]
te_2_d~f	**26,318**	**.0360726**	**.0003217**	**.0521904**	**.0354421**	**.0367032**

```
    mean = mean(te_2_diff)                                    t = 112.1278
H0: mean = 0                                     Degrees of freedom =    26317

    Ha: mean < 0                 Ha: mean != 0                 Ha: mean > 0
 Pr(T < t) = 1.0000         Pr(|T| > |t|) = 0.0000          Pr(T > t) = 0.0000
```

Table 4 - T test of Vaccine Difference Variable

One-sample t test

Variable	Obs	Mean	Std. err.	Std. dev.	[95% conf.	interval]
te_3_d~f	15,498	.0294249	.0003003	.0373826	.0288363	.0300135

mean = mean(**te_3_diff**) t = **97.9901**
H0: mean = **0** Degrees of freedom = **15497**

Ha: mean < **0** Ha: mean != **0** Ha: mean > **0**
Pr(T < t) = **1.0000** Pr(|T| > |t|) = **0.0000** Pr(T > t) = **0.0000**

Table 5 - T test of Workplace Closure Difference Variable

One-sample t test

Variable	Obs	Mean	Std. err.	Std. dev.	[95% conf.	interval]
te_4_d~f	3,191	.088154	.0013863	.0783097	.0854359	.0908721

mean = mean(**te_4_diff**) t = **63.5902**
H0: mean = **0** Degrees of freedom = **3190**

Ha: mean < **0** Ha: mean != **0** Ha: mean > **0**
Pr(T < t) = **1.0000** Pr(|T| > |t|) = **0.0000** Pr(T > t) = **0.0000**

Table 6 - T test of Mean Model Difference Variable

One-sample t test

Variable	Obs	Mean	Std. err.	Std. dev.	[95% conf.	interval]
te_5_d~f	29,192	.0402572	.0002621	.0447801	.0397435	.0407709

mean = mean(**te_5_diff**) t = **153.5996**
H0: mean = **0** Degrees of freedom = **29191**

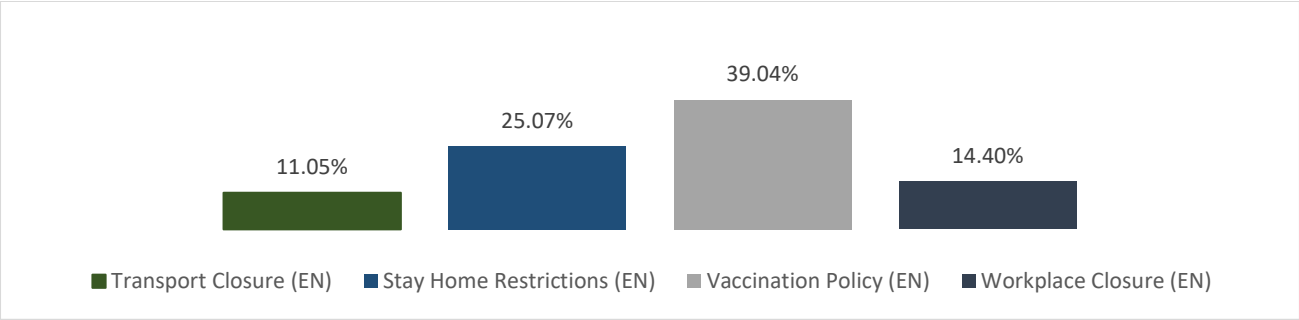
Ha: mean < **0** Ha: mean != **0** Ha: mean > **0**
Pr(T < t) = **1.0000** Pr(|T| > |t|) = **0.0000** Pr(T > t) = **0.0000**

From the t-tests results in Table 3 – Table 7, we see that all difference variables are statistically significantly different from zero as we reject the null hypothesis in all cases. This is evidence that the difference between the TE score that corrected for endogeneity and those that did not is statistically significant, and there is therefore an argument that correction for endogeneity in the input variables is indeed necessary to obtain unbiased estimates of TE.

6.1.2 NPI Channel Results

The key focus of this study is to estimate the technical efficiency with which the sample countries responded to the pandemic through different policy measures. There were 4 models, each of which investigated a different NPI channel:

Figure 3 - Average Technical Efficiency Score per Policy Model



Looking at mean TE score per NPI channel in Figure 3, we see that vaccination policies recorded the highest (39%) of the four, followed by Stay Home Restrictions (25%). Closure of workspace (14%) and transport hubs (11%) were less effective when measured by comparative technical efficiency. These findings are in line with historical trends and expectations.

