THE
ECONOMIC
JOURNAL

The Economic Journal, 1–24 https://doi.org/10.1093/ej/ueaf051
Advance Access Publication Date: 20 June 2025

TOWARDS AN UNDERSTANDING OF REFERENCE-DEPENDENT LABOUR SUPPLY: THEORY AND EVIDENCE FROM A FIELD EXPERIMENT*

Steffen Andersen, Alec Brandon, Uri Gneezy and John A. List

We experimentally study the effects of transitory incentives and income predicted by models of reference-dependent labour supply. Visiting an open-air market in India, we induce experimental variation in: (i) incentives by offering vendors a supplemental wage over the course of two days and (ii) income by, one morning, administering a substantial overpayment for a good sold in the market. We find that incentives have a staggered effect on hours worked: initially vendors do not respond, but by the second day their hours worked increases. In response to the overpayment, we find that both hours worked and a measure of effort are unaffected. Collectively these findings suggest vendors supply their labour according to a model where reference dependence is moderated by experience with an incentive regime or by the time at which income is accumulated.

JEL codes: J22, D91, C93

Theories of reference-dependent preferences predict that labour supply may respond negatively to transitory incentives (Camerer *et al.*, 1997; Farber, 2005, 2008, 2015; Köszegi and Rabin, 2006; DellaVigna, 2009). Underlying this prediction are income effects that, once a worker approaches their reference level of earnings, can dominate an incentive's substitution effect. While a large collection of studies on incentive and income effects are consistent with this prediction, few of these studies are experimental. Experimental estimates of these effects can overcome endogeneity concerns with observational and quasi-experimental estimates and help advance our understanding of reference-dependent labour supply.

We experimentally estimate incentive and income effects by designing, conducting and analysing two field experiments on the labour supply of vendors in an open-air market in India. These vendors face a similar decision problem to the samples of bike messengers, farmers, fishermen, rickshaw drivers, taxicab drivers, salespeople and undergraduates studied in the literature on reference-dependent labour supply. Much like these other samples of workers, the vendors in our experiments have flexible work scheduling with which they can respond to the highly variable earnings they face from one day to the next.

To estimate the effects of transitory incentives, we conducted the Market Survey Incentive Experiment with 250 vendors. The vendors in this experiment were randomly assigned to a

* Corresponding author: Alec Brandon, Carey Business School, Johns Hopkins University, 100 International Drive, Baltimore, MD 21202, USA. Email: alec.brandon@jhu.edu

This paper was received on 30 March 2021 and accepted on 18 June 2025. The Editor was Steffen Huck.

The data and codes for this paper are available on the Journal repository. They were checked for their ability to reproduce the results presented in the paper. The replication package for this paper is available at the following address: https://doi.org/10.5281/zenodo.15626937.

Stefano DellaVigna, Justin Holz, Casey Mulligan, David Novgorodsky, Zongjin Qian and George Wu provided helpful comments and suggestions. James Perry provided excellent research assistance. The experiments studied in this article received approval from the Ethics Council at Copenhagen Business School (CBS-24-016).

treatment or control group. While vendors in the control group were left untouched, over the course of two consecutive days, treatment group vendors had the hourly return on their labour supply increased by INR10, 30 or 60. Relative to the prevailing wage, these transitory incentives were substantial. Conversations with vendors suggested that, in a typical hour, vendors earned INR10, which indicates that our smallest wage supplement increased hourly earnings by 100%. We frame the wage supplements as an incentive for hourly completion of the market survey, which tasked treatment vendors with keeping a tally of the number of customers visiting their shop in each hour. In response, we collected hourly observations of whether treatment and control vendors worked in the market.

We estimate income effects with seventy-nine vendors in the Betel Nut Income Experiment. In this experiment, vendors selling betel nuts were selected and randomly assigned to a treatment or control group. Control group vendors were again left untouched, while one morning treatment vendors had their income altered by a Westerner paying INR500 for a pack of betel nuts that typically sold for INR17. Such an overpayment was approximately a week's worth of earnings. Administering such a large overpayment was straightforward because vendors in the market did not post prices and instead haggled over prices with potential customers. In response to the overpayment, we collected hourly observations of whether treatment and control vendors worked in the market and a measure of the effort exerted in haggling with potential customers. To elicit this measure of effort, we hired a team of locals to act as confederates and incentivised their haggling with vendors over betel nut transactions. In total, on the day of the overpayment, our confederates elicited the initial price offer and final price of 255 transactions with fifty-five of the seventy-nine vendors in the experiment.

We motivate the design and analyses of these experiments by developing econometric models that characterise our two parameters of interest: the effect of incentives on hours worked and the effect of income on the decision to quit working for the day. Observational and quasi-experimental estimates of these parameters are frequently cited as evidence of reference-dependent labour supply (see, e.g., DellaVigna, 2009; O'Donoghue and Sprenger, 2018). Our econometric models formalise the link between these parameters and the variation an experiment must induce to identify these parameters. The models also formalise the threats to the identification of the parameters. We show that a mean independence assumption is needed to overcome these threats. Intuitively, this assumption places three restrictions on labour supply behaviour. First, randomisation of the experimental incentives and income must successfully balance the non-experimental determinants of labour supply; second, the effects of the experimental incentives and income must not intertemporally spillover onto the labour supply of their receipients; and, third, the effects of the experimental incentives and income must not spillover onto the labour supply of non-receipients.

Using data generated by our experiments, we scrutinise the validity of each of these restrictions. The first restriction is evaluated by comparing treatment and control group labour supply outcomes before the two experiments were administered. The second restriction cannot be so easily tested in either experiment, but intuition suggests it is most likely to be violated in the Betel Nut Income Experiment because receipt of the overpayment could lead treatment vendors to believe that an additional overpayment from a Westerner would be earned if they remained in the market. This limitation motivates our elicitation of vendor effort by locals. Fehr and Goette (2007) show that the effects on hours worked predicted by reference-dependent labour supply models extend to effort. Furthermore, effort responses to the overpayment, as measured by haggling with locals, are less likely than hours worked to be biased by changes in beliefs of a subsequent appearance by a Westerner. As a result, we can scrutinise whether the second restriction holds by testing whether the effect of the overpayment on quitting differs from the effect on effort as measured by the prices elicited by the haggling of our confederates. To address the third restriction, we examine whether control vendor labour supply is altered by spillover effects induced by variation in the share of nearby vendors receiving the wage supplement in the Market Survey Incentive Experiment. When we undertake each of these exercises, we find no evidence of the data violating the restrictions necessary for identification.

We report two sets of findings. First, in the Market Survey Incentive Experiment, we find a staggered hours worked response to transitory incentives. On the first day that treatment vendors received the wage supplement, we find that for every INR10 of the wage supplement vendors reduce their time at work by 0.02 to 0.03 hours, although this effect is statistically indistinguishable from a null effect with standard errors of 0.03. However, on the second day of the experiment, the hours worked by treatment vendors is highly responsive to the wage supplements. We find that every additional INR10 of the wage supplement, which corresponds to a 100% increase in the wage, increases time in the market by approximately 10 minutes, which equates to a 2% increase in hours worked. While this implies that labour is supplied inelastically, the estimated parameter is highly statistically significant. Driving this effect is that treatment vendors show up earlier and quit later in the day. Additional analyses of the Market Survey Incentive Experiment help rule out explanations for the staggered response to the incentives such as the credibility of the incentive varying by day and vendors needing time to plan their labour supply response to the incentives.

Second, in the Betel Nut Income Experiment, we find that, after treatment vendors received the overpayment, they quit working at rates that are statistically indistinguishable from control vendors. Depending on the controls used, the overpayment altered vendor quit rates by −0.04 to 0.02 percentage points, with standard errors of 0.12. We also find that, after the overpayment, treatment vendors did not exert more or less effort than control vendors in their haggling with our confederates.

Our study offers contributions to two strands of research on reference dependence. The first is the experimental literature on the incentive and income effects predicted by reference-dependent models of labour supply. Most of the experimental research on reference-dependent labour supply studies the effects of varying reference earnings (Abeler *et al.*, 2011; Gill and Prowse, 2012; Hossain and List, 2012; Levitt *et al.*, 2016; Pierce *et al.*, 2020; Fryer *et al.*, 2022; see Ferraro and Tracy, 2022, for an overview). Two important exceptions are Fehr and Goette (2007) and Dupas *et al.* (2020). Fehr and Goette (2007) randomise a two-week long increase of 25% in the commission rate earned by forty-two bicycle messengers. While these bicycle messengers work longer hours in response to the experimental incentive, they also reduce their effort and, consistent with reference dependence, the total effect of the experimental incentive on revenue generated is moderated by an elicited measure of loss aversion. Dupas *et al.* (2020) recruit 259 bicycle taxi drivers in Kenya to participate in periodic lotteries conducted by their research staff. They find that winning a lottery, which on average pays out a day's worth of earnings, has a statistically indistinguishable effect on labour supply. Relative to these two studies, we contribute to the literature by formalising the restrictions needed for identification, more extensively testing the validity of these restrictions, designing a more natural framing of experimental variation in income, and inducing higher-stakes variation in incentives and income.

