Is Tipping Incentive Relevant?

Jacob Kohlhepp*

September 8, 2025

Abstract

Tips represent an important part of compensation for many American workers. Understanding whether tips incentivize service quality is critical for deciding whether tips should be treated differently than wages under tax and minimum-wage laws. For tipping to be incentive relevant, average tips must be positively correlated with service quality. I develop a test for incentive relevant tipping that consists of a simple regression of first-time tip percentages on an indicator for whether a customer returns in the future. A positive coefficient reveals either a behavioral quality-based tipping norm, rational forward-looking relational incentives or both. I implement the test using more than 200,000 first-time tips from the beauty industry, a setting where the test is high-powered because customers return to the same worker 30.7% of the time. I find no evidence that tipping is incentive relevant, even when the test is conducted among customers visiting the same worker. Across 2,953 workers, I fail to reject incentive-irrelevant tipping in 98.5% of cases. I instead find evidence that social norms play a role. As an example, 31.2% of first-time tips are exactly 20%, the likely default of the default built into the management software. Among customers with an observed second tip who tipped 20% during their first visit, 63.0% also tipped exactly 20% during their second visit.

Keywords: tipping, social norms

JEL Codes: J32, D91

*Department of Economics, UNC Chapel Hill. Contact: kohlhepp@unc.edu. I thank Daniel Ober-Reynolds, Daniel Haanwinckel, Ofer Azar, and Michael Lynn for their excellent feedback.

1 Introduction

Tips represent a large share of compensation for many service workers in the United States (Azar 2020). A crucial question, the focus of this paper, is whether tipping is incentive relevant. Do tips serve as a form of implicit performance pay, encouraging high-quality service, or are tips simply norm-based transfers unrelated to quality? The answer is critical for ongoing debates on how tips should be treated in tax and minimum-wage policy. For example, if tips are incentive irrelevant, making them tax exempt creates an easy-to-exploit loophole where two essentially equivalent types of earnings (wages and tips) are treated differently with little economic benefit.[1][2] If tips are incentive relevant, making them tax exempt effectively strengthens performance pay, potentially improving productivity via enhanced service quality. In this case, the productivity benefits could be worth the distortions caused by differential treatment.

Is tipping incentive relevant? Three challenges, largely driven by lack of rich data on tipping, make it difficult to give a general and definitive answer using past work. First, most past work does not account for the fact that incentive-relevant tipping can arise either due to strategic concerns about future interaction or a behavioral quality-based social norm, and these two channels are often observationally equivalent with limited data. Second, most past work focuses on specific settings, like restaurants or ride sharing, where tipping is ex ante less likely to be incentive relevant. Third, most past work that studies the quality-tipping connection is not well-powered to detect an effect. Low power arises because of small sample sizes, ordinal comparison of self-reported measures of service quality, and a low probability of repeat interaction in many contexts like ride sharing.

This paper overcomes these three challenges to show that tipping is incentive irrelevant. I develop a test that allows for both quality-based social norms and strategic concerns about future interaction. I apply the test to the beauty industry, where tipping is ex ante more likely to be incentive relevant because repeat interaction is common and interactions involve extended physical contact. The test avoids using self-reported service quality measures, instead relying on revealed preference via the decision to return. I use a large administrative data set from management software which allows me to focus only on first-time tips while maintaining a sample size of over 200,000 tips. I show there is little evidence of a quality-tip connection even if one compares only customers who see the same worker. That tipping is incentive irrelevant in the beauty industry suggests it is unlikely to be incentive relevant in other settings.

The test for incentive-relevance leverages the observation that both quality-based social norms and strate-

[1] A tax exemption on tips was proposed by both major US presidential candidates and eventually put in place by the Big Beautiful Bill (Internal Revenue Service 2025). In response to the proposals of the presidential candidate, the Brookings published an article pointing out the distortion from differential tax treatment (Berlin and Gale 2025).

[2] This flavor of argument has been made to support equating the federal tipped minimum wage and the regular federal minimum wage (Neumark and Wohl 2024)

gic concerns about future interaction manifest in the data as a positive correlation between first-time tips and whether a customer returns in the future. I derive a necessary and sufficient condition for incentive relevant tipping, and show it can be be tested via a simple regression of first-time tip percentages on an indicator for whether a customer returns. I show the test gains power (can better detect incentive-relevant tipping) as the customer return rate rises from 0 to 0.5. The return rate in my data is 0.307, so the test is relatively well-powered to detect incentive-relevant tipping.

I apply the test to over 200,000 first-time tips recorded by a salon-management software. I conduct the test at different levels of aggregation, from full aggregation to disaggregated by individual salon worker. I fail to reject the null hypothesis of incentive-irrelevant tipping in most cases at all levels of aggregation. My preferred level of aggregation is at the worker level. I fail to reject incentive irrelevance in 98.5% of the 2,935 workers. I perform a number of robustness analyses and similarly fail to reject incentive irrelevance in most cases.

Given that tips do not appear to be explained by service quality, I then ask whether some important patterns might suggest what drives observed variation in tips. I first show suggestive evidence that tipping prompts and defaults play a role. Although firms can customize their tipping prompts, a blog post by the company suggests the default tipping prompt built into the management software is 18%, 20%, and 22%. In the data, 31.1% of first-time tips are exactly 20%, the default of the default since it is the middle option for customers. Further, among customers with an observed second tip who tipped the default of the default during their first visit, 63.0% also tipped exactly the default of the default during their second visit. I also show that the average tip percentage of other customers in a zip code explains a similar amount of variation as the average tip left by other customers at an establishment, suggesting geographic norms may play a larger role than firm-specific tipping norms.

Finally, I show that the context of the visit is associated with the tip percentage. Specifically, salons can allow customers to prebook appointments online using the management software. If customers prebook, salons can choose to prompt customers to tip prior to the visit with the opportunity to revise at the end of the visit. As a result, some prebooked customers are prompted to tip twice, with an endogenous default the second time determined by the customer's preliminary tip from prebooking. Among customers that do not request a worker, prebooked appointments have 1.63 percentage point higher tips on average. Among customers that specifically request a worker, prebooked appointments have 0.79 percentage point lower tips on average. These associations are consistent with double prompting with an endogenous default increasing tips when a customer allows the salon to match them with a worker, and decreasing tips when a customer chooses the worker they want to see.

This paper contributes to several strands of research. One strand asks why customers tip. The approach in much of this work has been to directly ask customers (Lynn 2018, Azar 2010b), except for Azar 2007, which compares the sensitivity of tips to self-reported quality across repeat and non-repeat customers. An important finding in this strand has been that concerns about future interaction do not seem to determine tipping behavior (Azar 2010a,Azar 2020). I argue that the force the literature describes as future interaction is best thought of as the implicit use of tipping as part of a relational contract. The test developed in this paper for incentive relevance can detect such relational tipping, and I find no evidence for this force, which confirms the findings in the literature.