Figure 4 - Mean TE Score per NPI Channel vs. Mean Cases per Million Trend for 2021

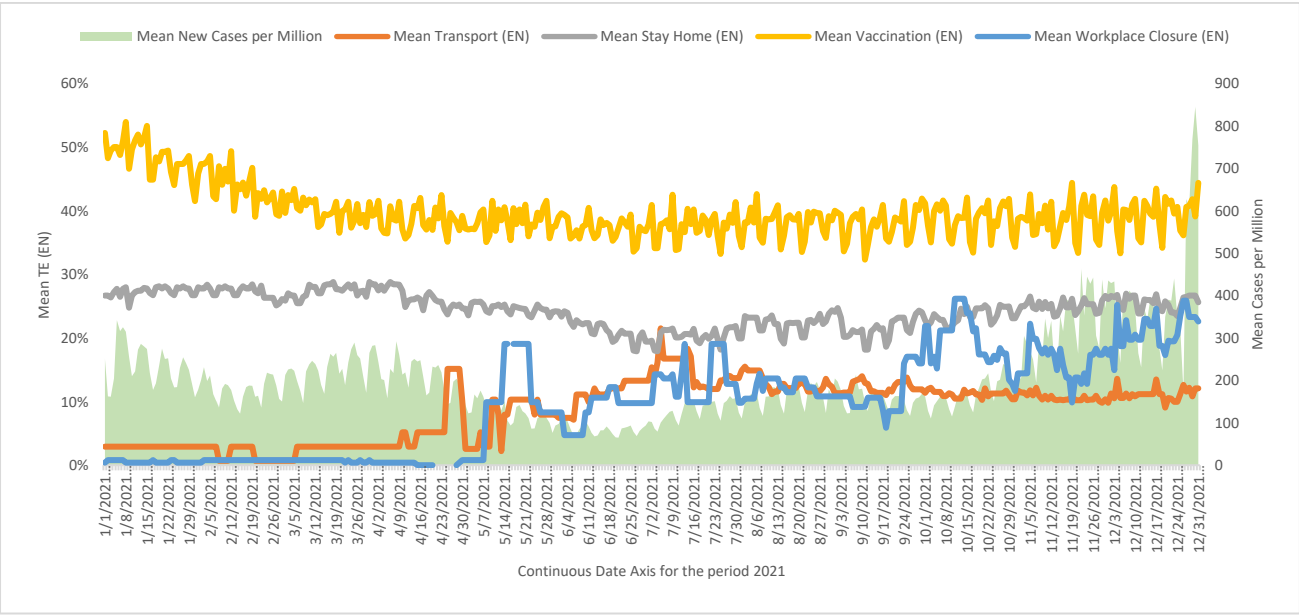
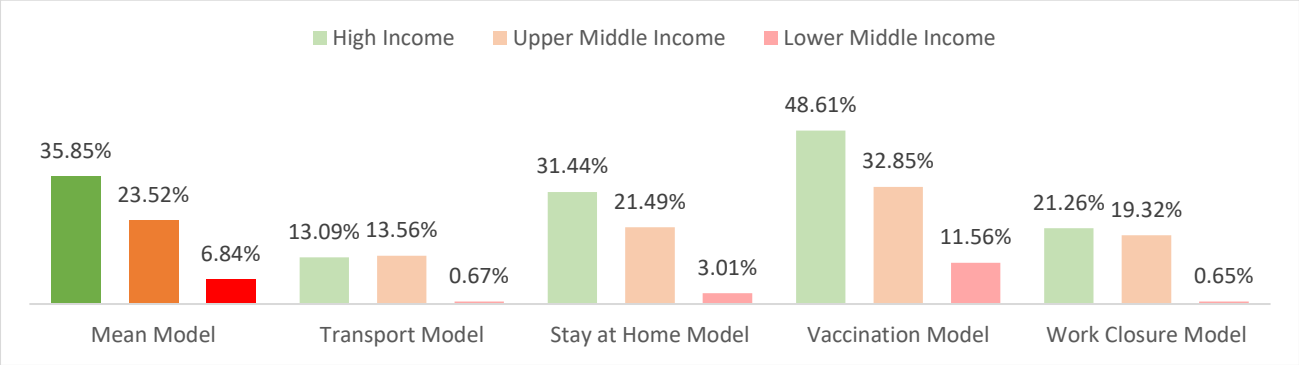


Figure 4 plots the mean TE scores for each NPI channel against the mean new cases per million daily trend to illustrate temporal fluctuation in TE. From the graph we see that vaccination lies above the other three NPI trend lines and is a clear leader in TE. The remaining three NPI trends display a more erratic pattern crossing each other at certain points in time. Up until around mid-April TE trends were mostly stable except for a few dips in transport and workplace closure TE. After May, these two TE

scores begin to grow, with growth in transport closure TE tapering off toward the end of the year, while stay at home TE remains fairly constant throughout the year.

6.1.3 Income Level Results

Figure 5 - Mean TE by Income Level: Mean Model + Individual Models



Looking at TE scores by model in Figure 5, we see that the Mean Model TE (which comprises the average of the four models) is 35.9% for high income countries, while upper and lower middle-income countries scored 23.5% and 6.9% respectively.