Second, our findings contribute to the theoretical literature on reference dependence. Understanding the relationship between labour supply, incentives and income is crucial for understanding labour market dynamics (Lucas and Rapping, 1969). Yet conventional models of labour supply cannot explain the staggered response to incentives we find in the Market Survey Incentive Experiment. Furthermore, leading reference-dependent models of labour supply cannot explain the non-response to the overpayment that we find in the Betel Nut Income Experiment (e.g., Köszegi and Rabin, 2006). Two alternative approaches to modelling reference-dependent labour supply are attractive.[1] The first is motivated by experiments on experience effects in reference dependence (List, 2003, 2004, 2011; Engelmann and Hollard, 2010) and explains the staggered response in the Market Survey Incentive Experiment with reference dependence that attenuates when vendors gain experience with the wage supplement. The second is motivated by observational evidence on intraday adaptation of reference levels of earnings in Thakral and Tô (2021). Such adaptation explains the null effect in the Betel Nut Income Experiment as an artefact of the time at which the overpayments were administered. Because treatment vendors received the overpayments early in their workdays, their reference earnings level could adapt.

We also offer smaller contributions to other strands of the literature. The staggered response to incentives in the Market Survey Incentive Experiment is a novel contribution to a large body of empirical research on incentive effects in labour supply (Rizzo and Blumenthal, 1996; Camerer *et al.*, 1997; Oettinger, 1999; Chou, 2002; Rizzo and Zeckhauser, 2003, 2007; Goette *et al.*, 2004; Contandriopoulos and Perroux, 2013; Nguyen and Leung, 2013; Doran, 2014; Chen and Sheldon, 2015; Farber, 2015; Stafford, 2015; Goldberg, 2016; Sheldon, 2016; DellaVigna *et al.*, 2017; Giné *et al.*, 2017; M.K. Chen *et al.*, 2019; K.-M. Chen *et al.*, 2020; Richards, 2020; Angrist *et al.*, 2021; Fisher, 2022; He *et al.*, 2022). Furthermore, the results of the Betel Nut Income Experiment contribute to the literature that measures the labour supply effects of daily income accumulation with observational data (Farber, 2005, 2015; Crawford and Meng, 2011; Ran *et al.*, 2014; Agarwal *et al.*, 2015; Morgul and Ozbay, 2015; MacDonald and Mellizo, 2017; Martin, 2017; Hammarlund, 2018; Schmidt, 2018; Avoyan *et al.*, 2020; Thakral and Tô, 2021; Dodini, 2022; Saia, 2022). These studies consistently find income effects that cannot be detected in the Betel Nut Income Experiment.

The remainder of this paper is organised as follows. In Section 1 we formalise our strategy for identifying the effects of incentives and income on labour supply outcomes. Section 2 describes the design of our field experiments and evaluates the restrictions imposed on the data by our identification strategy. In Section 3 we present our findings and Section 4 concludes.

1. Identification

In this section, we formalise our strategy for identifying the effects of transitory incentives and income. In so doing, we highlight threats to identification and consider how data can be used to evaluate these threats. Throughout we take the predictions of reference-dependent labour supply models as given and we point readers interested in the derivation of these predictions to Online Appendix B.

[1] A third possibility is that vendors evaluate their gain–loss utility by selecting some of the income they accumulate from vending and comparing the sum of these selected sources of income to a target. For example, vendors may not select the overpayment for the measure of daily income they use to determine their gain–loss utility. We do not emphasise this possibility because it is hard to reconcile with the staggered response to incentives that we see in the Market Survey Incentive Experiment.

1.1. *Setting and Notation*

Consider a sample of vendors who operate their own businesses in an open-air market. Each day, when the market is open, vendors are free to decide their hours of work. Let $j \epsilon \{1, \cdots, J\}$ index each vendor in the sample, $s \epsilon \{1, \cdots, S\}$ index each day, and $t \epsilon \{1, \cdots, T\}$ index time over the course of each day. Throughout, we assume that t indexes each hour of the day.

The outcomes of interest are daily hours worked and the intraday decision to quit working for the day. Let Y_{js} denote the daily hours of work decision for vendor j on day s and let D_{jst} indicate whether vendor j quits working on day s at time t.

These outcomes may be influenced by a transitory incentive and the accumulation of daily income. Let I_{js} denote the incentive to work for vendor j on day s. Implicit in this representation is that vendors face different work incentives on different days, but the incentive to work is stable over the course of a given day. Furthermore, let M_{jst} denote the income accumulated by vendor j on day s at time t.

1.2. *Parameters of Interest*

Our objective is to identify two parameters of interest. These parameters are motivated by models of reference-dependent labour supply, which predict that transitory increases in the rate of hourly earnings can negatively affect hours worked and that accumulated income positively affects the probability of quitting. These propositions can be modelled with the following equations,

$$Y_{js} = \alpha I_{js} + U_{js}, \tag{1}$$

where U_{js} captures non-incentive determinants of hours worked, and,

$$D_{jst} = \beta M_{jst} + V_{jst}, \tag{2}$$

where V_{jst} captures non-income determinants of quitting work for the day.

Our parameters of interest are the coefficients α and β. Relating these coefficients to the predictions of reference-dependent models of labour supply is straightforward. While conventional economic models predict $\alpha > 0$ and $\beta \approx 0$ (e.g., Lucas and Rapping, 1969), models of reference dependence predict that $\alpha < 0$ is possible and that $\beta > 0$ (Camerer *et al.*, 1997; Farber, 2005, 2008, 2015; Köszegi and Rabin, 2006; DellaVigna, 2009). Some models of reference-dependent labour supply also propose specific mechanisms that underly the formation of reference earnings. Such mechanisms can be incorporated into the models in (1) and (2) by allowing the coefficients to vary by vendor, day, time, or other relevant features. We discuss this possibility in greater detail in Section 3.3.

1.3. *Assumptions*

We identify our parameters of interest by experimentally varying a supplement to hourly earnings and a shock to daily income. The success of this strategy depends on the validity of mean independence assumptions, which we present formally in this subsection. We also consider the restrictions that these assumptions place on vendor labour supply and strategies to empirically evaluate these restrictions.

Let W_{js} denote the magnitude of a wage supplement and X_{jst} denote the magnitude of a shock to daily income. These variables are respectively related to incentives and income according to,

$$I_{js} = W_{js} + \tilde{I}_{js},$$

where $\tilde{I}_{js}$ denotes non-experimental work incentives, and

$$M_{jst} = X_{jst} + \tilde{M}_{jst},$$

where $\tilde{M}_{jst}$ denotes non-experimental daily income accumulation.

To identify our parameters of interest, we must invoke a mean independence assumption between the non-experimental determinants of labour supply and an experimental wage supplement or income shock. Such an assumption to identify the effect of incentives on hours worked, α, requires that,

$$E\left[\tilde{U}_{js}|\boldsymbol{O}_{js},\ W_{j1}, \cdots,\ W_{jS}\right] = E\left[\tilde{U}_{js}|\boldsymbol{O}_{js}\right] \text{ for } s\epsilon\,\{1,\ \cdots,\ S\}, \tag{3}$$

where $E[\cdot]$ is the expectations operator, $\tilde{U}_{js} = U_{js} + \alpha\tilde{I}_{js}$ captures the non-experimental determinants of hours worked, and $\boldsymbol{O}_{js}$ is a vector of observables that controls the variation that identifies α. For example, if $\boldsymbol{O}_{js}$ includes a fixed effect for each vendor, then within vendor variation will identify α. To identify the effect of income on the decision to quit working, β, we must make a similar assumption for the income shock variable,

$$E\left[\tilde{V}_{jst}|\boldsymbol{O}_{jst},\ X_{j11},\ \cdots, X_{jST}\right] = E\left[\tilde{V}_{jst}|\boldsymbol{O}_{jst}\right] \text{ for } s\epsilon\,\{1,\ \cdots, S\} \text{ and } t\epsilon\,\{1,\ \cdots,\ T\}, \tag{4}$$

where $\tilde{V}_{jst} = V_{jst} + \beta\tilde{M}_{jst}$ captures the non-experimental determinants of quitting and $\boldsymbol{O}_{jst}$ is a vector of observables.

Intuitively, the assumptions characterized by equations (3) and (4) respectively require that treatments alter labour supply only through their impact on incentives and income. At least three threats to this requirement seem plausible. First, randomisation may fail to balance the non-experimental determinants of labour supply. Second, the treatments may unintentionally induce intertemporal spillovers. For example, vendors may work long hours to earn a wage supplement and then rest on subsequent days because of the effort they exerted prior. In the same vein, after receiving an income shock, vendor beliefs about receiving a subsequent shock may change and alter their supply decisions. Third, the treatments may unintentionally spillover from treated to controlled vendors.[2] For example, if incentives or income cause treated vendors to exit the market, it may alter the return on working for control vendors and bias findings.

