Data Task Write-Up

Hude Hude

November 6, 2024

This is a complete write-up for the data task, revised and expanded from the initial submission that was completed on November 4 within the required time constraint.

Understanding the Task and Key Assumption

In this analysis, we aim to investigate whether financial constraints in U.S. hospitals have led to a shortage in technology adoption. Specifically, we want to determine if financial support—through increased revenue enabling hospitals to invest in technology—leads to an increase in both the quantity and quality of technology adoption. To identify this effect, we leverage the Medicare Modernization Act (MMA) as a quasi-experimental setting. This is because the policy selected hospitals for reimbursement rate adjustments purely based on political considerations. Then, from the hospitals' perspective, the policy acts as an exogenous shock. This assumed randomness in assignment allows us to treat the policy as a shifter, helping to isolate and identify whether financial support leads to increase in technology adoption.

Task 1: Merging Hospital Data with Treatment Status Indicator

To begin, I merged the main dataset (`datatask_main.dta`) containing hospital characteristics and technology variables with the treatment status dataset (`datatask_treat.dta`). This merge used `prov_id` as the unique identifier to link records across the two files.

Code Explanation:

- **Standardizing `prov_id`:** I observed that `prov_id` was in lowercase in one file and uppercase in the other. To address this mismatch, I converted all `prov_id` values to uppercase using `replace prov_id = upper(prov_id)`. This ensured a consistent format, allowing for accurate matching between datasets.
- **Merging:** After standardizing `prov_id`, I performed a one-to-many merge (`merge 1:m prov_id using datatask_main.dta`).

- **Excluding Unmatched Hospitals:** Some hospitals in the main dataset lacked treatment status information, meaning they did not appear in the treatment file. Since they could not contribute relevant data to our analysis, I dropped these observations.

- **Saving the Merged Dataset:** Finally, I saved the merged dataset as `datatask_merged.dta` to facilitate future tasks.

Task 2: Constructing the Saidin Index

To measure technology adoption across hospitals, I calculated a Saidin Index for each hospital in each year, following the approach described in the assignment. This index is intended to capture the extent of technology adoption and the relative rarity of each technology.

Code Explanation:

- **Sorting the Dataset:** First, I sorted the dataset by year to ensure that technology weights were calculated consistently across years.

- **Calculating Weights for Each Technology:** For each technology variable (e.g., `tech_1`, `tech_2`), I calculated the proportion of hospitals with that technology each year. The weight for each technology ($a_{k,t}$) was then calculated as $a_{k,t} = 1 -$ proportion of hospitals with technology k in year t. This step ensures that technologies adopted by fewer hospitals have higher weights.

- **Calculating the Saidin Index:** Using the weights, I calculated the Saidin Index for each hospital and year as a weighted sum of the technologies the hospital adopted. This was achieved by summing the product of each technology's presence (1 if the hospital has it, 0 otherwise) and its weight.

- **Saving the Dataset:** I saved the dataset with the Saidin Index as `datatask_saidin.dta`.

Intuition for the Saidin Index: The Saidin Index captures two main dimensions:

1. **Number of Technologies Adopted:** A hospital with more technologies has a higher Saidin Index.

2. **Rarity of Technologies:** Technologies that are less commonly adopted across hospitals contribute more to the index, meaning that hospitals adopting rare, advanced, or expensive technologies have a higher score.

Concern About the Saidin Index: A limitation of the Saidin Index is that it may inflate scores for hospitals adopting rare, low-quality technologies. For instance, a hospital with limited resources might choose a low-cost, low-quality technology that most hospitals avoid due to its limited effectiveness. This rarity would assign it a high weight, raising the hospital's index despite the technology's low value. Thus, the index may not fully capture true technological advancement if rare but low-quality technologies are included.

Task 3: Analysis of Saidin Index Distribution and Determinants of Technology Adoption in Hospitals (2004)

(a) Distribution of Saidin Index Scores

The distribution of the Saidin Index for hospitals in 2004 (Figure 1) is right-skewed, indicating that most hospitals have relatively low levels of technology adoption. The summary statistics (Table 1) reveal that the mean Saidin Index score is 1.91, with a standard deviation of 1.99, suggesting significant variability across hospitals. The scores range from a minimum of 0 to a maximum of 11.59, showing that while many hospitals have limited technology adoption, a smaller number achieve much higher scores.