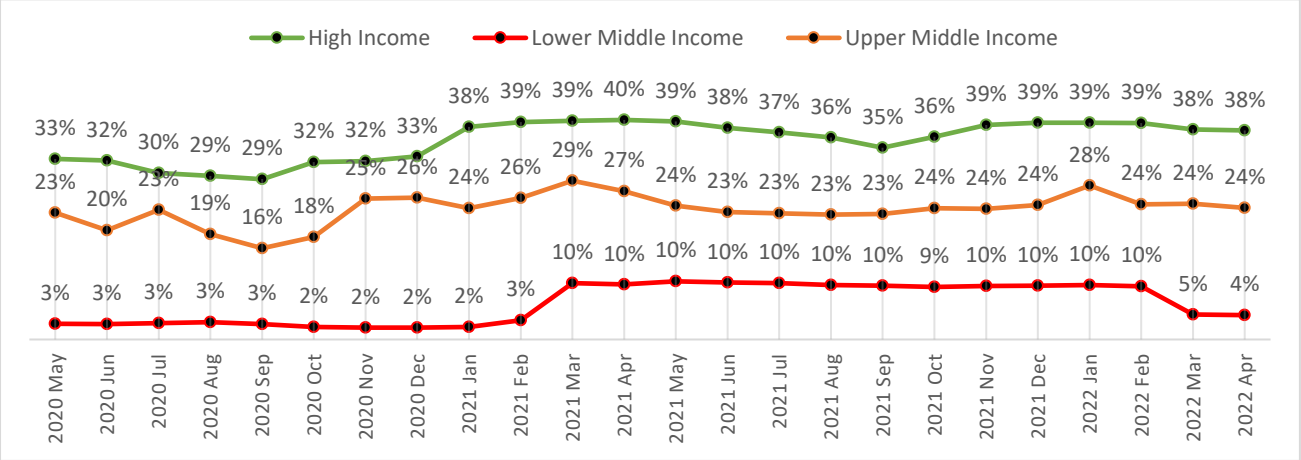
Looking at the individual models, we see similar trends across each of them, with high income countries scoring higher than upper and lower middle-income countries in descending order. The only exception is the transport model in which upper middle-income countries score better than high income countries.

The vaccination model scored the highest overall with TE scores of 48.6% (high income), 32.9% (upper middle-income), and 11.6% (lower middle-income) respectively. Lower middle-income countries scored particularly low on the transport and work closure models. This is likely owing to informal sectors of employment and transport systems being much more prevalent in lower income countries, giving these governments less control over them when trying to implement NPIs like closing transport hubs or workplaces.

Figure 6 plots the mean TE scores for each income level over time to illustrate any temporal fluctuation in TE. High income countries display more consistency with mean TE scores of around 32% in 2020 and around 39% afterward with a small dip in September 2021.

In contrast, upper middle-income countries displayed a more erratic trend indicating a higher policy reactiveness to COVID-19 waves. With the January 2022 wave being particularly high in the sample data, it is interesting to see that only upper-middle income countries ramped up TE during this time further indicating this cohort's reactiveness to changes in the number of cases.

Figure 6 - Mean Technical Efficiency Scores by Income Level and Date



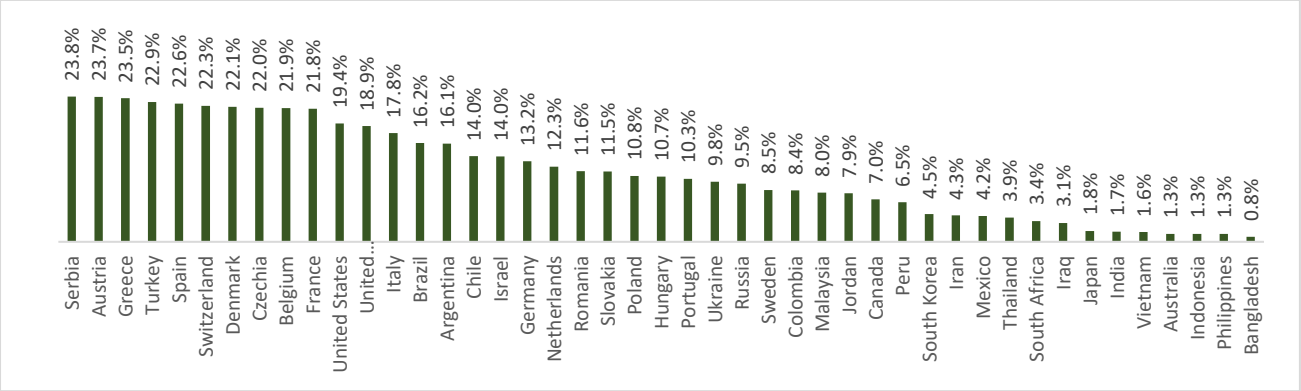
Lower middle-income countries also ramped up TE in early 2021, jumping from around 3% to 10% for the majority of the sample period. Both upper and lower middle-income countries began reducing TE in late 2022, particularly lower-income. This is likely due to the costly burden of NPI implementation over extended periods of time that lower middle-income countries cannot afford.

6.1.4 Country Level Results

The country level results provide two dimensions of analysis, the first comparative between countries, and the second considering the temporal effects of TE against waves of COVID-19 in select countries.