While these restrictions are strong, they have empirical content that can be tested. The first restriction can be tested with labour supply outcomes observed prior to the receipt of a wage supplement or income shock. These outcomes should be balanced across recipients and non-recipients. Statistically, significant differences in pre-receipt labour supply outcomes would reject this restriction. The second restriction, which is on intertemporal spillovers, can be easily navigated for the identification of incentive effects, α, by omitting labour supply outcomes observed after the wage supplements are offered. However, no such workaround is possible for the identification of income effects on the subsequent decision to quit working for the day, β. Instead, we adopt the following strategy: we design the administration of the income shock to minimise the risk of violating the restriction and we generate auxiliary data to evaluate the effects

[2] For simplicity, this restriction is implicit in equations (3) and (4). Adding the experimentally induced incentive and income variables for all vendors to the condition in the conditional expectation on the left hand side of equations (3) and (4) would make this restriction explicit.

of income on labour supply outcomes that are less susceptible to violations of this restriction. Under (4), we should find consilience across our analyses of the quitting and auxiliary data. We discuss the design of the income shock's administration and auxiliary data in greater detail in Section 2.2. The third restriction, which is on spillovers between vendors, can be evaluated by examining whether the vendors who form the control groups in our wage supplement experiment are influenced by the share of vendors near them who receive the supplement. Finding statistically significant differences in the hours worked by control group vendors as the share of vendors receiving the supplement increases would reject this restriction.

Under these restrictions, identification of our parameters of interest is straightforward. The effect of incentives, α, is identified by the coefficient δ in,

$$Y_{js} = \delta W_{js} + \boldsymbol{O}_{js}{}'\boldsymbol{\kappa} + \tilde{U}_{js}, \tag{5}$$

and the effect of income accumulation, β, is identified by the coefficient γ in,

$$D_{jst} = \gamma X_{jst} + \boldsymbol{O}_{jst}{}'\boldsymbol{\kappa} + \tilde{V}_{jst}. \tag{6}$$

Next, we describe the experimental design that allows for the estimation of our parameters of interest.

2. Design

To estimate our parameters of interest, we design and conduct a transitory wage supplement and overpayment experiment with 329 vendors in the Burra Bazar in Shillong, India (for more on this setting see Andersen *et al.*, 2018).[3] The open-air market is the city's main poultry, meat, vegetable, produce and merchandise market. Vendors in this market work as independent contractors. They are free to set their hours of work and their hourly earnings are determined by their shop's hourly revenues. Prices in the market are not posted. Instead, vendors haggle with customers to determine the price of each transaction, which are conducted in cash. Next, we discuss the design of our wage supplement and overpayment experiments in greater detail. Both discussions also provide empirical tests of the mean independence assumptions that underly our identification strategy of incentive and income effects.

2.1. *Market Survey Incentive Experiment*

We experimentally vary transitory wage for 250 vendors with the Market Survey Incentive Experiment. While this experiment was conducted over the course of two sessions in 2006 and 2007 and there were slight differences in administration between the two sessions (e.g., number of days observed), this subsection describes the major features common to both sessions and a more thorough description of each session is provided in Online Appendix A. Across both sessions, vendors in the experiment were randomly assigned to a treatment or control group. Vendors in the control group were left untouched, while treatment vendors were offered a financial incentive for each hour in which the market survey was completed. To complete the market survey in a given hour, vendors had to record a tally of customers who visited their shop and turn the survey in when research staff made their hourly visit to collect the survey. Completing this survey was

[3] See also Andersen *et al.* (2008), Gneezy *et al.* (2009), Andersen *et al.* (2011, 2013, 2018) and Hoffman *et al.* (2011) for studies conducted in this region.

intended to require minimal effort. We did not evaluate the quality of surveys that were completed and, as a result, we do not make use of the collected responses.[4] Research staff attempted to check these treatment vendor tallies in each hour that the market was open. Each tally that was checked earned the vendor the incentive for that hour. Treatment vendors received one of three incentive levels for each hour that the market survey was completed: INR10, INR30 or INR60. While no administrative data exists to calculate a baseline of hourly earnings against which these incentives can be benchmarked, conversations between research staff and vendors suggested that hourly earnings were around INR10.

Before running the Market Survey Incentive Experiment, vendors were selected for the sample by research staff. To facilitate the administration of the experiment, vendors were selected in fourteen distinct geographic clusters. Treatment and control status was randomly assigned within each cluster, or stratum, and every treatment vendor in a stratum was offered the same level of incentive. That is, treatment vendors in the same geographic cluster were never offered a different incentive. All told, 112 of the vendors were assigned to the control group, while 138 vendors were assigned to the treatment group. Sixty of the treatment vendors were offered the INR10 incentive, fifty-one were offered the INR30 incentive, and twenty-seven were offered the INR60 incentive.

Time in the Market Survey Incentive Experiment was divided into three mutually exclusive periods: Pre-Incentive, Incentive, and Post-Incentive. During the Pre-Incentive Period research staff hourly recorded whether vendors were present in the market. At the end of the final day of the Pre-Incentive Period, treatment vendors were recruited to complete the market survey over two consecutive days. If a treatment vendor was absent from the market on the final day of the Pre-Incentive Period, then recruitment occurred when the vendor returned to the market. Because completion of the survey required negligible effort and the incentives were so lucrative, only two vendors declined participation. Treatment and control vendors then began a two-day long Incentive Period. During this period, research staff continued their hourly observation of whether vendors were present in the market and attempted an hourly collection of each treatment vendor's market survey. Treatment vendors were paid the incentive for each market survey collected when they quit working for the day. Then the Post-Incentive Period began, during which research staff continued hourly observation of whether vendors were present in the market.

The data observed in the Market Survey Incentive Experiment can be easily related to the empirical model in (5) that we seek to estimate. Hours worked, Y_{js}, is measured by summing the number of times vendor j was observed by research staff on their hourly visits on day s. The wage supplement variable, W_{js}, is 0 for all vendors during the Pre-Incentive Period. During the Incentive Period, this variable remains 0 for control vendors, whereas the variable switches to 10, 30 or 60 for treatment vendors, depending on the incentive level offered for completing the market survey. If a treatment vendor declined the incentive offer or was absent from the market on the final day of the Pre-Incentive Period, then their wage supplement variable is still set to the incentive offered. The vector of observables, $\boldsymbol{O}_{js}$, dictated by the design of the Market Survey Incentive Experiment is a fixed effect for each stratum. This vector can also be augmented to determine the sources of variation that identify the effect of the wage supplement. For example,

[4] Several examples of filled out market surveys can be found in Online Appendix D. These examples are provided because they appear representative of the surveys collected over the course of the Market Incentive Experiment. Copies of all the collected surveys are available in our replication files. There, researchers can find scans that have integer counts and tallies of the number of customers who visited a treatment group vendor in each hour of a given day. Please note that because control group vendors were left untouched, they did not fill out any surveys.

Table 1. *Pre-Incentive Period Balance in Market Survey Incentive Experiment.*

	(1) Worked	(2) Hours worked	(3) Start time	(4) Break hours	(5) Quit time
Treatment diff.					
INR10	0.08	0.42	0.24	−0.05	−0.04
	(0.05)	(0.51)	(0.20)	(0.05)	(0.13)
INR30	−0.02	−0.39	0.30	−0.08	0.00
	(0.02)	(0.21)	(0.17)	(0.05)	(0.08)
INR60	0.00	0.29	−0.29	0.00	0.00
	(0.00)	(0.24)	(0.24)	(0.00)	(0.00)
Control	0.97	8.28	9.66	0.05	18.28
	(0.02)	(0.18)	(0.09)	(0.02)	(0.05)
Vendors	250	250	245	245	245

Notes: This table reports the coefficients from a regression of a labour supply outcome on the incentive offered to treatment group vendors and stratum fixed effects. The control group average is reported by taking the weighted average of control vendor outcomes across each stratum. The regression is fit with outcomes that are observed during the first day of the Pre-Incentive Period, during which treatment group vendors have yet to be notified of the incentive they will be offered during the Incentive Period. See Online Appendix Table C1 for estimates of the same coefficients obtained with all Pre-Incentive Period data. Clustered standard errors that are robust to heteroscedasticity and vendor autocorrelation are reported in parentheses under each coefficient.

populating $\boldsymbol{O}_{js}$ with vendor fixed effects would only use within vendor variation to identify our parameter of interest.

