

Project: The incidence of tax evasion: experimental evidence from Italy.

Authors: Albrecht Bohne, Giacomo Brusco, Leonardo M. Giuffrida¹

A Short Note Regarding Power Calculations

All our calculations are based on the Stata command `power twomeans`, with the following options specified:

- Sample size, `n` = 1000;
- Statistical power, `power` = 0.8;
- Statistical significance, `alpha` = 0.05;
- Ratio of treatment to control observations, `nratio` = 1;
- Baseline mean, `m1` = 31.8879.

Our baseline mean is based on the observed average price of the meal, excluding beverages, we found in our pilot survey.

The computed standard deviation from our pilot results is 15.45. When we use this standard deviation, `sd` = 15.45, we obtain a Minimum Detectable Effect (MDE) of 2.74, which corresponds to about 8.5% of the baseline average price.

Given we have several observable characteristics of restaurants and our main results will be able to condition on these characteristics, the most relevant standard deviation is the one conditional on these observables. We therefore estimate the following specification:

$$P_i = \beta_1 + \sum_{u=1}^{U-1} \beta_{2u} \mathbf{1}\{\text{Urban Area} = u\}_i + \sum_{f=1}^{F-1} \beta_{3f} \mathbf{1}\{\text{Firm Type} = f\}_i + \beta_4 \text{TenPpl}_i \quad (1)$$
$$+ \sum_{c=1}^{C-1} \beta_{6c} \mathbf{1}\{\text{Revenue Class} = c\}_i + \sum_{k=1}^{K-1} \beta_{7k} \mathbf{1}\{\text{Profits Class} = k\}_i + \epsilon_i,$$

where P_i is the price elicited from restaurant i , $\mathbf{1}\{\text{Urban Area} = u\}_i$ is an indicator for being part of urban area u of U total urban areas in our sample, $\mathbf{1}\{\text{Firm Type} = f\}_i$ is an indicator for being legally organized as firm type f (e.g. “sole proprietorship”, “family firm”, “partnership”, etc.) of F firm types in our sample, TenPpl_i is a dummy for whether the enumerator who visited restaurant i asked for a dinner for 10 people (as opposed to 14), $\mathbf{1}\{\text{Revenue Class} = c\}_i$ is a dummy for being part of revenue class c out of four total classes in our sample, $\mathbf{1}\{\text{Profits Class} = k\}_i$ is a dummy for being part of profit class k , out of six total classes in our sample, and ϵ_i is an error term. Note revenue and profits have been coded up as classes to preserve sample anonymity.

The standard deviation of the residuals resulting from estimating Equation 1 on the sample from our pilot survey is 13.48. When we perform our power calculations with this standard deviation, `sd` =

¹Bohne: ZEW Mannheim; Brusco: School of Business and Economics, University of Tübingen; Giuffrida: ZEW Mannheim, MaCCI, and CESifo.

13.48, we obtain an MDE of 2.39, which corresponds to about 7.5% of our baseline average price. This is the number we report in our AEA pre-registration.

Whether or not we control for restaurant observables, our MDE calculations imply we should be able to find full evasion pass-through if it occurred, given a VAT rate of 10% of restaurants in Italy. If we consider instead evasion not just on VAT but on income taxes as well, the point is made even stronger. Hopefully, however, our main survey will be able to detect a lot more. In our pilot survey, we did not re-sample restaurants, while we plan to do this in our main survey. Such re-sampling will allow us to run regressions using restaurant fixed effects, or using the within-restaurant price difference as the dependent variable. In both cases, we expect the remaining standard deviation to be considerably smaller than the one we obtained in the cross-sectional regression in the pilot study. Such a smaller standard deviation would increase our statistical power and thereby shrink the MDE.

In order to estimate a valid standard deviation on which to base our power calculations, we plan to retrospectively estimate the standard deviation on the final dataset. Since this regression is based on data we have not collected yet at the point of writing and submitting this document, we hereby pre-specify the exact regression in order to increase the credibility of the results:

$$P_{rv} = \alpha_r + \varepsilon_{rv}. \tag{2}$$

In Specification 2, P_{rv} is the price quote obtained from the v th visit of restaurant r , α_r is a fixed effect for restaurant r , and ε_{rv} is an error term. The relevant standard deviation for us is the one resulting from the estimation of Equation 2. We can use this standard deviation together with the values specified at the beginning of this document to retrospectively estimate the power of our analysis in a credible manner.