From a comparative perspective, Serbia, Austria and Greece all scored the highest with TE scores just over 23%. On the lower end of the spectrum are Bangladesh, Philippines, and Indonesia, all scoring under 1.4% mean TE.

Figure 7 - Mean Technical Efficiency Score per Country



Due to the fact that waves occurred in different countries at different times, it is difficult to see a sample wide trend of TE against number of cases. This is because the number of cases trend is averaged out over time due to differing waves in each region at different times. As such, it is more conducive to look at a few specific country casers instead. In this regard, one high income, two upper middle-income

and one lower middle-income country has been chosen, each from a different continent and with a sizeable population. The countries chosen are the USA as the high-income country, South Africa and Serbia as the upper middle-income country, and Bangladesh as the lower middle-income country.

In Figure 8, the high-income case of the USA, we see that for most of 2020, an average rate of 11% TE was maintained. At the end of November 2020, in response to the first wave, TE was rapidly boosted up to 23% where it stayed for the remained of the period under consideration. This sort of rapid upgrading of TE followed by long periods of sustained efficiency are typical of the larger high-income countries like the US or UK with larger health budgets and strong health institutions. The mean TE over the period for the USA was 19.3%.

Figure 8 - Mean New Cases per Million vs Mean TE Trend: USA

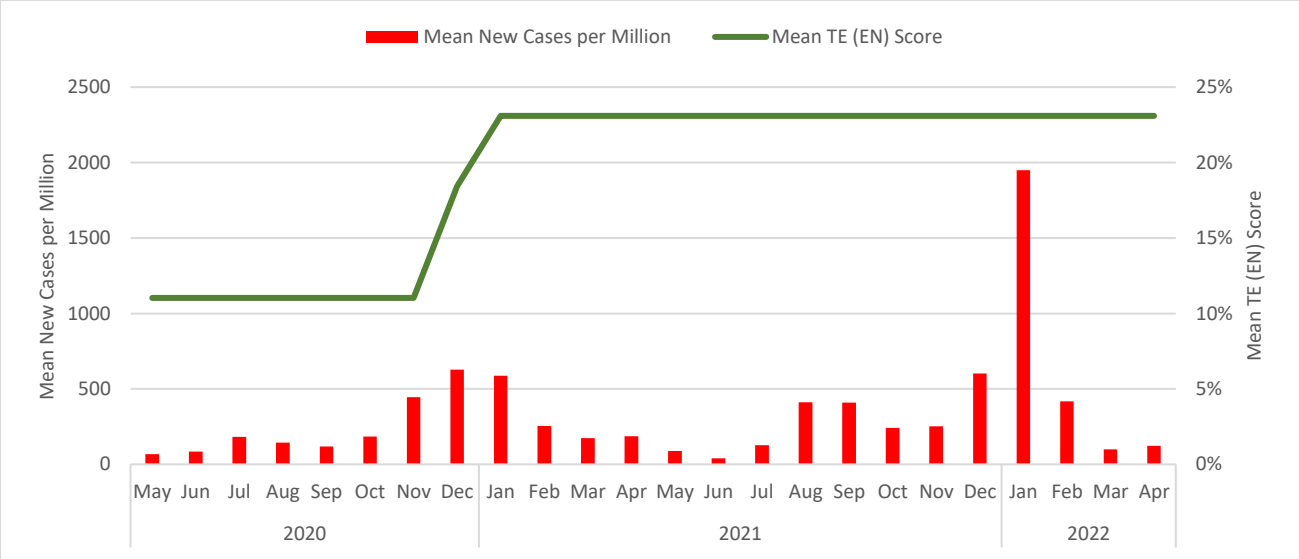
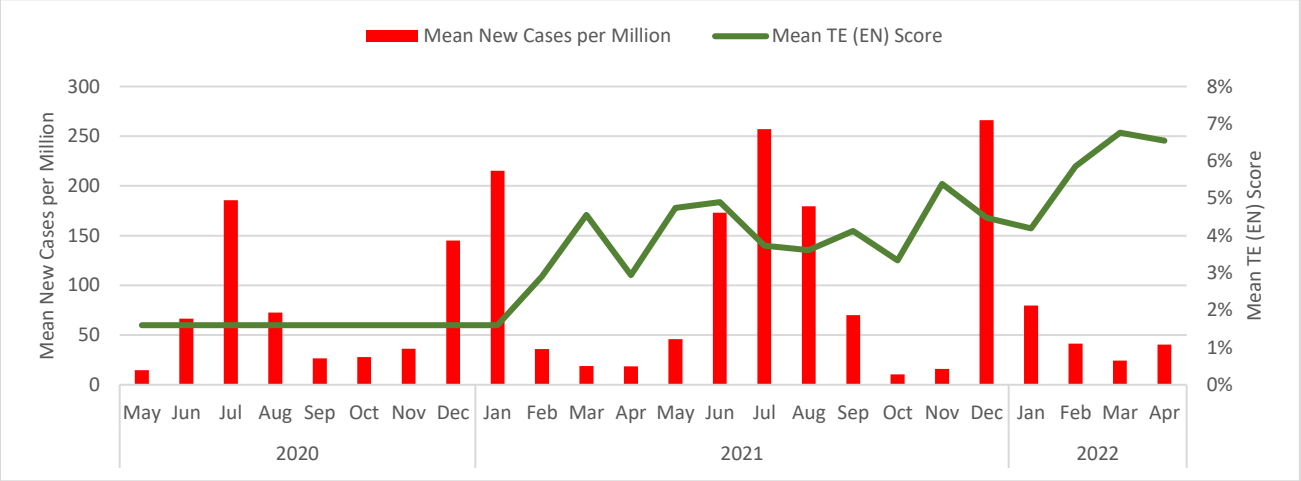