The data observed in the Market Survey Incentive Experiment also allow us to evaluate the mean independence assumption that underlies our strategy for identifying the parameter of interest. In Section 1.3 we discussed three restrictions that this assumption places on vendor labour supply. Table 1 presents an evaluation of the first restriction. This restriction requires that, before the wage supplements are offered, the non-experimental determinants of labour supply are balanced across the control and treatment vendors receiving the different wage supplements. To evaluate this restriction, we regress several labour supply outcomes constructed from data observed in the Pre-Incentive Period on the incentive level offered in the Incentive Period, with fixed effects included for each stratum. Inference is conducted with standard errors clustered by vendor.

Table 1 shows that the Pre-Incentive Period data does not reject the first restriction of the mean independence assumption. This is shown for the probability of a vendor working, total hours worked, start time of day, total break hours taken and quit time of day. Across each of these outcomes, we find that the treatment levels are statistically indistinguishable from the control group at the 5% level of significance. This provides an empirical basis for the first restriction of the mean independence assumption needed to identify the labour supply effect of incentives.

Tables 1 also provides an empirical description of the vendors in the Market Survey Incentive Experiment. On average, the vendors in our sample work nearly every day that we collect data in the market, they work nearly 8.3 hours per day, show up to the market just before 10 a.m., take very few breaks, and exit the market just after 6 p.m. These estimates highlight the margins over which vendors could respond to the wage supplements. Intensive margin responses were somewhat constrained, but possible, as the market opened between 6 a.m. and 8 a.m. and closed at 7 p.m. Extensive margin responses, though, were highly constrained because vendors nearly

Table 2. *Incentive Period Spillovers on Control Group Vendors in Market Survey Incentive Experiment.*

	(1) Hours worked	(2) Start time	(3) Break hours	(4) Quit time
Share treated	−0.07	−0.53	0.45	−0.15
	(0.49)	(0.42)	(0.25)	(0.28)
Constant	8.31	9.87	−0.08	18.33
	(0.20)	(0.13)	(0.06)	(0.08)
Vendors	112	109	109	109
Observations	224	218	218	218
R^2	0.11	0.21	0.27	0.43

Notes: This table reports the coefficients from a regression of control vendor labour supply outcomes on the share of vendors receiving the treatment in a stratum and stratum fixed effects. The reported constant is calculated by taking the weighted average across each stratum of the predicted outcome when the share of vendors receiving the treatment is 0. The regression is fit with outcomes that are observed during the first two days of the Incentive Period, during which treatment group vendors receive a wage supplement for each hour that they complete the market survey. See Online Appendix A for more on the timing of the Incentive Period during the two sessions of the Market Survey Incentive Experiment. Clustered standard errors that are robust to heteroscedasticity and vendor autocorrelation are reported in parentheses under each coefficient.

always worked. While these features of the market likely diminished the extent to which vendors could work more in response to the incentives, they allow our analysis to speak more directly to theories of reference-dependent labour supply, as such theories diverge most sharply from neoclassical models on the extensive margin of incentive effects.

In Table 2, we provide an evaluation of the third restriction imposed by the mean independence assumption that underlies our identification strategy.[5] This restriction requires that the effect of the wage supplement offered to a treatment vendor does not have spillover effects on control vendor labour supply. We evaluate this restriction by examining whether, during the Incentive Period, control vendors alter their labour supply as the share of vendors who are treated increases in each stratum. Such an examination is conducted by regressing total hours worked, start time of day, total break hours taken, and quit time of day on the share of vendors in a stratum receiving the treatment, with stratum fixed effects also included. We omit a test with the probability of a vendor working as the outcome because the share treated is perfectly collinear with the strata fixed effects. Inference is conducted with standard errors clustered by vendor. Table 2 shows that, consistent with the mean independence assumption, the share of vendors receiving the treatment did not have a statistically significant effect at the conventional threshold of 5% on the labour supply outcomes of control vendors observed during the Incentive Period. For example, column 1 shows that a 1 percentage point increase in the share of vendors receiving the treatment reduces control vendor hours worked by 0.07 hours, with a standard error of 0.49 on this estimated effect.

2.2. *Betel Nut Income Experiment*

We experimentally vary the income of seventy-nine vendors with the Betel Nut Income Experiment. While our intended sample size was eighty-five vendors, six were absent from the market on the morning in which the overpayment was administered. These vendors met the same

[5] See Section 1.3 for a discussion of the bearing of the second restriction on the Market Survey Incentive Experiment.

eligibility criteria as the vendors in the Market Survey Incentive Experiment and one additional criterion: they all sold a non-perishable good called betel nuts. Vendors in the experiment were randomly assigned to a treatment or control group. One morning in 2006, the twenty vendors assigned to treatment had their earnings altered by a member of our research staff who paid INR500 for a pack of betel nuts, while the fifty-nine vendors assigned to control received no such change to their earnings. Paying INR500 for a pack of betel nuts was a substantial overpayment, as the typical transaction for betel nuts was approximately INR17.

This overpayment was likely to have to caused treatment vendors to determine that they reached their daily level of reference income. As discussed in Section 2.2, conversations with vendors indicated that the typical hourly rate of earnings was INR10, which suggests that, in isolation, the overpayment constituted nearly a week of earnings. Furthermore, the context in which the overpayment was received was likely to have caused treatment vendors to incorporate the income shock into their mental account of daily earnings. Unlike Dupas *et al.* (2020), which uses a lottery conducted by research staff to evaluate reference-dependent labour supply, the income shock in our experiment is received in the same context as their typical daily earnings. While it is still possible that vendors omit the overpayment in their evaluation of gain–loss utility, such an omission is not a feature of leading models of reference dependence where income that accumulates from completing a job is evaluated relative to a targeted level of earnings (e.g., Köszegi and Rabin, 2006).

In response to the overpayment, we collected hourly observations of whether treatment and control vendors were in the market. These observations were collected the day of the overpayment and the day after the overpayment. These observations allow us to observe the data necessary to estimate the empirical model of income effects in (6). The decision to quit working variable, D_{jst}, is constructed by creating a vector that starts at 0 in the first hour that each vendor is observed in the market and switches to 1 when the vendor exits the market for the day. The income shock variable, X_{jst}, takes on one of two values. All vendors in the experiment start the day with the income shock variable equal to 0. Once the income shock is administered, this variable switches to 500 for treatment vendors and remains at 0 for control vendors. The data we observe also allow us to populate the vector of observables, $\boldsymbol{O}_{jst}$, with controls such as when the vendor started their day.

These data also allow us to evaluate the first restriction imposed by the mean independence assumption required for the identification of our parameter of interest. Recall that in Section 1.3 we discussed how the first restriction of the mean independence assumption in (4) requires that non-experimental labour supply determinants are mean independent of future receipt of the income shock. We can evaluate this by testing whether, on average, treatment and control group vendors started their days at different times. Column 1 of Table 3 presents the results of this test, which finds that, on average, control group vendors started their day at the 9.29th hour of the day (i.e., 9:21 a.m.), while treatment group vendors started at the 9.05th hour of the day (i.e., 9:03 a.m.). Consistent with the mean independence assumption, the p-value of 0.11 on the 18 minute difference between start times is not statistically significant at the conventional threshold of 5%.

These data, however, do not allow us to evaluate the second restriction required by the mean independence assumption. This restriction holds that the non-experimental determinants of labour supply are mean independent of prior receipt of the overpayment. A major concern is that this restriction will not hold because, after receiving the overpayment, treatment vendors will perceive that there is a greater chance of receiving an additional overpayment if they remain in the market. To diminish this concern with dynamic confounds, we had the only Westerner on our research

Table 3. *Pre-Income Shock Balance in Betel Nut Income Experiment.*

	(1) Start time	(2) Initial offer	(3) Final price
Control	9.29	17.87	17.16
	(0.09)	(0.91)	(0.91)
Treatment	9.05	15.95	15.13
	(0.15)	(0.99)	(0.85)
Control = Treatment	p-value = 0.182	p-value = 0.093	p-value = 0.067
Vendors	79	52	52

Notes: This table reports the average outcomes for treatment and control group vendors prior to the overpayment in the Betel Nut Income Experiment. The outcomes considered are the time vendors arrived to the market on the day of the income shock, and the initial price offer and final price elicited by confederates before the income shock. Average initial and final prices control for the type of betel nut. The p-value reports a test of equality between treatment and control group averages. This inference is conducted with clustered standard errors that are robust to heteroscedasticity and vendor autocorrelation.

staff administer the overpayments. Because the odds of receiving an additional overpayment from a Westerner scale with the probability of a Westerner entering the market, we hoped this design decision would limit concerns with dynamic incentives, as Westerners were rarely seen in the market.

We also collected additional labour supply outcomes that are less susceptible to concerns with dynamic confounds. To do so, we hired six locals to act as confederates. These confederates, who were kept blind to vendor treatment status, were tasked with eliciting a measure of the effort exerted by vendors. To elicit such a measure, our confederates were periodically assigned a vendor and endowed with INR30 to haggle over the price of a pack of betel nuts. Upon approaching a vendor, our confederates were instructed to first elicit an initial price offer for one of two types of betel nuts. The confederates were then free to haggle with the vendor until a final price was agreed upon. Haggling was incentivised, with the confederates earning the difference between their endowment and the agreed upon price. In total, confederates executed 258 transactions with fifty-eight vendors in the experiment: nineteen from treatment and thirty-nine from control. These transactions occurred on three days: before the overpayment, the day of the overpayment (after the overpayment had been administered), and the day after the overpayment. Columns 2 and 3 of Table 3 provide a description of the data collected before the day of the overpayment.