Another strand asks whether tips are related to service quality. This question is distinct from the last, because forces like a social norm can generate quality-based tips even if customers are not concerned about future interaction. Four main approaches have been used to test for a quality-tip link: randomization of service quality (Malcman et al. 2025), directly asking customers (workers) if they tip (are tipped) based on quality (Azar 2010b, Kwortnik Jr, Lynn, and Ross Jr 2009), measuring self-reported quality and regressing this on tips (Lynn and McCall 2000, Conlin, Lynn, and O'Donoghue 2003, Azar 2007), and analyzing the association between tipping policies and service ratings across firms or within firms across time (Kwortnik Jr, Lynn, and Ross Jr 2009, Lynn and Brewster 2018, Lynn and Kwortnik 2015). Results have been mixed, with workers and customers reporting that tips are based on service quality and a negative association between policies which stop tipping and service quality, while evidence from regressions and randomization of service quality show a weak connection between tips and service quality. The puzzling nature of these findings is well summarized by Azar 2009.

I contribute by studying a setting where ex ante tipping is most likely to be associated with service quality, using data that allow me to connect many tips from distinct customers to individual workers. I implement a test for quality-based tipping that is both powerful due to the high frequency of repeat interaction in the beauty industry and avoids needing to implicitly aggregate self-reported service quality measures. I find no evidence that tips are related to quality, which supports the weak or absent link found from regressing tips on self-reported quality, but goes against past work using direct surveys of workers and customers. A key limitation of approaches which use survey data to study the tip-quality relationship is that one must implicitly make cardinal comparisons of self-reported quality across people. Using the return decision allows me to avoid this limitation using revealed preference. The main assumption is that of a single index for quality, which determines the return decision via a single threshold and potentially the tip percentage.

This paper also contributes to the literature by addressing the issue of statistical power directly. Much of the work studying the connection between service quality and tips uses relatively small sample sizes. When the quality-tip relationship is assessed using return behavior, the minimal detectable effect will generally depend on the average probability a customer returns to the same firm/establishment/worker. In many settings studied in the literature, like ride sharing, hotels and restaurants, this return rate is low. In the beauty industry, it is high, which makes detecting an effect with the same sample size more likely.

Several papers have shown in different ways that tipping appears to be connected to social norms that differ across people and places. Chandar et al. 2019 show that variation among Uber riders accounts for most tip variation, while Haggag and Paci 2014 and Chandar et al. 2024 show how tips react to a change in default. McCall and Belmont 1996 show that the tip amount is impacted by the presence of a major credit-card company emblem on the payment tray. This paper contributes by showing that this view of tipping generalizes to the beauty industry, a place where ex ante it might be the least likely to hold due to the extended personal contact involved in salon work. This paper also contributes by showing that the social norms that undergird tipping appear to be on average not quality-based, which suggests that if workers have rational expectations and can improve service quality at a cost, they would not choose to do so.

2 A Statistical Framework for Tipping

This section presents a statistical framework to study whether tips are incentive relevant. This framework captures two potential drivers of incentive-relevant tipping: quality-based norms and relational contracts, which are described at length in Section 4.

A customer experiences service quality q, which I assume has weakly positive support. Based on service quality, the customer makes a return decision, $r(q)$, and then leaves a tip $b(q)$. Tips can be expressed as $b(q) = r(q)\tilde{b}_1(q) + [1 - r(q)]\tilde{b}_0(q) + \epsilon$. Under this specification, a customer may have a potentially quality-based tipping strategy that differs depending on whether they plan to return. Idiosyncratic variation in the baseline level of the tip is captured by ϵ, which I assume is mean 0 and satisfies $E[q\epsilon] = 0$. I specify a linear tipping strategy both when the customer plans to return, $\tilde{b}_1(q) = \beta_0 + \beta_1 q$, and when the customer plans not to return, $\tilde{b}_0(q) = \alpha_0 + \alpha_1 q$. The return decision is given by $r(q) = \mathbb{I}\{q + \omega \geq 0\}$. I assume that (q, ω, ϵ) are mutually independent.

Throughout the paper, I maintain two assumptions on the parameters. First, tips are non-decreasing in service quality ($\beta_1 \geq 0$ and $\alpha_1 \geq 0$). Second, if service quality and customer identity are held fixed, tips cannot decrease when a customer plans to return ($\beta_0 \geq \alpha_0$ and $\beta_1 \geq \alpha_1$). I specify that service quality (q) can be additively decomposed into a potentially stochastic component ψ that a worker cannot influence and a component e that the worker can influence. If a worker can directly improve service quality by exerting costly effort, this would be a part of e. A worker will exert costly effort to improve service quality if they expect it will increase tips. This motivates the following definition for incentive-relevant tipping.

Definition 1 *Tipping is incentive relevant if* $\frac{\partial \mathbb{E}[b(q)|e=x]}{\partial x} > 0$*, and incentive irrelevant if* $\frac{\partial \mathbb{E}[b(q)|e=x]}{\partial x} = 0$.

The statistical framework laid out earlier provides a precise necessary and sufficient condition under which tipping is incentive irrelevant.

Proposition 1 *Tipping is incentive irrelevant if and only if* $\alpha_1 = \beta_1 = 0$ *and* $\alpha_0 = \beta_0$.

Proof. I prove the "if' portion first. Expanding the expression for tips, I obtain

$$\begin{aligned}
b(q) &= (\beta_0 + \beta_1 q)r + (1-r)(\alpha_0 + \alpha_1 q) + \epsilon \\
&= \alpha_0 + \alpha_1 q + \underbrace{[\beta_0 + (\beta_1 - \alpha_1)q - \alpha_0]}_{\Delta(q)} r + \epsilon \\
&= \alpha_0 + \mathbb{E}[\Delta(q)]r + \alpha_1 q + (\Delta(q) - \mathbb{E}[\Delta(q)])r + \epsilon.
\end{aligned}$$

If $\alpha_1 = \beta_1 = 0$ and $\alpha_0 = \beta_0$, $\Delta(x) = 0$ for all x. Thus we have that $b(q) = \epsilon$ and $\mathbb{E}[\epsilon] = 0$; therefore $\frac{\partial \mathbb{E}[b(q)]}{\partial e} = 0$. I now prove the "only if" portion. Suppose $\frac{\partial \mathbb{E}[b(q)]}{\partial e} \leq 0$. I take the conditional expectation:

$$\begin{aligned}
\mathbb{E}[b(q)|q = x] &= \mathbb{E}[\alpha_0 + \mathbb{E}[\Delta(q)]r + \alpha_1 q + (\Delta(q) - \mathbb{E}[\Delta(q)])r + \epsilon | q = x] \\
&= \alpha_0 + \mathbb{E}[\Delta(q)]F_\omega(x) + \alpha_1 x + \Delta(x)F_\omega(x) - \mathbb{E}[\Delta(q)]F_\omega(x) \\
&= \alpha_0 + \alpha_1 x + \Delta(x)F_\omega(x).
\end{aligned}$$

Taking derivatives:

$$\frac{\partial \mathbb{E}[b(q)|q = x]}{\partial x} = \alpha_1 + \Delta(x)f_\omega(x) + \Delta'(x)F_\omega(x)$$

