Hey Shreya!

I think the presentations all went very fast for me so I might have missed some things you said that weren’t written down on the slides – I’d love to chat more in person! I think the PMUY subsidy is an interesting policy to evaluate and might be a good setting to investigate how effective the cash-back style subsidy can be with this population. I also like that it can help address a more broad question about these types of rebate subsidies and would love to see some reasoning on external validity. I have some concerns listed below, but there’s a good chance I might be misunderstanding the suggested model / estimated effects, so if it would be helpful, we could chat over coffee and discuss. I’ve numbered the feedback points in case it helps with later discussion, and put brief comments first then detailed comments after. I’m often overly verbose, so you probably only need to read the detailed comments if the brief ones don’t make sense.

Brief comments:

1. The first research question hints at comparing this cash-back subsidy to other subsidy designs. Is there a way you plan to make that comparison? This setting seems to only allow for comparison of this subsidy vs no subsidy.

2. When you ask “does demand increase or decrease with the subsidy?”, is this asking if the subsidy is *shifting* demand or just *moving along* the demand curve? Perhaps: “does quantity of gas refills demanded increase with the subsidy?”

3. Is there supposed to be an `h` subscript on the X in the regression equation?

3.a Is there a reason to use X = domestic price of gas refills? What does beta\_1 in this case tell us about the subsidy? Perhaps there’s an indicator term missing to create different slope or intercept coefficients for those under the subsidy?

3.b The identifying assumption seems hard to believe in the price vs demand setting. If this is particularly true here because of competition or something that’s unusual to this situation, there should be a clear description of this in the paper (there might have been in the presentation that I missed!)

3.c Related to 3.a – if you’re using X = subsidy, it seems important for identification of the desired effect to be comparing non-PMUY households to PMUY households, thus some interaction with a PMUY indicator would be necessary. Covariate overlap might be a concern that could be addressed with pruning or pre-analysis matching.

3.d Income effects – I’m interested to see how you include income in the regression equation. If you’re looking at differential subsidy effects across different income levels. I assume this means you’d need to have a beta\_2 \* X \* income \* Indicator(PMUY) term?

Detailed comments

**Regarding your research questions:**

*Q1: How does the cash-back design of gas subsidies affect take up?*

**1. Comparing to other subsidy designs**

Is this question about “compared to other subsidy designs?” Would you need to compare this to other designs to know this effect? If it is just about measuring the effectiveness of this particular subsidy, perhaps a different framing could make it more clear: “Does a cash-back natural gas subsidy increase take-up of natural gas as a cooking fuel?”

If you were able to compare this to other subsidies on the up-front price of the gas equipment, that comparative study sounds interesting.

*Q2: Does demand increase or decrease with subsidy?*

**2. Shifting demand or moving along it?**

Is this question about shifting demand or moving along the curve? If it’s moving along the curve, it seems hard to imagine a situation where offering a subsidy would shift demand down the curve, so the suggested question in point 1 above combines these two questions. I could imagine a situation where the subsidy appears to decrease demand for gas stoves: if the subsidy was only available for a particular brand of gas stove that was known by the community to be lesser quality, the demand for that stove under the subsidy could be less than the average demand for gas stoves. But I don’t think that would be a decrease in demand, just a measurement of demand for a subset of the available gas stoves.

If Q2 is about shifting the demand curve downward, I can’t think of a solid reason why a subsidy would shift a demand curve. Perhaps it could be large enough in scale to change the population’s information/beliefs and make them reconsider their preferences for stove types? I’m not sure how many people of the 911,563 are eligible for the subsidy.

**3. The regression equation**

I might have missed when you talked about the regression equation slide, so apologies if you discussed this! If I’m interpreting the equation correctly (that the X has no `h` subscript), then this regression using X=”the subsidy” would only be over the PMUY households (since X = 0 for the non-PMUY households). If X=”price of gas”, then this could include all the households.

**3.a Do the effects of gas price help determine the effects of the subsidy?**

If X=price of gas: Assuming the identifying assumption of cov(X, u) = 0 holds, I’m not sure why measuring the response of gas refills to the price would be helpful in determining the effect of the subsidy. This is just trying to estimate the local demand for gas refills, right? Would love to chat more about this to understand the reasoning.

**3.b The identifying assumption**

I can imagine that global gas prices can be a large driver of domestic gas prices here. However, it seems hard to believe that the domestic retailers of natural gas aren’t setting prices (at least a little) with respect to domestic demand. In the US, for example, utilities are a classic example of a natural monopoly because of the infrastructure required to maintain the utility (the pipes to delive the gas to households). However, I could imagine the assumption would be true in an extremely competitive retail gas industry. If the natural gas retail industry in India was more competitive (if there were many small businesses driving pressurize gas trucks around offering gas refills and competing with each other), then perhaps this is a reasonable assumption. This seems like a critical assumption to explain well in the paper if you use the subsidy directly as your main explanatory variable (since the subsidy is highly correlated with the price).

**3.c The effects of the subsidy on refills**

If X = “the subsidy on the price of a gas refill”: We can see from your graphs that the subsidy is highly correlated with the price of gas since the goal/outcome of the program was to keep the effective price relatively constant. So, if X does not have a subscript `h` on it, and the regression is only including the PMUY households, using the variation in subsidy to explain refills is essentially the same as using the unsubsidized price of gas since all the households are facing the same subsidy (and same up-front price).

Perhaps you were planning to use the non-PMUY households as well? If so, a PMUY indicator on a second subsidy or price term seems necessary. There is also the concern of covariate overlap in estimating the treatment effect: using income for example, the PMUY population is going to have a distribution of incomes on the lower tail of income, but the non-PMUY population could have either income over the full distribution, or specifically over the upper part of the income distribution and none of the lower part (if the subsidy is available for all low-income households).

To address this covariate overlap issue, I can suggest two different approaches based on the level of covariate overlap issues (there are probably others, but these are what I’m familiar with).

- if there is some overlap in the lower part of the income distribution, then simply pruning the non-PMUY HHs to only include HHs in the part of the income distribution shared with the PMUY HHs. This could just be removing any HHs with income above the PMUY threshold, or matching the non-PMUY HHs to the PMUY HHs to attempt good covariate balance.

- if there is little or no overlap in income between the non-PMUY HHs and PMUY HHs, this could be a good reason to attempt regression discontinuity at the PMUY income threshold. This would remove your ability to look at differential effects by income, but if there was no overlap in income anyway, I don’t think this would have been viable anyway.