Regressing the initial offers and final transaction prices on receipt of the overpayment allows us to estimate income effects that are less susceptible to concerns with dynamic confounds. This claim can be substantiated with a reference-dependent model of labour supply and bargaining effort where receiving the treatment increases vendor beliefs of an additional overpayment. While such a model has an ambiguous prediction on the effect of an overpayment on probability of exiting the market for the day, the prediction on effort is unambiguous. Relative to control, treatment vendors are predicted to exert less effort. These predictions are formally developed in Online Appendix C.

2.3. *Power Analysis*

The Market Survey Incentive and Betel Nut Income Experiments were run in 2006 and 2007. As a result, our experiments pre-date norms of *ex ante* power analyses and pre-analysis plans.[6]

[6] For example, power analyses in economics were first formalised in the journal *Experimental Economics* in 2011 (List *et al.*, 2011) and the AEA RCT Registry did not launch until 2012.

Nonetheless, we conduct an *ex post* power analysis to give readers a sense of the rate at which the Market Survey Incentive Experiment and Betel Nut Income Experiment can detect estimates of their respective parameters of interest. Online Appendix Figure C1 plots the Type 1 and 2 error rates under different values of the effect of incentives on hours worked. Therein, readers can see that Type 1 error rates are around 5% regardless of the value assumed for the parameter of interest under the alternative hypothesis. Readers can also see that Type 2 error rates imply power of 80% when the parameter of interest implies that each Indian rupee of the wage supplement causes vendors to work 0.015 additional (or fewer) hours.[7] To put the magnitude of this effect into perspective, it suggests that doubling vendor hourly earnings with an INR10 wage supplement must increase (or reduce) hours worked by 9 minutes to be detected 80% of the time. This equates to an approximate 2% increase (or decrease) in daily hours worked, which is small relative to many estimates reported in the observational studies that motivate our experimental analysis. For example, the ordinary least squares estimates reported in Camerer *et al.* (1997), Farber (2015), Richards (2020) and He *et al.* (2022), respectively, suggest that a 100% increase in wage changes daily hours worked by −62%, −10%, −11% and −2%. The minimum detectable value of our parameter of interest is also small relative to the instrumental variables estimates in these papers, which respectively suggest that a 100% increase in wage changes daily hours worked by −93%, 59%, −20% and 8%.

The Type 1 and 2 error rates under different values of the effect of overpayments on quit rates are plotted in Online Appendix Figure C2. This figure reveals that Type 2 error rates imply power of 80% when the parameter of interest implies that each Indian rupee of the overpayment causes vendors to increase their probability of quitting for the day by approximately 0.001 percentage points.[8] To relate this minimum detectable value of the parameter of interest to the observational findings that motivated our experimental analysis, we convert the value into the minimum response to a 10% increase in daily income. Assuming vendors earn INR10 per hour and work 8.95 hours per day as control vendors did on the day of the overpayment, a 10% increase in daily income must increase the quit rate by at least 0.01 percentage points. Farber (2005, 2015), Crawford and Meng (2011) and Thakral and Tô (2021) find that increasing daily income by 10% increases the share of workers quitting for the day by 0.18 percentage points, 0.32 percentage points, 0.58 percentage points and 0.40 percentage points, which are all larger than the 0.01 percentage points our experiment is powered to detect.

3. Results

In this section, we present estimates of our parameters of interest and then discuss their implications for theories of labour supply.

3.1. *Market Survey Incentive Experiment*

Before presenting our estimates of the parameter of interest identified by the Market Survey Incentive Experiment, we investigate the reduced form effect of each wage supplement on hours

[7] Online Appendix Figure C1 only considers positive values of the parameter of interest, but power is symmetric, so the results also apply to negative values of the parameter of interest.

[8] To put the magnitude of this effect into units of hours worked, it suggests that to be appropriately powered an INR500 overpayment must cause treatment vendors to quit working approximately 0.45 hours earlier than control vendors. Relative to the hours worked by control vendors, such an effect equates to a 5% reduction in average hours worked.

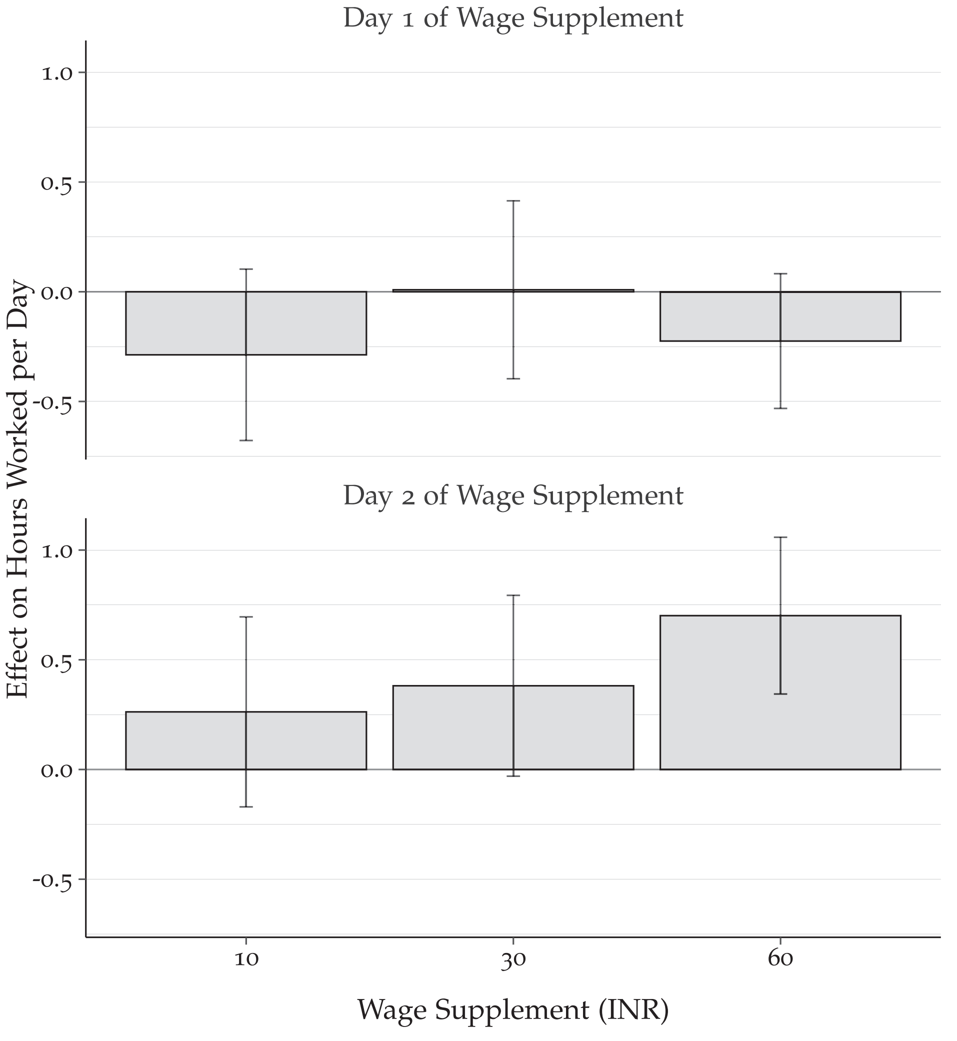

Fig. 1. *Effect of Incentives in Market Survey Incentive Experiment. This figure reports the effect of each wage supplement and day of the wage supplement on hours worked. Effects are obtained by regressing daily hours worked on an indicator for each day and the level of the wage supplement in the Market Survey Incentive Experiment, with a fixed effect for each strata also included. This regression is run on data observed in the Pre-Incentive and Incentive Periods. Brackets indicate 95% confidence intervals, which are constructed with clustered standard errors that are robust to heteroscedasticity and vendor autocorrelation.*

worked in each day of the experiment. The objective of this investigation is to illustrate the labour supply responses that underlie our estimated parameter of interest. Figure 1 presents the estimated effects, with brackets indicating the 95% confidence interval on each estimate. These estimates are obtained by regressing total hours worked per day in the Pre-Incentive and Incentive Periods on an indicator for each level of the wage supplement and day of the Incentive Period,

Table 4. *Effect of Incentives on Hours Worked in Market Survey Incentive Experiment.*