By the dominated convergence theorem and law of iterated expectations, I have that

$$\begin{aligned}
\frac{\partial \mathbb{E}[b(q)]}{\partial e} &= \int \frac{\partial \mathbb{E}[b(q)|q = s + e]}{\partial e} dF_\psi(s) \\
&= \int \left.\frac{\partial \mathbb{E}[b(q)|q = x]}{\partial x}\right|_{x=s+e} dF_\psi(s) \\
&= \int \alpha_1 + \Delta(s+e)f_\omega(s+e) + \Delta'(s+e)F_\omega(s+e)dF_\psi(s) \\
&= \alpha_1 + \int \Delta(s+e)f_\omega(s+e) + \Delta'(s+e)F_\omega(s+e)dF_\psi(s).
\end{aligned}$$

Under the maintained assumptions, the integral expression and α_1 are both weakly positive. I also have that $\frac{\partial \mathbb{E}[b(q)]}{\partial e} = 0$. These imply $\alpha_1 = 0$ and $\int \Delta(s+e)f_\omega(s+e) + \Delta'(s+e)F_\omega(s+e)dF_\psi(s) = 0$. If $\alpha_1 = 0$, then because $\alpha_1 \geq \beta_1 \geq 0$, I have that $\beta_1 = 0$. This implies $\Delta'(s+e) = \beta_1 - \alpha_1 = 0$. This further implies $\int (\beta_0 - \alpha_0)f_\omega(s+e)dF_\psi(s) = 0$. Because $\beta_0 \geq \alpha_0$, it must be that $\beta_0 = \alpha_0$.■

3 A Regression-Based Test for Incentive Relevance

This section derives a simple test for incentive-relevant tipping that can be performed by running a regression of first-time tip percentages on an indicator for whether a customer returns. To understand the spirit of the test, recall that under the statistical framework, tips can be decomposed into a causal effect of returning and omitted variable bias:

$$b(q) = \alpha_0 + \underbrace{\mathbb{E}[\Delta(q)]r(q)}_{\text{causal effect of returning}} + \underbrace{\alpha_1 q + (\Delta(q) - \mathbb{E}[\Delta(q)])r}_{\text{omitted variable bias}} + \epsilon$$

The test leverages the fact that both forms of incentive-relevance manifest as a positive regression coefficient on the return indicator.

Proposition 2 *Given data on first-time tips and customer return behavior, tipping is incentive irrelevant if and only if*

$$\mathbb{E}\begin{bmatrix} 1 & r(q) \\ r(q) & r(q)^2 \end{bmatrix}^{-1} \mathbb{E}\begin{bmatrix} b(q) \\ r(q) \cdot b(q) \end{bmatrix} = \begin{pmatrix} \beta_0 \\ 0 \end{pmatrix}$$

Proof. It will be useful throughout the proof to expand the following matrix expression:

$$\mathbb{E}\begin{bmatrix} 1 & r(q) \\ r(q) & r(q)^2 \end{bmatrix}^{-1} = \frac{1}{\mathbb{E}[r(q)^2] - \mathbb{E}[r(q)]^2} \begin{bmatrix} \mathbb{E}[r(q)^2] & -\mathbb{E}[r(q)] \\ -\mathbb{E}[r(q)] & 1 \end{bmatrix}$$

I begin by showing the only if direction. I have that $b(q) = \alpha_0 + \alpha_1 q + \Delta(q) r + \epsilon$. If tipping is incentive irrelevant, $\alpha_1 = \beta_1 = 0$ and $\alpha_0 = \beta_0$. Then I have that $\Delta(q) = 0, \alpha_1 q = 0$ for all q; therefore $b(q) = \beta_0 + \epsilon$. Thus tips are independent of quality and the return decision, and by the law of iterated expectations $\mathbb{E}[b(q) \cdot r(q)] = \mathbb{E}[r(q)\mathbb{E}[b(q)|r(q)]] = \beta_0 \mathbb{E}[r(q)]$. Also $E[b(q)] = \beta_0$. Thus,

$$\mathbb{E}\begin{bmatrix} 1 & r(q) \\ r(q) & r(q)^2 \end{bmatrix}^{-1} \mathbb{E}\begin{bmatrix} b(q) \\ r(q) \cdot b(q) \end{bmatrix} = \frac{1}{\mathbb{E}[r(q)^2] - \mathbb{E}[r(q)]^2} \begin{bmatrix} \mathbb{E}[r(q)^2] & -\mathbb{E}[r(q)] \\ -\mathbb{E}[r(q)] & 1 \end{bmatrix} \begin{pmatrix} \beta_0 \\ \beta_0 \mathbb{E}[r(q)] \end{pmatrix} = \begin{pmatrix} \beta_0 \\ 0 \end{pmatrix},$$

the desired result.

I now show the if direction; that is, I assume the condition holds and show this implies tips are incentive irrelevant. I prove this by contradiction; that is I assume the condition holds and tips are incentive relevant. I can rewrite the matrix involving tips from the condition, separating it into two parts:

$$\mathbb{E}\begin{bmatrix} b(q) \\ r(q)\cdot b(q) \end{bmatrix} = \mathbb{E}\begin{bmatrix} \alpha_0 + \mathbb{E}[\Delta(q)]r(q) + \epsilon \\ r(q)\cdot(\alpha_0 + \mathbb{E}[\Delta(q)]r(q) + \epsilon) \end{bmatrix} + \mathbb{E}\begin{bmatrix} \alpha_1 q + (\Delta(q) - \mathbb{E}[\Delta(q)])r \\ r(q)\cdot(\alpha_1 q + (\Delta(q) - \mathbb{E}[\Delta(q)])r) \end{bmatrix}$$

Focusing on the first matrix, note that

$$\frac{1}{\mathbb{E}[r(q)^2] - \mathbb{E}[r(q)]^2}\begin{bmatrix} \mathbb{E}[r(q)^2] & -\mathbb{E}[r(q)] \\ -\mathbb{E}[r(q)] & 1 \end{bmatrix}\mathbb{E}\begin{bmatrix} \alpha_0 + \mathbb{E}[\Delta(q)]r(q) + \epsilon \\ r(q)\cdot(\alpha_0 + \mathbb{E}[\Delta(q)]r(q) + \epsilon) \end{bmatrix} = \begin{pmatrix} \alpha_0 \\ \mathbb{E}[\Delta(q)] \end{pmatrix}$$

because $r(q)$ is an indicator. Thus the left-hand-side of the condition from the proposition becomes:

$$\begin{pmatrix} \alpha_0 \\ \mathbb{E}[\Delta(q)] \end{pmatrix} + \frac{1}{\mathbb{E}[r(q)^2] - \mathbb{E}[r(q)]^2}\begin{bmatrix} \mathbb{E}[r(q)^2] & -\mathbb{E}[r(q)] \\ -\mathbb{E}[r(q)] & 1 \end{bmatrix}\mathbb{E}\begin{bmatrix} \alpha_1 q + (\Delta(q) - \mathbb{E}[\Delta(q)])r \\ r(q)\cdot(\alpha_1 q + (\Delta(q) - \mathbb{E}[\Delta(q)])r) \end{bmatrix}$$