Figure 9 depicts the first upper middle-income country case, namely South Africa. In this case, we see a much more erratic trend in mean TE over the period. Reaching only a maximum of 6.8% in March 2022, South Africa operated at a far lower lever than the US on average. With an initial TE score of around 1.5% in 2020, South Africa only ramped TE up in January 2021 after its second wave. In contrast to the US who increased TE at the time of the wave, South Africa seems to display a consistent lag in policy response, as the increases in TE comes directly after the four waves. In some cases, we even see a definite decrease in TE during a wave followed by a dramatic increase in TE as the wave starts to taper off.

Figure 9 - Mean New Cases per Million vs Mean TE Trend: South Africa



In Figure 10, the Serbian cases, the country that scored the highest overall mean TE score of 23.7%, we see a highly reactive government, with peaks in TE mirroring peaks in the number of cases. This trend is from early on in the pandemic and persists for the duration of the period under consideration.

Figure 10 - Mean New Cases per Million vs Mean TE Trend: Serbia

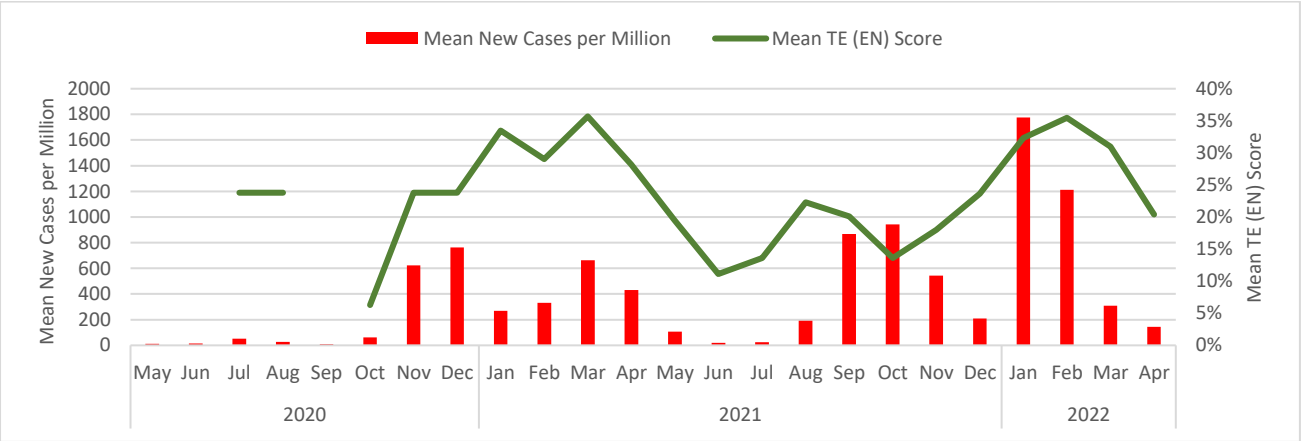
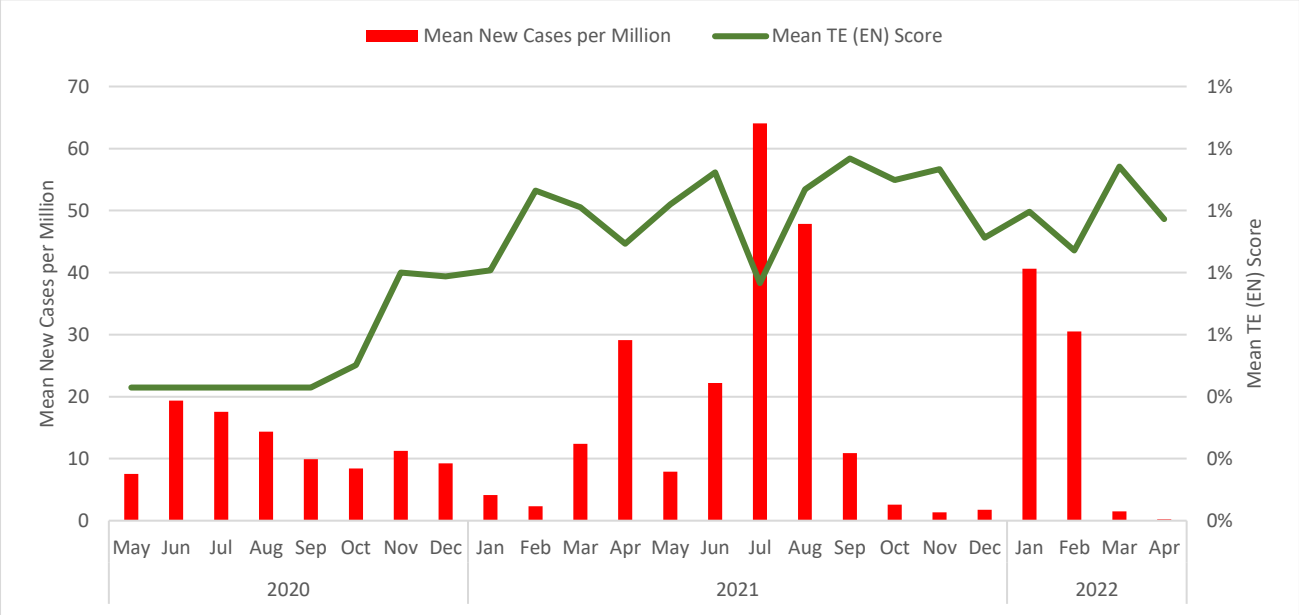