	(1) Hours worked	(2) Hours worked	(3) Hours worked	(4) Hours worked
Wage supplement on Day 1	−0.003	−0.002		
	(0.003)	(0.003)		
Wage supplement on Day 2	0.013	0.015		
	(0.003)	(0.003)		
Wage supplement			0.005	0.007
			(0.003)	(0.003)
Constant	8.443	8.427	8.443	8.427
	(0.087)	(0.026)	(0.087)	(0.026)
Fixed effects	Strata	Vendor	Strata	Vendor
Vendors	250	250	250	250
Observations	820	820	820	820
R^2	0.18	0.74	0.18	0.73

Notes: This table reports the coefficients from a regression of hours worked per day on the magnitude of the wage supplement in the Market Survey Incentive Experiment. This regression is estimated on data from the Pre-Incentive and Incentive Periods. Columns 1 and 2 report the coefficient for each day of the wage supplement, and columns 3 and 4 report the coefficient over both days of the wage supplement. Columns 1 and 3 report the coefficients after adjusting for strata fixed effects. Columns 2 and 4 report the coefficients after adjusting for vendor fixed effects. The reported constant is calculated by taking the weighted average across each stratum or vendor of the predicted outcome when the wage supplement is 0. Clustered standard errors that are robust to heteroscedasticity and vendor autocorrelation are reported in parentheses under each coefficient.

with strata fixed effects also included. Starting at the top panel of Figure 1 we see that, contrary to conventional economic models and consistent with models of reference-dependent labour supply, on the first day of the Incentive Period estimated effects are negative for two of the three levels of the wage supplement and that none of the estimates are statistically distinguishable from a null effect.[9] Moving to the bottom panel, which presents the same estimates for the second day of the Incentive Period, we see a very different response. Each wage supplement has a positive effect on hours worked, the estimates are monotonic in the size of the wage supplement, and, in isolation, the effect of the largest wage supplement rejects the null of hypothesis of no labour supply response.

We next present estimates of the parameter of interest in the Market Survey Incentive Experiment, the effect of incentives on hours worked. Table 4 presents our estimates. Given the heterogeneity in response to the wage supplements over the two days of the Incentive Period we report the effect of incentives on hours worked in each day of the Incentive Period, as well as the effect over both days of the Incentive Period. Columns 1 and 2 report the effect separately for each day, and columns 3 and 4 report the effect over both days of the Incentive Period. The vector of observables used in columns 1 and 3 is a fixed effect for each stratum, while in columns 2 and 4 the vector of observables is a fixed effect for each vendor.

Starting with the effect on each day of the Incentive Period in columns 1 and 2, we see a staggered response to the wage supplement. On the first day of the Incentive Period, each Indian rupee of the wage supplement reduced hours worked by 0.002 to 0.003. This is a small effect, equating to 0.1 to 0.2 fewer minutes of time spent at work, and cannot be statistically

9 The three estimates are not jointly statistically significant, with a p-value of 0.55.

distinguished from a null effect. However, this null effect is estimated precisely enough to rule out small increases and decreases in hours worked. The lower bounds of the 95% confidence intervals are −0.008 to −0.009 and the upper bounds are 0.003 to 0.005. To put these bounds into perspective, we suppose that in the absence of the wage supplement vendors earn INR10 and work 8.4 hours per day, an INR10 wage supplement increases hourly earnings by 100% and causes an effect on work hours that is no smaller than −1.1% and no larger than 0.6%. Therefore, we can reject possibility that the true effect of the wage supplement is equal to the estimates that motivated the experiment, as these estimates ranged from −2 to −62% for a 100% increase in hourly earnings (Camerer *et al.*, 1997; Farber, 2015; Richards, 2020; He *et al.*, 2022).

Moving to the second day of the Incentive Period, we see a different response to the wage supplement. Columns 1 and 2 show that on the second day, each Indian rupee of the wage supplement increased hours worked by 0.013 to 0.015. While still small in magnitude, this second day effect is estimated precisely enough to statistically significantly reject an effect of zero, with both p-values < 0.001. Furthermore, comparing the first and second day effects substantiates the claim that the response was staggered, with p-values < 0.001 on the comparison of the first and second day coefficients obtained with both specification of controls.

Moving to columns 3 and 4 of Table 4, we see that over both days of the Incentive Period each Indian rupee of the wage supplement increased hours worked by 0.005 to 0.007. The statistical significance of the effect depends on the specification of controls used. Relative to a null effect of zero, the estimate of 0.005 in column 3, which controls for the strata in the experiment has a p-value of 0.10 and the estimate of 0.007 in column 4, which has vendor fixed effects as controls, has a p-value of 0.02. Therefore, given the conventional significance threshold of 5%, there is some ambiguity over whether our estimated parameter of interest is non-zero. However, one-sided tests reject the possibility of a negative effect, with p-values of 0.01 and 0.05. As a result, we can also conclude that the aggregate effect of our experimental wage supplement is different from prior research that finds negative incentives effects with observational data (Camerer *et al.*, 1997; Farber, 2015; Richards, 2020; He *et al.*, 2022).

Our estimates also differ from the broader literature on the effect of incentives. To illustrate this point, we calculate the intertemporal elasticity of labour supply implied by our estimates. To do so, we again assume that, in the absence of the wage supplement, vendors earn INR10 and work 8.4 hours per day. Under these assumptions, we find that the −0.002 to −0.003 effect of each Indian rupee on hours worked on the first day of the Incentive Period implies an elasticity of −0.002 to −0.004, and the 0.013 and 0.015 effect of each Indian rupee on the second day of the Incentive Period implies an elasticity of 0.015 to 0.018. The magnitude of these elasticities is small relative to other estimates of the intertemporal elasticity of labour supply in the literature. For example, Chetty *et al.* (2011) reports an elasticity of 0.6 to 0.8 from a meta-analysis of the quasi-experimental literature. This divergence is likely the consequence of features specific to our setting. One such potential feature is that the extensive margin response to incentives was highly inelastic in our setting. Experimental studies on incentives find much larger responses on the extensive margin. For example, Goldberg (2016) finds an elasticity of 0.15 in a Malawian day labour market when extensive margin incentives are varied, and K.-M. Chen *et al.* (2020) find that 91% of the effect of an experimental wage multiplier on Uber is on the extensive margin of labour supply. Yet the vendors in our setting worked on nearly every day we collected data in the market, regardless of the wage supplement offered. Ultimately, we discourage readers from extrapolating the specific magnitude of our implied elasticities to other settings and encourage readers to instead focus on the relationship between the direction of our estimated effects and the comparative statics of reference-dependent models, which we discuss in Section 3.3.

Additional empirical analyses help isolate the margins of labour supply that determine the incentive effects on hours worked. Online Appendix Table C3 reports estimates of wage supplement's effect on the decision to work, when the day is started, break time, and when the day is ended. Therein, we find that on the first day of the Incentive Period, the incentive effect on the time vendors ended their day is negative and nearly statistically significant at the 5% threshold, with less precisely estimated evidence of vendors also showing up to work later. Moving to the second day of the Incentive Period, the effect of the incentive on both these margins flip. Each Indian rupee of the wage supplement causes vendors to show up earlier and exit later the second day of the Incentive Period, with both effects statistically significant. Regardless of day, Online Appendix Table C3 also shows that incentives do not statistically significantly alter the probability of a vendor working or the time they take on breaks.

In the final column of Online Appendix Table C3, we also consider the effect of our wage supplement on hours worked in the Post-Incentive Period. Therein, we report that each Indian rupee of the wage supplement caused vendors to work 0.005 fewer hours per day in the Post-Incentive Period and this effect is not statistically distinguishable from a null effect. We also consider the effect of accumulated incentives on the vendor decision to quit working for the day. Online Appendix Table C4 reports that the day of the wage supplement also influenced the effect on the probability of quitting for the day. On the first day of the supplement, treatment vendors were more likely than control vendors to quit working for the day. This flips on the second day, with quitting marginally discouraged by the wage supplements. These findings are broadly consistent with the findings in Table 4 and Online Appendix Table C3.

One potential explanation for our estimated incentive effects is that, on the first day of the Incentive Period, vendors did not view the wage supplements as credible because they were not paid their earnings until the end of the first day. To examine this possibility, we exploit a quirk in the administration of the first session of the Market Survey Incentive Experiment. In this session, treatment vendors in each stratum were recruited in two waves. If the credibility of the incentives in the Market Survey Incentive Experiment was a determinant of our findings, then we should observe a different response across the two waves because by the time the second wave started, treatment vendors in the first wave had received their payments for each market survey that they completed. Online Appendix Table C5 shows that there was no such dynamic between the two waves. As a result, we conclude that our estimated incentive effects reflect vendor preferences and constraints, not time at which the wage supplement was paid.

Another potential explanation is that, because treatment vendors exited the market earlier on the first day of the wage supplement, they are better rested for a long day of work on the second day of the wage supplement. We test for this explanation by replicating Table 4 with a control for the hours worked by a vendor on the prior day. Online Appendix Table C6 shows that once lagged hours worked are controlled for, the same staggered response persists: Day 1 of the wage supplement has an effect on hours worked that is statistically indistinguishable from a null, and on Day 2 the supplement has a highly statistically significant positive effect on hours worked.

