Moral Hazard in Employer-Sponsored Health Insurance: Evidence from COVID-19*

Cameron M. Ellis† Shan Ge‡ Yiwen Lu§

August 20, 2024

Abstract

Keywords:
JEL classifications:

*We thank seminar participants at...
†Tippie College of Business, University of Iowa, cameron-ellis@uiowa.edu
‡Stern School of Business, New York University, sg3634@nyu.edu
§The Wharton School, University of Pennsylvania, yiwelu@wharton.upenn.edu

1 Introduction

2 Institutional Mechanics

3 Data

3.1 Data Sources

3.2 Sample Construction

3.3 Variable Construction

3.4 Summary Statistics

4 Methods

4.1 Differences-in-Differences Analysis

To investigate potential moral hazard effects in self-insured firms during the COVID-19 pandemic, we first employ a difference-in-differences (DiD) approach. This method compares changes between self-insured firms (treatment group) and fully-insured firms (control group) before and after the onset of COVID-19.

Our primary model is a two-way fixed effects DiD:

$$Y_{it} = \delta(\text{SelfInsure}_i \times \text{PostCOVID}_t) + \lambda_i + \theta_t + \epsilon_{it}, \tag{1}$$

where:

- Y_{it} is the outcome variable for firm i at time t, chosen to capture potential moral hazard behaviors.
- $\text{SelfInsure}_i \times \text{PostCOVID}_t$ is the interaction term representing the DiD effect.
- λ_i and θ_t are firm and year fixed effects, respectively.

- ϵ_{it} is the error term.

The coefficient of interest, δ, estimates the average treatment effect on the treated (ATT). It quantifies how the pandemic differentially impacted self-insured firms compared to fully-insured firms. A positive δ could indicate increased moral hazard behavior among self-insured firms, while a negative δ might suggest a mitigation of such behavior. This approach allows us to examine whether self-insured firms exhibited different patterns in key outcomes during the pandemic, potentially reflecting moral hazard effects. By focusing on changes in behaviors and outcomes that could be influenced by the firm's insurance status, we can gain insights into how self-insurance might affect a firm's response to a major health crisis like COVID-19.

It is important to note that, unlike a traditional DiD approach, our primary interest does not lie in estimating the overall "treatment effect" of COVID-19 on firms. Instead, our focus is on the difference in responses between self-insured and fully-insured firms, which we interpret as potential evidence of moral hazard. The coefficient captures this differential effect, allowing us to isolate how the insurance arrangement influenced firm behavior during the pandemic. By comparing the changes in outcomes between these two groups, we can identify whether self-insured firms exhibited behaviors consistent with moral hazard, such as potentially taking on more risk or altering their healthcare-related decisions in response to the pandemic. This approach enables us to disentangle the effects of self-insurance from the general impact of COVID-19, providing insights into how different insurance structures may influence firm responses to major health shocks.

4.2 Instrumental Variables

While the DiD method will correctly estimate the differential reactions of self-insured and fully insured firms to COVID-19, interpreting this difference as moral hazard requires additional assumptions. Consider a scenario where risk-averse managers are more likely to fully insure their firms and simultaneously more inclined to limit work-from-home arrangements due to fears of negative financial outcomes. In the DiD setup, this could create a spurious correlation between self-insurance and changes in work-from-home rates that does not genuinely reflect moral hazard.

To address these potential endogeneity concerns, we employ an instrumental variables approach (via a control function). In a sense, we are trying to estimate "demand" for self-insurance. This is the inverse of buying full-insurance, so we can think of identification through the lens of estimating demand for full insurance. So for an instrument, we need supply-side factors that change the price of full insurance relative to self-insurance. We use two primary state-level instruments which allow us to use a our full data set. In Appendix whatever, we use the "insurer cost shock" instrument of Gao, Ge, Schmidt, and Tello-Tril (Gao et al.) on the subset of firms who begin our sample fully-insuring and isolate the effect on those who switch to self-insuring prior to COVID-19.

Our full sample, supply-shock instruments are...