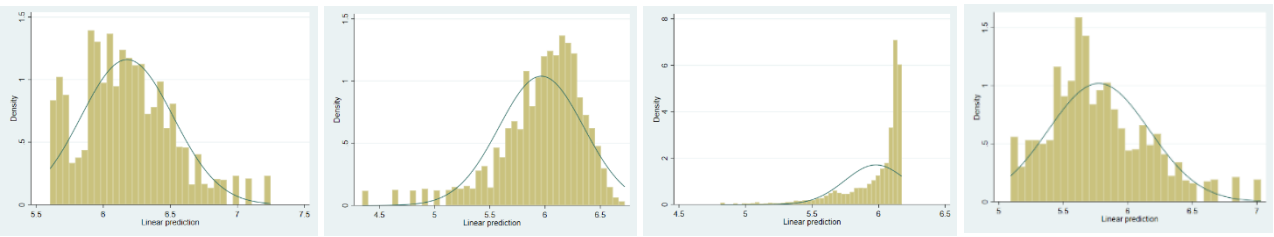
Figure 11, the case of the lower middle-income country depicts the Bangladeshi case, the lowest scoring country in mean TE. Similar to South Africa, but in contrast to the high-income case of the US, Bangladesh displays a greater degree of variance in TE over time, indicating a government or set of policies that is more reactive to fluctuations in the number of cases per day. Getting a fairly early start, the first noticeable spike in TE occurs in September 2020. In July 2021 during the country's worst wave, TE dropped noticeably to under 1%. Bangladesh scored a mean TE scored of 0.8% across the period under analysis.

Figure 11 - Mean New Cases per Million vs Mean TE Trend: Bangladesh



6.2 Tests for Normality in the Error Term

Figure 12 - Normal Distribution: Residual Density Plot (transport, stay at home, vaccine, work closure)



From Figure 12, we can see that besides the vaccine model residual plot (third panel from the right), the errors are adequately normal in distribution, with some degree of skewness in each. The models are therefore appropriate for use with the *xtsfkk* routine which uses the normal/truncated normal distribution.

6.3 Stage 2: TE Drivers Results

In the second stage regression, the mean annual TE score estimated in the first stage was regressed on four exogenous variables for each of the five models. Table 7 displays the regression results of a standard OLS regression for each of the five models.

Table 7 - Stage 2 Regression Results: Drivers of Technical Efficiency

Variable	Transport	Stay At Home	Vaccine	Work Closure	Mean
Health as a % of GDP	0.0229	0.00711	0.0284**	-0.0576**	0.0128**
	(1.92)	(1.42)	(3.21)	(-3.12)	(2.99)
Beds per 1000	-0.0224	0.0257**	0.00407	-0.0321	0.0206*
	(-1.19)	(2.73)	(0.36)	(-0.93)	(2.59)
Specialists per 100 000	0.00167*	0.00277***	0.00186*	0.00599	0.00193**

	(2.60)	(3.62)	(2.10)	(1.68)	(3.22)
Physicians per 1000	-0.00567	-0.0257	0.0189	0.172**	0.00225
	(-0.18)	(-1.33)	(0.85)	(3.87)	(0.16)
Observations	25	72	57	18	72
R-squared	0.265	0.346	0.326	0.610	0.433
Adjusted R-squared	0.118	0.307	0.274	0.490	0.400

t statistics in parentheses
** p<0.01, *** p<0.001

Health expenditure as a percentage of GDP was statistically significant at the 1% level for vaccinations, workplace closures and the mean model. The coefficient was also positive at 0.0128 indicating that higher levels of GDP expenditure are associated with higher levels of TE. Similarly, the number of beds per 1,000 population was also positively related to TE scores. This result was statistically significant for the Stay at Home and Mean models with an average result of 0.02. Specialists per 100,000 population was statistically significant for the transport, stay at home and vaccine models as well as the mean model with positive coefficients across the board. This indicates that an increase in either of these variables is associated with and increase in TE as one may expect. Note that the income level and nurses per 1000 population were omitted due to collinearity.