That is the association between tips and the return indicator can be decomposed into the causal effect of returning on tips plus omitted variable bias generated by unobserved service quality. Noting that $\mathbb{E}[r(q)^2] - \mathbb{E}[r(q)]^2$ is the variance of the return indicator, and again using the fact that $r(q) = r(q)^2$ because r is an indicator, I expand the expression for omitted variable bias

$$\frac{1}{Var(r(q))}\begin{bmatrix} \mathbb{E}[r(q)] & -\mathbb{E}[r(q)] \\ -\mathbb{E}[r(q)] & 1 \end{bmatrix}\mathbb{E}\begin{bmatrix} \alpha_1 q + (\Delta(q) - \mathbb{E}[\Delta(q)])r \\ r(q)\cdot(\alpha_1 q + (\Delta(q) - \mathbb{E}[\Delta(q)])r) \end{bmatrix} = \frac{1}{Var(r(q))}\begin{bmatrix} \alpha_1\mathbb{E}[r(q)](\mathbb{E}[q] - \mathbb{E}[r(q)q]) \\ F(q) \end{bmatrix}$$

Focusing on $F(q)$,

$$\begin{aligned} F(q) &= -\alpha_1\mathbb{E}[r(q)]\mathbb{E}[q] - \mathbb{E}[r(q)]\mathbb{E}[r(q)\Delta(q)] + \mathbb{E}[\Delta(q)]\mathbb{E}[r(q)]^2 + \alpha_1\mathbb{E}[r(q)q] + \mathbb{E}[\Delta(q)r(q)] - \mathbb{E}[\Delta(q)]\mathbb{E}[r(q)] \\ &= \alpha_1 Cov(r(q), q) + (1 - \mathbb{E}[r(q)])Cov(r(q), \Delta(q)) \end{aligned}$$

Combining all pieces, the left-hand side of the condition from the proposition is equal to

$$\begin{pmatrix} \alpha_0 \\ \mathbb{E}[\Delta(q)] \end{pmatrix} + \frac{1}{Var(r(q))}\begin{bmatrix} \alpha_1\mathbb{E}[r(q)](\mathbb{E}[q] - \mathbb{E}[r(q)q]) \\ \alpha_1 Cov(r(q), q) + (1 - \mathbb{E}[r(q)])Cov(r(q), \Delta(q)) \end{bmatrix}$$

The bottom scalar is strictly positive when tips are incentive relevant because of four facts. First, $E[\Delta(q)] = \beta_0 - \alpha_0 + (\beta_1 - \alpha_1)\mathbb{E}[q] > 0$ because q has positive support. Second, $Var(r(q)) \geq 0$ and $\alpha_1 \geq 0$. Third, $1 - \mathbb{E}[r(q)] \geq 0$ because it is the fraction of customers that do not return. Finally, the $Cov(r(q), q) \geq 0$ and $Cov(r(q), \Delta(q)) \geq 0$. To see this final fact, note that $r(q)|\omega = x = \mathbb{I}\{q + x \geq 0\}$ is an increasing function of

q, q is trivially an increasing function of q, and $\Delta(q) = \beta_0 - \alpha_0 + (\beta_1 - \alpha_1)q$ is an increasing function of q (because $\beta_1 - \alpha_1 \geq 0$). It can be shown that the covariance of two increasing functions applied to the same random variable is weakly positive (Schmidt 2003). Thus, I have that

$$\begin{aligned}
Cov(r(q), \Delta(q)) &= \mathbb{E}[Cov(\mathbb{I}\{q + \omega \geq 0\}, \Delta(q)|\omega)] + Cov(\mathbb{E}[\mathbb{I}\{q + \omega \geq 0\}|\omega], \mathbb{E}[\Delta(q)|\omega]) \\
&= \mathbb{E}[Cov(\mathbb{I}\{q + \omega \geq 0\}, \Delta(q)|\omega)] \\
&= \int Cov(\mathbb{I}\{q + \omega \geq 0\}, \Delta(q)|\omega = x) dF_\omega(x) \\
&\geq 0
\end{aligned}$$

where the first line follows from the law of total covariance and the second from the independence of ω and q. The same argument implies $Cov(r(q), q) \geq 0$. Thus the bottom scalar is strictly positive, which contradicts the assumption that it is 0 which was made at the beginning of the proof.■

An ordinary least squares regression of tip percentages on a constant and the return indicator converges in probability to the left-hand side of the condition. The coefficient should be strictly positive if tipping is incentive relevant, and it should be 0 if not. Therefore, the test can be implemented by running a regression of the form $b(q) = c_0 + c_1 r(q)$ and performing a Wald Test with the null hypothesis $c_1 = 0$ and alternative hypothesis $c_1 > 0$. Such a test is consistent because under the null hypothesis there is no omitted variable bias and the OLS estimator for the coefficient on the return indicator converges in probability to 0.

The test is powerful in settings where the fraction of customers who return is close to 0.5. A simple way to see this is to consider only alternative hypotheses where $\beta_1 = \alpha_1 = 0, \beta_a > \alpha_0$. Assume the variances of tips are equal for returning and non-returning customers. The tipping expression reduces to $b = \alpha_0 + (\beta_0 - \alpha_0)r(q) + \epsilon$. Omitted variable bias does not exist under this set of alternatives, and $r(q)$ is a binary indicator; therefore a regression of the form $b(q) = c_0 + c_1 r(q)$ is analogous to a standard treatment effect regression, where a customer is treated if they decide to return. Thus, I can use a standard derivation like that given by McConnell and Vera-Hernández 2015, to obtain the following expression for the minimal detectable effect:

$$MDE(\beta_0 - \alpha_0) = (z_{POWER} + z_{0.05})\sqrt{\frac{Var(\epsilon)}{N\mathbb{E}[r(q)](1 - \mathbb{E}[r(q)])}}$$

where the unconditional expectation of $r(q)$, $\mathbb{E}[r(q)]$, is the fraction of customers that return and N is the total sample size. The minimal detectable effect is lowest when exactly half of customers return, and it is decreasing between $\mathbb{E}[r(q)] = 0$ and $\mathbb{E}[r(q)] = 0.5$. Thus, the test requires fewer observations to detect the same effect size as the return rate rises from 0 towards 0.5. In many settings, the unconditional probability a customer returns to the same service provider (not the same firm) is low: Chandar et al. 2019 find only 1%

of rider-driver pairs on Uber match again in a one-month period. Even in industries like restaurants, where repeat customers are common, worker turnover and random matching may make the return rate to the same waiter/waitress low. The test has low power in such settings. In the beauty industry, repeat customers are common: in my data, 30.7% of customers are observed matching with the same worker again. Therefore the test is especially well suited for the beauty industry.

4 Mechanisms Behind Incentive Relevance

Past work on tipping in both economics and psychology provides several reasons why tips might be correlated with service quality. In this section, I classify these mechanisms into two broad categories: quality-based norms and relational contracts. I then show how each mechanism impacts tips under the statistical model, and how each mechanism is represented in the necessary and sufficient condition for tipping to be incentive irrelevant.