In the first stage, we estimate the likelihood of a firm self-insuring based on state-level premium taxes:

$$\text{SelfInsure}_i = \alpha + \beta \text{Z}_i + \gamma_1 X_i + \epsilon_i \tag{2}$$

Where SelfInsure_i is a binary variable that indicates whether firm i chooses to self-insure or not. Z_i represents a collection of instruments, which includes variables like premium tax rates and potentially other relevant factors that influence the self-insurance decision but are not directly related to our outcome of interest. X_i is a vector of firm-level control variables that may also affect the self-insurance decision, such as firm size, industry, or other relevant characteristics. Finally, ϵ_i represents the error term, which captures any unobserved factors that might influence the self-insurance decision but are not included in our model.

Next, we calculate the residuals from this first-stage regression:

$$\hat{\epsilon}_i = \text{SelfInsure}_i - (\alpha + \beta \text{Z}_i + \gamma_1 X_i) \tag{3}$$

These residuals, $\hat{\epsilon}_i$, capture the endogenous variation in the self-insurance decision that is not explained by the instruments or other control variables (i.e., unoberved selection effects). In the second stage, we include these residuals directly in the DiD model as an additional control:

$$Y_{it} = \delta \text{SelfInsure}_i \times \text{PostCOVID}_t + \theta \text{SelfInsure}_i + \lambda \hat{\epsilon}_i + \gamma_2 X_i + \theta_t + \nu_{it} \tag{4}$$

Where Y_{it} represents our outcome of interest, such as the rate of work-from-home arrangements for firm i at time t. $PostCOVID_t$ is a binary indicator that takes the value 1 for the period after the onset of the COVID-19 pandemic and 0 otherwise. The term $\hat{\epsilon}_i$ represents the residual from the first stage regression, which we include to account for potential endogeneity in the self-insurance decision. We also incorporate θ_t to represent time fixed effects, allowing us to control for any time-specific factors that might affect all firms similarly. Lastly, ν_{it} is the error term for this second stage equation, capturing any remaining unexplained variation in the outcome.

In this setup, δ captures the differential effect of self-insurance on the outcome, net of the endogenous selection into self-insurance. The inclusion of $\hat{\epsilon}_i$ in the model corrects for any endogeneity in the self-insurance decision, ensuring that the estimated coefficient on the interaction term accurately reflects moral hazard.