A final potential explanation our data can address is the possibility that vendors respond to the wage supplements on the second day, and not the first, because of the additional time they had to plan their response.[10] A quirk in the administration across the two sessions of the Market

[10] Of course, other potential explanations exist that our data cannot address. For example, vendors may have needed to learn from their own experience with an incentive regime to determine their best response. Explanations like this could

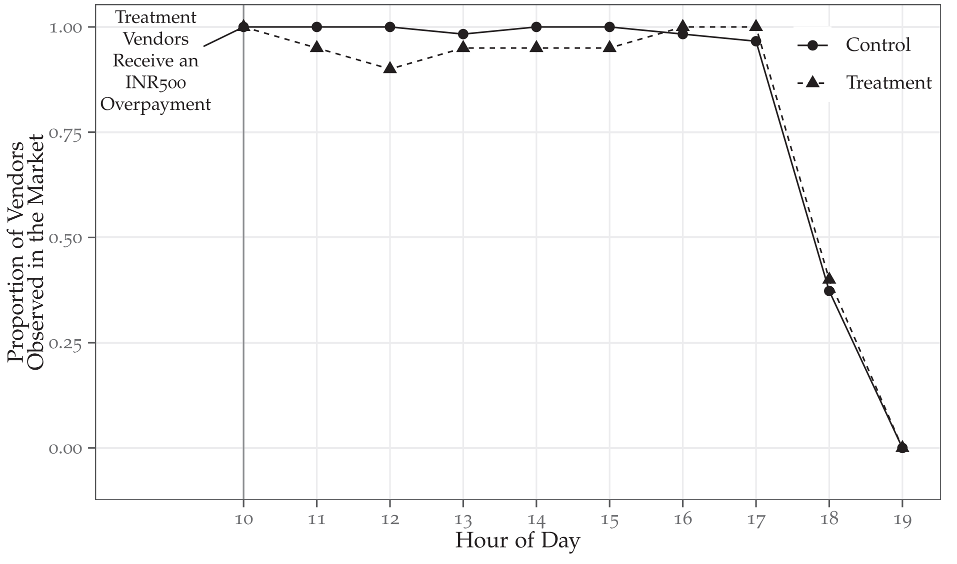

Fig. 2. *Share of Vendors in Market After Overpayment Administered in Betel Nut Income Experiment. This figure plots the proportion of vendors in the treatment and control group who were observed in the market after the overpayment was administered.*

Survey Incentive Experiment allows us to evaluate this explanation. In the first session, treatment vendors were informed about the wage supplements the evening before the supplements could be earned. In the second session, treatment vendors were informed two calendar days before the supplements could be earned. Online Appendix Table C7 shows that this differential timing did not influence staggered response to the wage supplements.

3.2. *Betel Nut Income Experiment*

Figure 2 plots the share of vendors in treatment and control who were active in the market on the day of the overpayment. When the overpayments were distributed around 10 a.m., all treatment and control vendors were in the market. Shortly after the overpayment, 5–10% of the treatment vendors exited the market, only to return around 4 p.m. Then at 5 p.m. control vendors started to quit working for the day at a marginally faster rate than treatment vendors. By 7 p.m. all the treatment and control vendors quit working for the day.

Table 5 uses the data plotted in Figure 2 to estimate our second parameter of interest: the effect of accumulated income on the decision to quit working for the day. Columns 1 and 2 report the coefficients from a regression of the binary quitting decision on an indicator for treatment. Therein, we see that the baseline quitting rate was approximately 9.7% and, depending on whether start time is controlled for, treatment had a −0.04 to 0.02 percentage point effect on the quit rate. These estimates are small in magnitude, and are statistically indistinguishable from the null effect predicted by a conventional economic model of labour supply. Columns 3 and 4 estimate (6) by

be evaluated best in settings where familiar incentive regimes can be exogenously varied and compared to unfamiliar incentive regimes, such as a variant of our wage supplement.

Table 5. *Effect of Income on Decision to Quit Working in Betel Nut Income Experiment.*

	(1) Quit (p.p.)	(2) Quit (p.p.)	(3) Quit (p.p.)	(4) Quit (p.p.)
Treatment	−0.04092	0.01957		
	(0.11973)	(0.12182)		
Overpayment			−0.00008	0.00004
			(0.00024)	(0.00024)
Constant	9.65630	9.64094	9.65630	9.64094
	(0.06271)	(0.06099)	(0.06271)	(0.06099)
Fixed effects		Start Time		Start Time
Vendors	79	79	79	79
Observations	819	819	819	819

Notes: This table reports the coefficients from a regression of the binary decision to quit working for the day on the overpayment in the Betel Nut Income Experiment. Columns 1 and 2 report the effect of assignment to receive the overpayment, and columns 3 and 4 report the effect of the magnitude of the overpayment. Columns 1 and 3 feature no controls, while columns 2 and 4 include a fixed effect for the time of day when the vendor started working. The reported constant in columns 2 and 4 are calculated by taking the weighted average across each start time. The R^2 for the models estimated in columns 1 and 3 is 0.00004 percent and for the models estimates in columns 2 and 4 is 0.00160 percent. See Online Appendix Table C11 for estimates from the same regression that also control for accumulated hours on break. Clustered standard errors that are robust to heteroscedasticity and vendor autocorrelation are reported in parentheses under each coefficient.

regressing the decision to quit working for the day on a variable that measures the magnitude of the overpayment. Therein, we see that each Indian rupee of the overpayment had an exceedingly small effect, in percentage points, on the baseline quitting rate. Much like the estimates in columns 1 and 2, the estimates in columns 3 and 4 are also statistically indistinguishable from a null effect.

A basic concern with the estimates in Table 5 is that treatment vendors may remain in the market because the overpayment causes them to believe that continuing to work will net them an additional overpayment. To evaluate an outcome that is less likely to be confounded by such beliefs, we collected auxiliary data on the effort exerted by vendors in their haggling over the price of betel nuts. Table 6 reports the effect of the overpayment on these outcomes. Columns 1, 2 and 3 show that, regardless of whether vendor fixed effects are included, average initial offers for a pack of betel nuts are INR18. After the overpayment, treatment vendor offers increase by an average of INR2 to INR3. Moving to columns 4, 5 and 6, we find similar effects of the overpayment on the final transaction price. While, at the conventional level of 5% significance, these effects are not statistically significantly different from a null effect, such a null is consistent with the effects reported on quitting and this consistency suggests that dynamic confounds did not bias our estimates in Table 5.[11]

3.3. *Implications*

The data generated by our two experiments cannot be fully explained by models of reference dependence or neoclassical alternatives. While a conventional model of labour supply predicts vendors will increase their hours worked on both days of the Market Survey Incentive Experiment,

[11] In Online Appendix Table C8, we show that auxiliary data observed the day after the overpayment is consistent with the results in Table 6.

Table 6. *Effect of Income on Betel Nut Haggling in Betel Nut Income Experiment.*

	(1) Initial price	(2) Initial price	(3) Initial price	(4) Final price	(5) Final price	(6) Final price
Treatment	−1.46			−1.94		
	(1.10)			(1.04)		
Day of overpayment	0.83	−0.16	0.12	0.64	0.19	0.37
	(1.12)	(1.16)	(0.93)	(1.04)	(1.12)	(0.90)
Treatment X day of overpayment	1.60	2.60	1.98	1.69	2.25	1.73
	(1.52)	(1.52)	(1.18)	(1.42)	(1.47)	(1.15)
Constant	18.27	18.05	18.04	17.83	17.19	17.21
	(0.68)	(0.53)	(0.41)	(0.66)	(0.50)	(0.41)
Fixed effects	Betel type	Betel type, Vendor	Betel type, Vendor, Confederate	Betel type	Betel type, Vendor	Betel type, Vendor, Confederate
Vendors	55	55	55	55	55	55
Observations	151	151	151	151	151	151
R^2	0.11	0.46	0.66	0.10	0.45	0.63

Notes: This table reports effects on the initial and final prices of betel nut transactions. Three effects are reported: the effect of assignment to treatment, the effect of the day on which treatment was administered, and the effect of receiving the treatment on the day the treatment was administered. Columns 1, 2 and 3 report these effects on initial price offered by a vendor, and columns 4, 5 and 6 report these effects on the final transaction price. Columns 1 through 6 control for the type of betel nut that was haggled over. Columns 2, 3, 5 and 6 also control for vendor. Columns 3 and 6 also control for the confederate who elicited the initial and final prices. Estimated are obtained with data observed on a day prior to the overpayment and the day of the overpayment. Estimates of the effect on the day after the overpayment can be found in Online Appendix Table C8, and estimates of the effect of the overpayment on logged initial and final prices can be found in Online Appendix Table C9. Clustered standard errors that are robust to heteroscedasticity and vendor autocorrelation are reported in parentheses under each coefficient.