Quality-Based Norms. There is a common belief in the United States that tip percentages should increase with the quality of service. I refer to this belief as a quality-based norm because tipping based on quality is not rational in one-shot interactions. Even if quality-based norms are behavioral rather than rational on the part of the client, they can encourage rational costly effort on the part of the worker.

If quality-based norms exist, the slope of tips with respect to service quality should be positive whether or not the customer plans to return; that is, $\beta_1 > 0$ and $\alpha_1 > 0$. This is precisely why $\beta_1 = \alpha_1 = 0$ is necessary for tipping to be incentive irrelevant. If quality-based norms exist and service quality is not observed, the proof of Proposition 2 shows that a regression of tip percentages on a return indicator will be plagued by omitted variable bias. The coefficient will be biased upward, because higher unobserved service quality causes both higher return rates and higher tip percentages. The upward bias will be correctly interpreted by the statistical test as evidence that tips are incentive relevant.

Relational Contracts. A key feature of the beauty industry is that customers and workers often interact repeatedly, sometimes over long time horizons. In such a setting, tips may play a role that they cannot play in other settings: they may stand in as an implicit bonus in a relational contract between the customer and the stylist. The literature on tipping often refers to this mechanism as "strategic forward-looking behavior." Because this force arises from two rational agents interacting dynamically, it is helpful to recast this mechanism as a relational contract in the spirit of Levin 2003.

To make the connection a bit more explicit, suppose each period, a customer proposes a base payment for the service as well as a contingent tip that cannot be enforced. The worker then decides whether to accept or reject the contract. If the worker rejects, both receive their outside options. If the worker accepts, the worker chooses effort which yields stochastic service quality. Privately observing service quality but not effort, the customer decides whether to follow through on their promised tip. While the base payment is contractually enforced, the tip is not. The stage game then repeats for an infinite number of periods. Such a model is equivalent to the subjective performance valuation model proposed by Levin 2003, and as a result there is a class of equilibria where the client pays a base wage and then a two-level contingent tip that is high when service quality is above some threshold and low when service quality is below some threshold.[3] Importantly, the relationship also terminates (the worker rejects the contract offered forever after) whenever output is low and the low tip is paid. In this way, tipping can function as a relational contract that sustains better service quality even if quality-based tipping norms do not exist.

If tips serve as a relational contract à la Levin 2003, this will manifest in the reduced form statistical model. The baseline level of tips left by the customer when they return should be higher than the baseline level when they do not: formally $\beta_0 > \alpha_0$. This is precisely why $\beta_0 = \alpha_0$ is necessary for tipping to be incentive irrelevant. If tips serve as a relational contract, the "contingent bonus" $\beta_0 - \alpha_0$ will directly generate a positive correlation between the return indicator and tip percentage. Even more, if quality-based norms do not exist, formally $\alpha_1 = \beta_1 = 0$, a regression of tips on a return indicator will consistently estimate the relational contingent bonus $\beta_0 - \alpha_0$. The positive sign of the coefficient on the return indicator will be correctly interpreted by the test as evidence that tips are incentive relevant.

An important caveat is that, although the test can detect either form of incentive relevant tipping, it cannot distinguish between them. This is fundamentally a limitation of using a revealed preference approach that assumes quality is unobserved. When quality is not observed and tips are incentive relevant, better service can lead to higher tips either because of a quality-based social norm or because the customer knows they are more likely to come back. When the researcher has access only to first-time tips and a return indicator, these two forces are observationally equivalent. Telling them apart requires either more structure or more data. It is possible that service quality is not a large determinant of the return decision. Formally, the variance of service quality is small relative to the idiosyncratic factors in the return decision (ω). If this is true, the test is effectively only a test of relational/forward looking behavior.

5 Data and Institutional Details

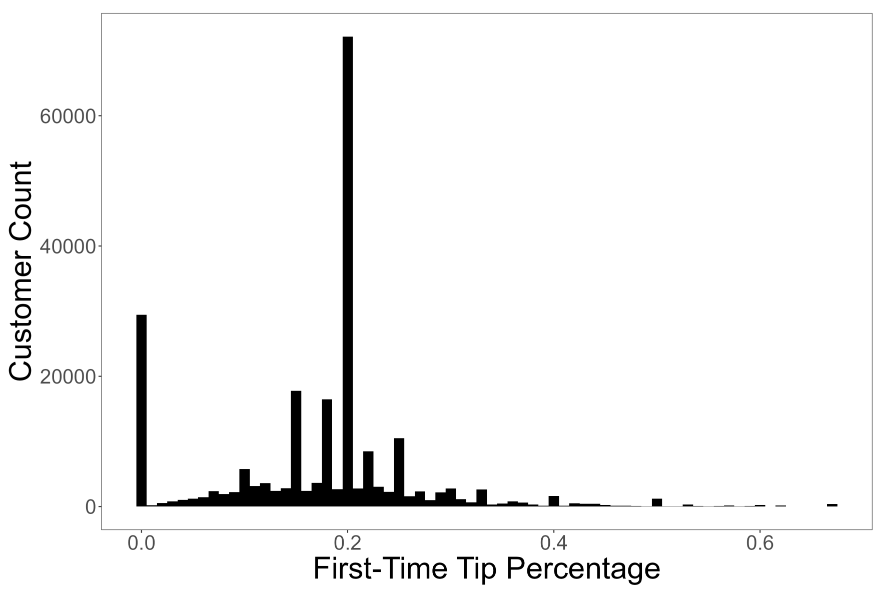
The data come from a salon-management software. Firms, which sometimes own multiple establishments, subscribe to the software to schedule client visits and process transactions. Most subscribing firms self-describe as hair salons, followed by barbershops, nail salons and a few other more niche salon types. The

[3]See Theorem 7 of Levin 2003.

data record, for each service purchased as part of an appointment, a text description of the service, the price, the staff member that performed the service, the start time and duration of the service, and details about how the appointment was booked. Anonymous identifiers allow customers and staff to be tracked within a firm across time, but not across firms. Data on tips were provided separately and linked to the main data at the appointment level.

I define a visit as all transactions associated with a unique customer-day. I limit the data to customers who never have multi-customer visits, who never have multi-location appointments, and whose first and second visit involve only a single worker. I focus on first-time visits, defined as the first visit by each unique customer identifier in the data. To make sure return behavior is well measured, I include only first-time visits that occur 180 days or more before July 31, 2021, the end of the data extract. I exclude first-time visits with a missing tip or where the total price of all services is less than $1. I exclude 23 out of 711 establishments which are missing location or firm information. A histogram of first-time tip percentages is provided in Figure 1. While there is significant dispersion, there is bunching at common tipping prompt thresholds, most notably 0.20.

Figure 1: Histogram of First-Time Tip Percentages



Note. Excludes the top 1% of tips for visual clarity. Bin width is set to 0.01. Percentages are the total tip amount collected from all services on a date divided by the total price of all services on a date.