7. Discussion

The results that pertain to the significance of correction for endogeneity lie at the heart of this thesis, as all subsequent analysis and results depend on its outcome. As such it makes a logical point of departure in the discussion of the results.

In the first stage, the SFA model was run for each NPI channel under two scenarios, ones that corrected for endogeneity in the input variables (referred to as the endogenous case), and one that did not (referred to as the exogenous case). To decide whether endogeneity indeed existed in the input variables, the eta tests for endogeneity was run as part of the SFA regression, a test that uses the same principles as the Durbin Wu-Hausman test for endogeneity (Karakaplan & Kutlu, 2017). The regression results from Section 6.1.1 show that in all four models, endogeneity did exist and needed to be corrected. Having's shown that, the study then proceeded to test whether the difference between the endogenous and exogenous TE scores was in fact statistically significant for each model. This was done using a one sample t-test, and in all cases proved the correction for endogeneity to be statistically significant, further supporting the evidence in favour of the endogenous case. This was an important result, as it shows that in studies which assess how COVID-19 containment policy affects the number of daily cases, the policy input variables are likely endogenous due to the effect of case rate on policy

intensity. It was shown that the model that did not correct for endogeneity over estimated TE score by about 4% based on the mean model results.

As an aside, the study made use of some of the latest available data on COVID-19 policy interventions and population behavioural responses to these policies around the world. This makes them of huge value to econometricians interested in COVID-19 policy analysis. The study was made possible by the Google movement index data which enabled one to model how effective governments were at reducing movement through key areas such as workplaces or transport hubs, or how effective stay at home restrictions were based on residential concentrations of people. The stringency index was another pivotal dataset that enriched the study by providing an appropriate instrument to correct for endogeneity in the input variables. Both of these datasets are high frequency and available at the daily grain, stretching over the entire pandemic. It is highly recommended that these datasets be applied to a variety of other policy related COVID-19 studies such as DEA applications as they are rich panel datasets which enable both dimensional and temporal analysis of policy efficiency. In contrast, the data on drivers that enabled the second stage was unfortunately sparsely available with little inter-temporal variation. This limited the use of panel data as mean values were calculated for many countries for these driver variables. This was the main data limitation encountered in the study, and perhaps further research that looks at new drivers of TE with more detailed data at the daily level is required.

Still looking at first stage SFA results, Section 6.1.2 presented the TE results for the various NPI and vaccine policy channels. From this perspective, one of the most obvious outcomes was that the race to develop a vaccine was indeed warranted as vaccination consistently proved to be the most technically efficient intervention when compared to the other NPIs considered in this study. Similarly, mean technical efficiency was greatest in high income countries, in most cases, except for the closure of transport hubs, in which upper middle-income countries were slightly more efficient than high income countries. One reason for this could be the higher reliance of high-income countries on high frequency public transport hubs such as underground trains and airports. This makes it more difficult to implement complete closure of these hubs and reduces the efficiency of this channel of policy intervention when compared with upper-middle income countries. Interestingly, the TE of transport policy was also the only one to show positive growth in the longer term.

The country level results also offered some interesting insight into the performance of high, upper, and lower-middle income countries. The key take-aways from the analysis were that large and wealthy countries such as the United States or United Kingdom with abundant resources and funding ramped

up TE early and maintained high levels of TE despite the drops in cases between waves. This was also seen in the constant and high levels of TE scores seen in the high-income country TE trends depicted in Figure 6. The results also line up with findings by (Breitenbach, Ngobeni, & Aye, 2021, p. 1) - "it is suspected that most countries literally "threw" resources at fighting the pandemic, thereby probably raising inefficiency through wasted resource use."

On the contrary, upper-middle income countries such as Serbia (the highest scoring country overall) displayed highly reactive trends in mean TE where rises and falls in cases per day were mirrored closely by changes to policies and hence the level of stringency and TE. In comparing the two cases, Serbia's reactive approach to setting policy is probably what enabled them to achieve the highest TE score as the approach reduces the economic costs of interventions by ensuring that the minimum required intervention is implemented at any given time. This is an important takeaway for all countries, especially given the devastating economic impact that lockdown intervention have caused for economies around the world. In future, countries should aim to be as reactive/flexible as possible when setting policy of this nature to reduce the detrimental impacts of restrictions on the economy over extended periods of time.