The kernel density plot highlights this skewness, with a large concentration of hospitals having low Saidin Index values, and only a few hospitals with higher scores. This skewness might reflect disparities in funding, size, or strategic priorities related to technology adoption across hospitals.

(b) Hospital Types and Technology Adoption

The simple OLS regression result (Table 2) indicates that hospital type, specifically teaching status and nonprofit status, significantly affects the Saidin Index, which measures technology adoption. Teaching hospitals have a positive and statistically significant coefficient of 1.56 ($p < 0.01$), suggesting that teaching hospitals tend to have higher Saidin Index scores, indicating higher technology adoption compared to non-teaching hospitals.

Nonprofit hospitals also exhibit a positive and significant relationship with the Saidin Index, with a coefficient of 0.34 ($p < 0.01$), implying that nonprofit hospitals adopt more or rare technologies than for-profit hospitals, on average. However, government ownership does not have a statistically significant effect on the Saidin Index (coefficient = 0.03, p = 0.82), suggesting that government-owned hospitals do not differ significantly in their technology adoption levels compared to private hospitals.

These results suggest that hospital characteristics like teaching and nonprofit status are correlated with higher rates of technology adoption, while government ownership does not have a clear impact on adoption levels.

(c) Bed Count and Saidin Index Scores

The OLS regression results (Table 2) also suggest that hospitals with more beds tend to have higher Saidin Index scores, which indicates greater technology adoption. The coefficient for the variable *beds* is 0.0048, which is positive and statistically significant ($p < 0.01$). This implies that, on average, each additional bed is associated with a 0.0048 increase in the Saidin Index score, holding other factors constant. Although the effect per bed is small, it accumulates across hospitals with larger bed counts, suggesting that larger hospitals[1] may adopt more or advanced technologies compared to smaller hospitals.

However, it is important to note that OLS assumes strict exogeneity, meaning that the explanatory variables should be uncorrelated with the error term. This assumption may not hold in this context due to potential unobserved factors influencing both hospital characteristics and technology adoption. Consequently, the OLS estimates may be biased.

Despite this limitation, OLS is useful at this stage for exploring correlations and providing initial insights into the relationship between hospital characteristics and technology adoption.

Task 4: Estimating the Yearly Treatment Effects on the Saidin Index

Table 3 shows the estimation result. The specified difference-in-differences (DID) model in this task aims to identify both the treatment effect and the year fixed effects, where the treatment effect is captured by β and the year fixed effects are captured by λ. Here, λ also serves as the control group trend because, for hospitals in the control group, the treatment indicator is 0. This means that the relative change in the Saidin Index for the control group across years is entirely captured by λ. For the treatment group, the relative change in the Saidin Index is given by $\beta + \lambda$, reflecting the combined effect of the treatment and the underlying trend.

To estimate a difference-in-differences (DID) panel regression model, I did the following:

1. **Creating Interaction Terms:** I generated interaction terms for each year from 2001 to 2010 by multiplying the treatment indicator with an indicator for each year.

2. **Panel Regression Setup:** Using the `xtreg` command with fixed effects, I estimated the model, clustering standard errors by hospital (`prov_ID`). I also tried an alternative specification with `reghdfe` to account for two-way fixed effects, though this method absorbs year fixed effects and doesn't allow for direct estimation of yearly treatment effects.

[1] Here, "larger" refers to hospitals with a greater number of beds.

3. **Extracting Coefficients with `parmest`:** To automate extraction, I used the `parmest` package. However, I encountered limitations with `parmest` in selectively retrieving only the coefficients of interest and formatting them in the required layout.

Challenges and Solution: While using `parmest`, I could only output all estimated coefficients into a `.dta` file without the desired organization. To overcome this within the time constraint, I implemented an alternative approach:

- I loaded the full set of coefficients and performed additional cleaning and merging operations to isolate the specific estimates of interest.
- After creating a dataset with only the relevant coefficients, I calculated the combined estimate (`tr_mean`) as the sum of the treatment effect (`tr_effect`) and the control group trend (`cr_mean`) for each year.

This approach allowed me to achieve the goal of generating a final `.dta` file (`Hude_Hude_estimates.dta`) with the estimates organized in the desired layout, ready for further analysis and plotting.