Summary statistics of several important variables are provided in Table 1. On average customers spend $122 during a visit and spend a little under 2 hours in the salon. Many of the salons in the data are higher-end, with some visits costing customers thousands of dollars. Slightly less than a quarter of visits are prebooked, and customers request a specific staff member during 60.5% of visits. While the vast majority of tips are small in dollar terms, there are rare examples of customers leaving tips of $1,000 or more. When tips are converted to percentages of the price, the average tip percentage is 18.4% with significant variation around the average.

Table 1: Summary Statistics

Statistic	Mean	St. Dev.	Min	Pctl(25)	Pctl(75)	Max
Tip Amount (USD)	18.711	21.177	0.000	7.200	24.000	3,000.000
Price (USD)	122.188	142.225	2.000	50.000	149.000	4,680.000
Percent Tip	0.184	0.162	0.000	0.139	0.200	13.840
Return Indicator	0.306	0.461	0	0	1	1
Duration (hours)	1.965	2.947	0.000	0.750	2.000	222.000
Prebooked Indicator	0.236	0.424	0	0	0	1
Staff Requested Indicator	0.605	0.489	0	0	1	1

A major limitation of the data is that many transactions do not have tip information recorded. One reason is that different establishments appear to have adopted the tipping feature of the salon-management software at different times. As a result, the fraction of first-time visits without tip information is 99.8% in 2017, the first full calendar year after the software company's founding, and decreases to 53.3% in 2021, the last year in my data. In 2021, among establishments that record one tip, the median fraction of first-time visits missing a tip is 21.8%. This suggests that a large share of the missing tip information comes from the gradual adoption of the feature by firms. The tests of incentive relevance should then be thought of as most directly applying to firms with formal tipping policies rather than firms that allow tips to be solicited informally.

Even if a firm activated the tipping feature, customers and workers can strategically circumvent the software. Workers might have reason to do this, either to avoid paying taxes on tips or to keep their tip income private from the salon owner. This is the most concerning group of missing tips, because strategic circumvention could alter the economic forces underlying tipping behavior in ways difficult to predict. One way to circumvent the software is via peer-to-peer payment apps. There is suggestive evidence that strategic substitution via such apps appears to be small. Venmo experienced an outage lasting several hours on December 30, 2019 (Tech 2020). Venmo is regarded as "the first mainstream P2P payment app to market" and had 24 billion transactions in 2019 Quarter 2, compared to 171 million on Zelle ("Strong Growth from Venmo and Zelle Drives P2P Transaction Volume" 2019). If strategic substitution was occurring, the Venmo outage should temporarily increase the fraction of visits tipped within the software. But this does not occur: the fraction of visits with observed tips remains stable before, during and after December 30, 2019.

Another source of missing tips and circumvention is cash tips. The data do not record the method of payment. However, when customers paid in cash, the software by default did not display a tipping prompt, and it is likely that these tips are recorded as missing. The tests of incentive relevance should then be considered as most directly applying to cashless tipping. Whether the results generalize to cash tipping is unclear. There is evidence in the literature that the method of payment impacts the level of tips (Malcman et al. 2025, McCall and Belmont 1996) but no evidence that it impacts the relationship between tips and quality.

6 Results and Robustness

I conduct the test for incentive relevance at various levels of aggregation. Within each group, I regress the first-time tip percentage on a return indicator. I implement the test as a one-sided Wald test of the coefficient on the return indicator, and recover a raw p-value for each unique group at a given level of aggregation. This raw p-value represents the probability of observing that coefficient or larger if tipping is incentive irrelevant. Because there are many groups once the data is disaggregated, I correct for multiple hypothesis testing by adjusting the p-values using the Benjamini-Yekutieli procedure (Benjamini and Yekutieli 2001).

I design the procedure to maximize power. In other words, I try to maximize the probability the procedure detects incentive-relevant tipping when it occurs. I do this two ways. First, I treat each first-time tip as an independent observation and do not cluster standard errors when conducting the Wald tests. Clustered standard errors are typically conservative (Abadie et al. 2023). Second, I control the false discovery rate rather than the family-wise error rate, and set it to be to be 0.05. The false discovery rate controls the expected fraction of incorrect rejections among all rejections, while the family-wise error rate controls the probability of rejecting at least one true null hypothesis. The concepts coincide when all null hypotheses are true (Benjamini and Yekutieli 2001). Thus, my implementation is more powerful than an implementation using either a Bonferroni correction or its improvements, like Hommel 1988. The false discovery rate is also appropriate for the context: this paper is concerned with whether tipping is generally incentive relevant, not whether there exists a single establishment or worker where tips are incentive relevant.

I conduct the test for each state, then each zip code, then each firm, then each establishment and finally each worker. I report the fraction of rejections of incentive-irrelevant tipping as well as the median and interquartile range of the coefficients on the return indicator. The results are presented in Table 2. I fail to reject incentive irrelevant tipping in the vast majority of instances at every level of aggregation. Across groups, the coefficient on the return indicator has a median close to 0 and an interquartile range between -0.0231 and 0.0316. One way to interpret this is that in the majority of groups, the difference in tip percentages between returning and non-returning customers is less than 3.16 percentage points, relative to an overall standard deviation of 19 percentage points.

Table 2: Tests of Incentive-Relevant Tipping

Group	First Time Tips	Groups	Median Obs. per Group	Frac. Reject	p25 Coef.	Median Coef.	p75 Coef.
All	226,017	1	226,017	0.0000	-0.0023	-0.0023	-0.0023
By State	226,017	37	1,312	0.1892	-0.0051	0.0045	0.0220
By Zip Code	225,842	305	331	0.0426	-0.0076	0.0068	0.0212
By Firm	225,819	271	233	0.0295	-0.0076	0.0066	0.0194
By Establishment	225,752	404	240	0.0347	-0.0071	0.0073	0.0222
By Worker	218,014	2953	38	0.0105	-0.0231	0.0041	0.0316

My preferred level of aggregation is at the worker level. The assumption that quality is independent of unobserved shocks to the decision to return is most credible after conditioning on the worker the customer sees. I fail to reject incentive irrelevance in 98.5% of the 2,935 workers. Of course, the number of unique first-time tips observed per group falls as the data are disaggregated, and as a result the power of the test falls. Despite this, it is reassuring I observe 38 first-time tips for the median worker, and that the fraction of cases where incentive irrelevance is rejected remains low even if the data are aggregated up to the establishment or firm level.

I show the result is robust by focusing on transactions which are more likely to be incentive relevant. I provide full tables displaying the results from all robustness exercises in the Online Appendix. I first conduct the test only among transactions where the customer specifically requested the worker for the visit, which represents 136,639 of the original sample of first-time tips. At all levels of aggregation, I fail to reject incentive irrelevance in the vast majority of instances, and at the worker-level I fail to reject in 99.3% of cases. I second conduct the test excluding prebooked appointments, resulting in a sample of 172,605 first-time tips. Firms using the software have the option of collecting tips prior to the service during online booking. Because this tipping happens at arm's length and before the service, it is less likely to be driven by quality or forward-looking behavior. The results are largely similar to the main analysis, with broadly similar rejection rates at all levels of aggregation. The notable exception is full aggregation, where I reject incentive-irrelevant tipping. I view full aggregation to be inappropriate, because sorting by customers based on quality is highly likely across different types of salons, and this quality sorting violates the maintained assumptions of the test.