In the second stage, data limitations pertaining to the drivers of TE forced the aggregation of the dataset to the annual level with average TE scores per annum. As such a panel model was no longer viable, and an OLS regression was used instead. The *Health Expenditure as a % of GDP* variable gives a good idea of the size and reach of a country's national health budget. The mean model coefficient of 0.0206 is highly statistically significant and indicates that larger health budgets are associated with higher levels of TE on average. This is likely due to these countries being able to purchase and distribute the vaccine quickest given their financial resources. *Beds per 1000* was chosen as a driver because it gives a good idea of the level of infrastructure relative to the size of the population, a factor that became very important given the demand for hospital beds and ventilators during the pandemic. As one might expect, the mean model shows a positive coefficient of 0.02 and is highly statistically significant. Lastly, the *Specialists per 100 000* variable displayed positive coefficients across all models, with only the workplace coefficients being statistically insignificant. At the mean model, the coefficient of 0.0012 is very small, but is positive non the less, indicating that the number of specialists had some small positive influence in the levels of TE. While the coefficients were mostly statistically significant, the results should be considered alongside the data limitations stated, as these may have compromised the accuracy of the indicators so some extent. Further research into the drivers of COVID-19 policy efficiency that uses data that exists at the daily level would hugely improve these estimates.

On the subject of potential extensions or improvements to this research, one exciting new application to frontier efficiency modelling is Machine Learning (ML). Although ML has been applied to some cases of DEA modelling, such as (Taherinezhad & Alinezhad, 2022) who run a performance evaluation on country outbreak handling during COVID-19, it has never been done in the SFA case. It is recommended that ML be incorporated into this or a similar SFA study to enrich the results and insights produced and better enable countries in their response to future pandemics.

8. Conclusion

This study estimated the TE of both pharmaceutical and non-pharmaceutical country responses to the COVID-19 pandemic using a Stochastic Frontier Analysis and a panel dataset combining indicators from the Our World in Data (OWID), Google Movement Index and World Bank Economic Indicators databases. A second stage OLS regression was run to investigate the effects of a priori driver on TE. Each of the four input variables were endogenous and required correction for endogeneity to obtain unbiased estimates. It was shown that the difference between the results for inputs corrected for endogeneity, and those not, was statistically significant proving that correction for endogeneity was indeed necessary. The results that did not correct for endogeneity overestimated TE scores by around 4% on average across the models. Other key insights were that vaccination policies proved to be the most technically efficient channel of prevention followed by stay-at-home restrictions. The drivers of TE analysis returned satisfactory results, but limitations in data mean that there is room for improvement in assessing the drivers in future. Lastly it was recommended that ML techniques be applied to the SFA regression in future research as this is an exciting new area of TE analysis that would benefit hugely from improved accuracy and additional insights.

END

References

Ahmad, W., Zainon, W., Nawi, M., & Ghazali, F. (2020). Malaysia's Efficiency in Dealing with COVID-19 Outbreaks Compared to Other Asian Countries by Using Stochastic Frontier Analysis (SFA). *Annals of King Edward Medical University*, 324-329.

Battese, G., & Coelli, T. (1992). Frontier Production Functions, Technical Efficiency and Panel Data: With Application to Paddy Farmsers in India. *Journal of Productivity Analysis*, 153-169.

Breitenbach, M., Ngobeni, V., & Aye, G. (2020). Efficiency of Healthcare Systems in the first wave of COVID-19 - a technical efficiency analysis. *Munich Personal RePec Archive*.

Breitenbach, M., Ngobeni, V., & Aye, G. (2021). Global Healthcare Resource Efficiency in the Management of COVID-19 Death and Infection Prevelance Rates. *Frontiers in Public Health*.

Fakir, A., & Bharati, T. (2021). Pandemic catch-22: The role of mobility restrictions and institutional inequalities in halting the spread of COVID-19. *PLoS ONE*.

Karakaplan, M. (2022). Panel stochastic frontier models with endogeneity. *Stata Journal*, 643-663.

Karakaplan, M., & Kutlu, L. (2017). Endogeneity in panel stochastic frontier models: an application to the Japanese cotton spinning industry. *Applied Economics*.

Karakaplan, M., & Kutlu, L. (2017). Handling Endogeneity in Stochastic Frontier Analysis. *Economics Bulletin*, 889-901.

Labrecque, J., & Swanson, S. (2018). Understanding the Assumptions Underlying Instrumental Variable Analyses: a Brief Review of Falsification Strategies and Related Tools. *National Library of Medicine*, 214-220.

Schmidt, P. (2011). One-step and two-step estimation in SFA models. *Journal of Productivity Analysis*, 201-203.

Shirouyehzad, H., & Jouzdani, J. (2020). Fight against COVID-10: A Global Efficiency Evaluation based on Contagion Control and Medical Treatment. *Journal of Applied Research on Industrial Engineering*, 109-120.

Taherinezhad, A., & Alinezhad, A. (2022). Nations performance evaluation during SARSCoV-2 outbreak handeling via data envelopment analysis... *International Journal of Systems Science: Operations & Logistics*,.

Wang, H., & Schmidt, P. (2002). One-Step and Two-Step Estimation of the Effects of Exogenous Variables on Technical Efficiency Levels. *Journal of Productivity Analysis*, 129-144.

Worthington, A. (2003). Frontier Efficiency Measurement in Health Care: A Review of Empirical Techniques and Selected Applications. *Medical Care Research and Review*, 135-170.