4.3 Accounting for Non-Compliance

Figure 1: Distribution of Insured Person Count: Self-Insured vs. Fully-Insured

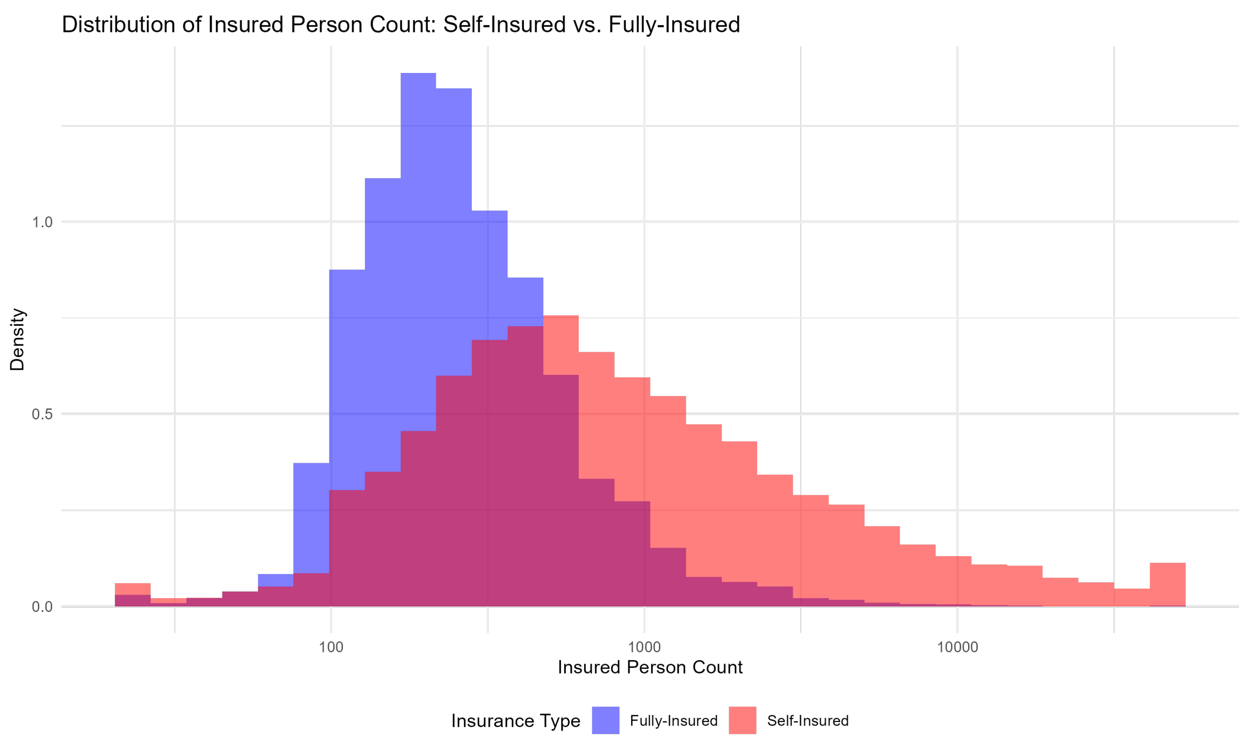

Figure 1 illustrates the distribution of insured person count for self-insured and fully-insured firms. While both distributions show a wide range of firm sizes, the self-insured category includes a substantial number of very large firms in the right tail of the distribution. This pattern presents a challenge for our instrumental variable approach.

The presence of these very large firms in the self-insured category suggests that beyond a certain size threshold, firms are highly likely to self-insure regardless of state-level premium tax rates. For these firms, the decision to self-insure is likely driven by factors such as economies of scale in risk management, rather than tax considerations. As a result, our instruments may have little to no effect on the self-insurance decision for this subset of very large firms. This lack of compliance with our instruments among the largest firms could lead to inefficient estimates and potentially weaken the validity of our instrumental variable strategy. It suggests a heterogeneity in the first-stage relationship between our instrument and the endogenous variable (self-insurance decision) that needs to be addressed in our empirical approach.

To address this issue, we apply the method proposed by Abadie et al. (2024). This approach allows us to improve the efficiency of our instrumental variable estimation when there is heterogeneity in the first-stage relationship. The method involves the following steps:

1. Divide the sample into groups based on firm size, as indicated by the insured person count.
2. Estimate group-specific first-stage relationships between the instruments and the endogenous variable.
3. Select groups with strong first-stage relationships using a data-driven procedure.
4. Interact the instruments with the indicators for the selected groups.

To implement this method, we modify our first-stage equation as follows:

$$\text{SelfInsure}_{ig} = \sum_{g=1}^{G} \beta_g \mathbf{Z}_i \cdot \mathbf{1}(i \in g) + \gamma X_i + \epsilon_{ig} \tag{5}$$

where g indexes the groups based on firm size. We then apply the Lasso selection procedure described in Belloni et al. (2012) to determine which interaction terms to include. This approach allows us to focus on the subpopulations where the instruments are most informative, potentially leading to more precise estimates of the moral hazard effects we aim to identify. Finally, after selecting the optimal combination of instruments, we use the UJIVE method of Kolesár (2013) to account for the bias induced by the large number of instruments.

By addressing the issue of low compliance in the right tail of the distribution, we can improve the validity and efficiency of our IV estimation, thereby strengthening our analysis of how insurance arrangements affect firm behavior during the COVID-19 pandemic.

5 Results

5.1 Main Findings

5.2 Robustness

6 Conclusions

References

Abadie, A., J. Gu, and S. Shen (2024). Instrumental variable estimation with first-stage heterogeneity. Journal of Econometrics 240(2), 105425.

Belloni, A., D. Chen, V. Chernozhukov, and C. Hansen (2012). Sparse models and methods for optimal instruments with an application to eminent domain. Econometrica 80(6), 2369–2429.

Gao, J., S. Ge, L. D. Schmidt, and C. Tello-Tril. How do health insurance costs affect low-and high-income workers?

Kolesár, M. (2013). Estimation in an instrumental variables model with treatment effect heterogeneity. Working Paper.

Appendix A Description of Variables

Online Appendix B   Appendix