I also conduct a modified version of the test, where I include date fixed effects in all regressions. In the full sample with date fixed effects and the sample where staff are requested, I continue to fail to reject in the vast majority of cases. In the sample excluding prebooked appointments, I fail to reject in the vast majority of cases with the exception of the full aggregation test, where I reject incentive irrelevance. For the worker level analyses with fixed effects, the rejection rate is higher than in the main analysis but still low. Because of the changing composition of the sample due to gradual adoption, I also conduct the test separately for

2018, 2019, 2020 and 2021. I find low and broadly similar rejection rates across all levels of aggregation, with the exception of full aggregation.

7 Accounting for Tip Variation

The last section argued that tips are not incentive relevant. Still, there is significant variation in tip percentages across customers. This section argues that the data are consistent with tips being driven by social norms that vary across zip codes, rather than business practices that vary across firms.

I begin by showing that defaults appear to be salient, both across customers and within customers across time. The management software allows businesses to designate tip prompts for customers. A large number of tips occur near percentages that were common recommendations at the time. A blog post by the software company suggests that the default tipping prompt is 18%, 20%, 22%. If customers default to the middle suggestion, and firms do not adjust the settings, the default of the default would be 20%, which is the most common tip percentage observed in the data: 70,413 or 31.1% of first-time tips are 20%. To illustrate the salience of the default of the default, I examine the 17,175 returning customers who leave a first-time tip of 20% and who also have an observed second tip. Among these customers, 10,825 or 63.0% leave a second tip of 20%.

I next show that norms appear to be associated with zip codes rather than firms. For each first-time tip I compute the leave-self-out average tip percentage among all other tips at each level of aggregation (state, zip code, firm, establishment, worker). I then regress the first-time tips on the leave-self-out average (LSOA) tip, controlling for date fixed effects and clustering standard errors at the establishment level. The results are presented in Table 3. The leave-self-out tip is positive and statistically significant across all levels of aggregation. The point estimate is typically a little less than 1, which can be interpreted as a 1-percentage-point higher tip of other customers in the same state/zip code/firm/establishment is associated with a little less than a 1-percentage-point higher tip of a focal customer on average. What is surprising is that the amount of variation explained by these regressions is relatively low, suggesting that even among customers who go to the same worker for beauty services there are large unexplained differences in tipping patterns. The tests for incentive irrelevance also imply these differences are not well explained by return behavior.

Further, the amount of variation explained does not rise much when the data are disaggregated below the zip-code level.[4] One possibility is that this is mechanical, because many zip codes have only one establishment using the software.[5] However, the result remains true when the data are limited only to zip codes with at

[4]Some firms have establishments in multiple zip codes, so firms are not fully nested in zip codes.
[5]There are also three zip codes with more than 10 establishments using the software.

Table 3: Associations Between Tips and the Tips of Others

Dependent Variable:	Percent Tip				
Model:	(1)	(2)	(3)	(4)	(5)
Variables					
State LSOA Tip	0.9585***				
	(0.1295)				
Zip Code LSOA Tip		0.9668***			
		(0.0180)			
Firm LSOA Tip			0.9707***		
			(0.0297)		
Establishment LSOA Tip				0.9656***	
				(0.0075)	
Worker LSOA Tip					0.7842***
					(0.0317)
Fixed-effects					
Date	Yes	Yes	Yes	Yes	Yes
Fit statistics					
Observations	226,017	226,005	225,997	225,994	225,638
R^2	0.03138	0.09042	0.09290	0.10196	0.11116

Clustered (Establishment) standard-errors in parentheses
*Signif. Codes: ***: 0.001, **: 0.01, *: 0.05*

least 2 establishments. This pattern is consistent with social norms that vary geographically, but is largely inconsistent with firms using the software or other business practices to differentially manipulate social norms at their establishments.

Ruling out a causal link between specific business practices (i.e. prompt engineering) and tip percentage requires exogenous variation, which is not known to exist in this setting for the time period. I can say that if a causal link does exist, the data are inconsistent with firms differentially exploiting the link. That is, either all firms are using the same business practices with respect to tipping or there is no easy way to influence tips that is profitable for firms. This interpretation is consistent with the earlier result. If tips were incentive relevant, one would expect to see some firms using tips as a way to pass on the monitoring role of the firm to consumers (Azar 2004).

I conclude the section by examining whether the context of the visit predicts tip percentage. To do this, I use several additional pieces of information. As discussed previously, the software records whether an appointment is prebooked and whether the staff assigned was requested by the customer.[6] I also construct the total duration of all services during the visit, to measure whether more contact with a worker is associated with tip percentage. To understand whether peer pressure plays a role, I compute the total number of customers seen by the worker on the date of the appointment. When constructing this variable, I use all transactions, not just those with tip information. I regress first-time tip percentages on the context variables,

[6] I classify a visit as prebooked or staff requested if at least one service on the date is prebooked or had staff requested.

controlling for both worker and date fixed effects, and present the results in Table 4. I cluster standard errors at the establishment level to account for unobserved aspects of the tipping context.

Table 4: Tipping Percentages and Context

Dependent Variable:	Percent Tip				
Model:	(1)	(2)	(3)	(4)	(5)
Variables					
Prebooked Service	0.0095***		0.0163***		
	(0.0019)		(0.0034)		
Staff Requested		7.36×10^{-5}	0.0027		
		(0.0012)	(0.0014)		
Prebooked Service × Staff Requested			-0.0106***		
			(0.0031)		
Total Duration (hours)				-0.0043***	
				(0.0004)	
Customers Seen on Date					0.0004*
					(0.0002)
Fixed-effects					
Worker	Yes	Yes	Yes	Yes	Yes
Date	Yes	Yes	Yes	Yes	Yes
Fit statistics					
Observations	226,017	226,017	226,017	226,017	226,017
R^2	0.17144	0.17110	0.17160	0.17514	0.17116

Clustered (Establishment) standard-errors in parentheses
*Signif. Codes: ***: 0.001, **: 0.01, *: 0.05*

The results show that tip percentages are systematically different based on the context of the visit. After one controls for worker and date fixed effects, prebooking is associated with 0.95-percentage-point higher tips on average. The effect is statistically significant at the 0.05 level and is an 5.16% increase relative to the unconditional mean of 0.201. Requesting staff is not associated with higher or lower tips; however, an interesting pattern arises when both prebooking and requesting are considered together. Among customers who do not request a worker, prebooked appointments have 1.63 percentage point higher tips on average. Among customers that specifically request a worker, prebooked appointments have 0.79-percentage-point lower tips on average. All of these effects are statistically and practically significant, and they suggest that tipping context may impact tip levels.