Task 5: Plotting the Yearly Changes in Saidin Index for Control and Treatment Groups

Figure 2 shows the resulting plot. The objective of this task is to plot the average change in the Saidin Index over time, comparing the treatment and control groups, with 95% confidence intervals (CIs) for the relative change in the index for the treatment group.

1. **Data Preparation:** After loading the dataset `Hude_Hude_estimates.dta`, I sorted it by year to ensure the data points appear in chronological order on the plot.

2. **Construction of Confidence Intervals:**

    - The relative change in the Saidin Index for the treatment group in each year is captured by `tr_mean`, which is $\beta + \lambda$. Here, λ represents the yearly fixed effects, which reflect the trend for the control group, and β captures the additional effect of the treatment.
    - To represent the 95% CI for the relative change ($\beta + \lambda$), I performed a linear transformation on the lower (`tr_lo`) and upper (`tr_hi`) bounds of the CI for β. By adding the yearly control group trend (`cr_mean`, which is λ) to `tr_lo` and `tr_hi`, I created adjusted bounds `tr_lo_adj` and `tr_hi_adj`. These adjusted bounds represent the CI for $\beta + \lambda$, allowing a more accurate visual comparison of treatment and control trends with the 95% CI for the relative change in the treatment group.

3. **Plotting:**

   - I used `twoway` to plot the control group trend (`cr_mean`, or λ) as a solid blue line and the relative change for the treatment group (`tr_mean`, or $\beta + \lambda$) as a dashed red line.
   - The adjusted 95% CIs are displayed as red vertical bars around the treatment line, reflecting the uncertainty in the relative change for the treatment group.
   - The x-axis represents the years from 2001 to 2010, with 2004 as the baseline year (where parameters are normalized to zero).

Task 6: Discussion & Limitation

(a) Conclusion remarks

The plot (Figure 2) indicates a clear effect of the MMA reimbursement rate adjustments (the "treatment") on hospital technology adoption, as measured by the Saidin Index. Prior to the 2004 policy change, the treatment group's Saidin Index was significantly below that of the control group, with non-overlapping confidence intervals, suggesting that treated hospitals were initially behind in technology adoption. After 2004, this trend reverses, with the treatment group experiencing a steeper increase in the Saidin Index, eventually surpassing the control group. The separation between the treatment and control groups post-2004, along with the non-overlapping confidence intervals from 2005 onward, suggests a positive impact of the policy on technology adoption in treated hospitals.

While the graph and confidence intervals provide evidence of a positive effect, it is important to recognize that there may be threats to the model and the identification of the treatment effect. These potential limitations will be discussed in the next part.

(b) Potential Threats to the Validity of the DID Model

1. Selection Bias from Political Influence: The reimbursement rate adjustments under the Medicare Modernization Act (MMA) of 2003 were politically motivated, introducing potential selection bias into the treatment group. Since the selection was not random, hospitals receiving the adjustment may differ in unobserved ways from control hospitals. For instance, politically-connected hospitals could have greater access to resources, more established networks, or closer relationships with technology suppliers. These differences could independently influence technology adoption, confounding the estimated treatment effect. Moreover, network advantages in treated hospitals might affect access to discounted or advanced technologies and lower overall technology costs. Such systematic differences between treated and control groups could bias the estimated impact of the MMA policy adjustment.

Based on the graph (Figure 2), we observe that the selected hospitals were indeed systematically different from the non-selected hospitals before the policy took effect, as indicated by the lower Saidin Index in the treatment group prior to 2004. This initial difference suggests that the selected hospitals may not represent the general hospital population, potentially biasing the estimated impact of the MMA policy adjustment.

2. Two-Sided Market Effects and Demand Response: The analysis focuses primarily on the supply side (i.e., the hospitals), yet the policy adjustment could have also influenced demand. For instance, increased funding may create public perceptions of improved healthcare quality in treated hospitals, thereby attracting more patients. This demand-side response could drive additional revenue, which hospitals might reinvest into adopting new technologies. Consequently, the observed technology adoption may not solely result from the policy's direct effect on hospital funding; rather, it may also reflect an indirect effect through increased patient demand for higher-quality care. Ignoring this demand feedback loop could lead to an overestimation of the policy's direct impact on technology adoption.