The software allows tips to be collected both before the appointment (if prebooked) and after the appointment during checkout. Firms can choose either to prompt the customer after, or to prompt the customer twice, before and after. If customers are prompted twice, they are allowed to revise their tip after the appointment. The results are suggestive evidence that double prompting increases tips on average, potentially because it shifts up the effective default for customers. The interaction effect is suggestive evidence that double prompting has the opposite effect on customers who do not request specific staff compared to those who do. Duration of the visit and the number of customers seen by the worker have statistically significant associations with tip percentages, but the point estimates are economically insignificant. An important limitation of all of these estimates is that they very well might be the result of selection of customers, and may not actually represent the causal effect of the context variables on tips holding the customer pool fixed.

8 Conclusion

This paper defines tipping as incentive relevant if tips are expected to increase with service quality. I provide a test for incentive-relevant tipping and apply it to the beauty industry, a setting where repeat interaction is common and relational incentives are strong. I find no evidence of incentive-relevant tipping across a wide range of specifications and levels of aggregation. I then ask what generates the underlying variation in tips. I show evidence consistent with social norms that vary at the zip-code level. I also show evidence that tip percentages are associated with the context of the tip, specifically the default and whether there was a single prompt or two prompts with the opportunity to revise.

A puzzling aspect of these findings is that firms can influence tips, the management software gives them the option to do so, but the data are not consistent with their taking advantage of this opportunity in differential ways. Perhaps there exists a dominant strategy in terms of tip design that most beauty firms follow. Given the relatively scarce evidence in the academic literature and the conflicting views online, this seems unlikely. Alternatively, perhaps firms do not try to systematically influence tipping behavior at all because they are indifferent. If the only potential role of tipping is to offload employee monitoring onto customers, the fact that tipping is not incentive relevant supports this conclusion. On the other hand, because tips benefit workers, showing firm indifference also requires showing that engineering more generous tips comes at an equivalent cost, for example, reduced demand. This paper does not have the exogenous variation needed to verify whether this is true, and therefore leaves the question open for future work.

References

Abadie, Alberto, Susan Athey, Guido W Imbens, and Jeffrey M Wooldridge. 2023. "When should you adjust standard errors for clustering?" *The Quarterly Journal of Economics* 138 (1): 1–35.

Azar, Ofer H. 2004. "Optimal monitoring with external incentives: the case of tipping." *Southern Economic Journal* 71 (1): 170–181.

———. 2007. "Do people tip strategically, to improve future service? Theory and evidence." *Canadian Journal of Economics/Revue canadienne d'économique* 40 (2): 515–527.

Azar, Ofer H. 2009. "Incentives and service quality in the restaurant industry: the tipping–service puzzle." *Applied Economics* 41 (15): 1917–1927.

———. 2010a. "Do people tip because of psychological or strategic motivations? An empirical analysis of restaurant tipping." *Applied Economics* 42 (23): 3039–3044.

———. 2010b. "Tipping motivations and behavior in the US and Israel." *Journal of Applied Social Psychology* 40 (2): 421–457.

———. 2020. "The economics of tipping." *Journal of Economic Perspectives* 34 (2): 215–236.

Benjamini, Yoav, and Daniel Yekutieli. 2001. "The control of the false discovery rate in multiple testing under dependency." *Annals of Statistics,* 1165–1188.

Berlin, Ian, and William Gale. 2025. *No tax on tips: An answer in search of a question.* Accessed on August 25, 2025. September. https://www.brookings.edu/articles/no-tax-on-tips-an-answer-in-search-of-a-question/.

Chandar, Bharat, Uri Gneezy, John A List, and Ian Muir. 2019. *The Drivers of Social Preferences: Evidence from a Nationwide Tipping Field Experiment.* Technical report. National Bureau of Economic Research.

———. 2024. *When and Why Defaults Fail to Change Aggregate Outcomes: Field Experimental Evidence from 40 Million Tipping Observations.* Technical report. Stanford Digital Economy Lab.

Conlin, Michael, Michael Lynn, and Ted O'Donoghue. 2003. "The norm of restaurant tipping." *Journal of Economic Behavior & Organization* 52 (3): 297–321.

Haggag, Kareem, and Giovanni Paci. 2014. "Default tips." *American Economic Journal: Applied Economics* 6 (3): 1–19.

Hommel, Gerhard. 1988. "A stagewise rejective multiple test procedure based on a modified Bonferroni test." *Biometrika* 75 (2): 383–386.

Internal Revenue Service. 2025. *One Big Beautiful Bill Act: Tax deductions for working Americans and seniors.* Accessed on August 25, 2025. https://www.irs.gov/newsroom/one-big-beautiful-bill-act-tax-deductions-for-working-americans-and-seniors.

Kwortnik Jr, Robert J, W Michael Lynn, and William T Ross Jr. 2009. "Buyer monitoring: A means to insure personalized service." *Journal of Marketing Research* 46 (5): 573–583.

Levin, Jonathan. 2003. "Relational incentive contracts." *American Economic Review* 93 (3): 835–857.

Lynn, Michael. 2018. "How motivations for tipping vary with occupational differences in descriptive tipping norms." *Journal of Behavioral and Experimental Economics* 77:1–10.

Lynn, Michael, and Zachary W Brewster. 2018. "A within-restaurant analysis of changes in customer satisfaction following the introduction of service inclusive pricing or automatic service charges." *International Journal of Hospitality Management* 70:9–15.

Lynn, Michael, and Robert J Kwortnik. 2015. "The effects of tipping policies on customer satisfaction: A test from the cruise industry." *International Journal of Hospitality Management* 51:15–18.

Lynn, Michael, and Michael McCall. 2000. "Gratitude and gratuity: a meta-analysis of research on the service-tipping relationship." *The Journal of Socio-Economics* 29 (2): 203–214.

Malcman, Merav, Tal Shavit, Mosi Rosenboim, and Ofer H Azar. 2025. "Tipping Food-Delivery Agents: A Field Study." *Journal of Quality Assurance in Hospitality & Tourism,* 1–17.

McCall, Michael, and Heather J Belmont. 1996. "Credit card insignia and restaurant tipping: Evidence for an associative link." *Journal of Applied Psychology* 81 (5): 609.

McConnell, Brendon, and Marcos Vera-Hernández. 2015. "Going beyond simple sample size calculations: A practitioner's guide." *IFS Working Paper Series.*

Neumark, David, and Emma Wohl. 2024. *Do Higher Tipped Minimum Wages Reduce Race, Ethnic, or Gender Earnings Gaps for Restaurant Workers?* Technical report. National Bureau of Economic Research.

Schmidt, Klaus D. 2003. "On the covariance of monotone functions of a random variable." *Dresdner Schriften zur Versicherungsmathematik* 2003.

"Strong Growth from Venmo and Zelle Drives P2P Transaction Volume." 2019. Accessed: 2025-07-28, November. https://www.emarketer.com/content/strong-growth-from-venmo-and-zelle-drives-p2p-transaction-volume.

Tech, Quicker Better. 2020. "Venmo Goes Down... And Other Small Business Tech News." Accessed: 2025-07-28, *Forbes* (January). https://www.forbes.com/sites/quickerbettertech/2020/01/05/venmo-goes-downand-other-small-business-tech-news/.