In conclusion, the DID approach relies on two key assumptions: that the selection of hospitals for MMA adjustments was effectively random, making the treated group representative, and that the policy did not trigger a feedback loop in patient demand that could independently drive technology adoption. These assumptions are essential to attribute observed changes in technology adoption directly to the policy.

Table 1: Summary Statistics for Saidin Index in 2004

Variable	Obs	Mean	Std. Dev.	Min	Max
Saidin Index	1,182	1.9147	1.9945	0	11.5948

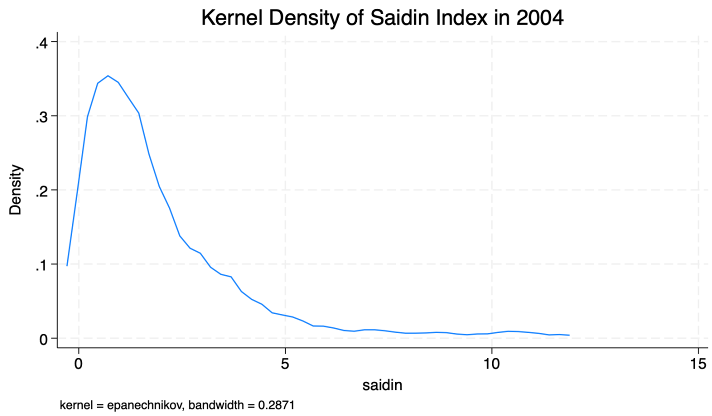

Figure 1: Kernel Density of Saidin Index in 2004

Table 2: OLS Regression: Saidin Index and Hospital Characteristics

	Saidin Index
teach	1.5558***
	(0.1762)
beds	0.0048***
	(0.0003)
nonprof	0.3431***
	(0.1301)
govt	0.0348
	(0.1503)
treat	0.4783
	(0.3239)
Constant	0.4255***
	(0.1253)
Observations	1,182
R-squared	0.4495
Adjusted R-squared	0.4471
F-statistic	192.03

Note: *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$. Standard errors are in parentheses.

Table 3: DID: Saidin Index and Yearly Treatment Effects

Year Effects		**Treatment-Year Effects**	
	Saidin Index		**Saidin Index**
Year 2001	-0.915***	*Treat_yr2001*	-1.856***
	(0.053)		(0.480)
Year 2002	-0.987***	*Treat_yr2002*	-1.682***
	(0.053)		(0.435)
Year 2003	-0.473***	*Treat_yr2003*	-0.964***
	(0.044)		(0.367)
Year 2005	0.863***	*Treat_yr2005*	0.948***
	(0.043)		(0.316)
Year 2006	1.056***	*Treat_yr2006*	1.427***
	(0.049)		(0.418)
Year 2007	1.753***	*Treat_yr2007*	1.966***
	(0.066)		(0.548)
Year 2008	1.983***	*Treat_yr2008*	2.242***
	(0.069)		(0.469)
Year 2009	2.212***	*Treat_yr2009*	2.881***
	(0.073)		(0.509)
Year 2010	2.321***	*Treat_yr2010*	2.809***
	(0.072)		(0.575)
Constant	1.934***	Observations	11,761
	(0.031)	R-squared (Within)	0.3991

Note: *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$. Robust standard errors are in parentheses. Estimated coefficients are relative to 2004.

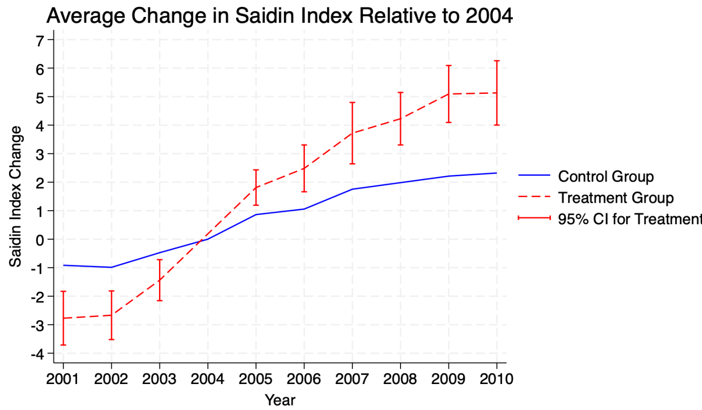

Figure 2: Comparison of Saidin Index Changes Over Time