$\alpha > 0$, we cannot reject the null hypothesis of no response to the wage supplements on the first day of the Market Survey Incentive Experiment, and we cannot reject the null hypothesis of no response over both days of the experiment at the 5% level unless vendor fixed effects are used. Additional analyses fail to find evidence that the null effect on the first day of the Incentive Period can be attributed to a conventional model that also features certain types of learning, either via observation of the behaviour of nearby vendors who start receiving the wage supplement a day earlier, or via a longer time horizon to respond to the first day of the wage supplement. Similarly, theories of reference-dependent preferences predict that a shock to income will cause vendors to exit the market early, $\beta > 0$, but our data cannot reject the null hypothesis of no response in quit rates and elicited effort.

Allowing heterogeneity in reference dependence is an attractive approach to explaining our findings. The first potential dimension of heterogeneity is across vendors and is described by the experience vendors have with the wage supplements in the Market Survey Incentive Experiment. Such experience effects have been found in studies on reference-dependent preferences more broadly. For example, Engelmann and Hollard (2010) find that exchange asymmetries predicted by reference-dependent preferences disappear after a 15 minute period in which undergraduates are forced to make trades in a lab experiment (see also List, 2003, 2004, 2011, for evidence on experience gained over longer time horizons). This type of heterogeneity can be incorporated into (1) by allowing the parameters of interest to vary by vendor,

$$Y_{js} = \alpha_j I_{js} + U_{js}, \tag{1'}$$

where a_j covaries positively with a measure of vendor experience. This positive covariance between α_j in ($1'$) and experience can explain why, on the first day of the Market Survey Incentive Experiment when vendors have no experience with the wage supplements, they do not work a longer day and why, on the second day once experience has been gained, they increase their hours worked. This positive covariance can also explain why Farber (2015) finds labour supply elasticities increase with experience, and why recent studies on incentive effects with rideshare drivers do not find a staggered response (e.g., M.K. Chen *et al.*, 2019; K.-M. Chen *et al.*, 2020). The drivers in these samples already have a tremendous amount of experience with the incentive regime that is varied. Comparing the effect of an incentive that vendors are used to receiving, such higher earnings due to a demand shock, to an unusual incentive like the market survey could provide a test of this explanation.

The other dimension of heterogeneity is intertemporal. A recent observational study on New York City taxicab drivers, Thakral and Tô (2021), proposes that this type of heterogeneity can be described by the following dynamic process. Around the start of a day, workers respond to surprises by updating their reference level of income. But as a day continues, it is increasingly harder for workers to adjust their reference levels of income. Incorporating this type of heterogeneity into (2) can be accomplished by allowing the parameters of interest to vary by time,

$$D_{jst} = \sum_{k=1}^{t} \beta_k M_{jsk} + V_{jst}, \tag{2'}$$

where β_t grows in magnitude as time in the day, t, grows. This formulation offers an explanation for the results of the Betel Nut Income Experiment. Because we administer the overpayments early in the day, under ($2'$), treatment vendors are able adapt their reference level of earnings and stay at work. Overpayments administered later in the day would then provide a test of the dynamic formation of reference levels of earnings.[12]

4. Conclusion

How do labour markets respond to transitory changes in demand? Conventional economic models predict that supply responds positively to the incentives created by transitory shifts that increase demand, and negatively to the disincentives generated by transitory shifts that reduce demand (Lucas and Rapping, 1969). Underlying this prediction is an assumption about the nature of income effects proposed by Friedman (1957). This assumption holds that, absent credit constraints, income effects are determined by the extent to which they alter lifetime earnings. As a result, when transitory shifts in demand create incentives that do not meaningfully alter expected lifetime earnings, the quantity of labour supplied should increase. Nearly three decades ago, Camerer *et al.* (1997) argued that this prediction may not hold if workers have reference-dependent preferences. Under such preferences, small incentives can create big income effects that inhibit the supply of labour when demand peaks. While a large collection of studies report evidence of this prediction, little of it is experimental.

[12] However, one concern with overpayments at different points in the day is that they may bias results in the direction of being consistent with theories of adaptive reference dependence. For example, vendors may be compelled to leave the market to secure the overpayment they received. If the overpayment is received later in the day, then the process of securing their overpayment could lead to data that appears consistent with dynamically formed reference levels of earnings, but instead reflects the fixed cost of leaving the market. Digital payments offer an attractive way of avoiding this type of concern.

We experimentally study the effects of transitory incentives and income on the labour supply of 329 vendors in an open-air market in India. The effects we estimate fail to conform with the predictions of either model of labour supply. In the Market Survey Incentive Experiment, we find a staggered response to high-stakes incentives. On the first day of the wage supplement, consistent with reference dependence, treatment vendors reduce their hours worked. While this effect is statistically indistinguishable from a null effect, such an effect can be explained by reference dependence, but not a conventional model of labour supply. By the second day of the wage supplement, though, treatment vendors behave according to a conventional labour supply model. The higher the level of an incentive offered on this day the more hours that vendors work, a trend that is highly statistically significant. In the Betel Nut Income Experiment, we find a null effect on labour market participation in response to a high-stakes overpayment received by treatment vendors. We also fail to find evidence that this null response is driven by treatment vendors remaining in the market in hopes of earning an additional overpayment. While this null response is consistent with the predictions of a conventional labour supply model, theories of reference-dependent preferences predict otherwise.

These findings can be explained by two variants of reference-dependent labour supply. The first has vendors assign a large weight to their reference-dependent preferences when first responding to an incentive arrangement. Once experience with an incentive arrangement is gained, though, vendors diminish the weight placed on reference dependence and behave according to their conventional economic desires. The second variant allows for intraday updating of reference levels of earnings (Thakral and Tô, 2021). Under this theory, our overpayment did not influence treatment vendors because it was conducted in the morning, which allowed vendors to update their reference level of earnings. A useful exercise for future research is to experimentally introduce the variation suggested by these competing explanations into the experimental designs developed in this study.

Copenhagen Business School, Denmark
Johns Hopkins University, USA
University of California, San Diego, USA
University of Chicago & NBER, USA

Additional Supporting Information may be found in the online version of this article:

Online Appendix
Replication Package

References

Abeler, J., Falk, A., Goette, L. and Huffman, D. (2011). 'Reference points and effort provision', *American Economic Review*, vol. 101(2), pp. 470–92.

Agarwal, S., Diao, M., Pan, J. and Sing, T.F. (2015). 'Are Singaporean cabdrivers target earners?', Working paper, SSRN.

Andersen, S., Bulte, E., Gneezy, U. and List, J.A. (2008). 'Do women supply more public goods than men? Preliminary experimental evidence from matrilineal and patriarchal societies', *American Economic Review*, vol. 98(2), pp. 376–81.

Andersen, S., Ertaç, S., Gneezy, U., Hoffman, M. and List, J.A. (2011). 'Stakes matter in ultimatum games', *American Economic Review*, vol. 101(7), pp. 3427–39.

Andersen, S., Ertaç, S., Gneezy, U., List, J.A. and Maximiano, S. (2013). 'Gender, competitiveness, and socialization at a young age: Evidence from a matrilineal and a patriarchal society', *Review of Economics and Statistics*, vol. 95(4), pp. 1438–43.

Andersen, S., Ertaç, S., Gneezy, U., List, J.A. and Maximiano, S. (2018) 'On the cultural basis of gender differences in negotiation', *Experimental Economics*, vol. 21, pp. 757–78.

Angrist, J.D., Caldwell, S. and Hall, J.V. (2021) 'Uber versus taxi: A driver's eye view', *American Economic Journal: Applied Economics*, vol. 13(2), pp. 272–308.

Avoyan, A., Khubulashvili, R. and Mekerishvili, G. (2020). 'Call it a day: History dependent stopping behavior', Working Paper 8603, CESifo.

Camerer, C., Babcock, L., Loewenstein, G. and Thaler, R. (1997). 'Labor supply of New York City cabdrivers: One day at a time', *Quarterly Journal of Economics*, vol. 112(2), pp. 407–41.

Chen, K.-M., Ding, C., List, J.A. and Mogstad, M. (2020). 'Reservation wages and workers' valuation of job flexibility: Evidence from a natural field experiment', Working Paper 27807, National Bureau of Economic Research.

Chen, M.K., Rossi, P.E., Chevalier, J.A. and Oehlsen, E. (2019). 'The value of flexible work: Evidence from Uber drivers', *Journal of Political Economy*, vol. 127(6), pp. 2735–94.

Chen, M.K. and Sheldon, S. (2015). 'Dynamic pricing in a labor market: Surge pricing and flexible work on the Uber platform', Working paper, UCLA Anderson School of Management.

Chetty, R., Guren, A., Manoli, D. and Weber, A. (2011). 'Are micro and macro labor supply elasticities consistent? A review of evidence on the intensive and extensive margins', *American Economic Review*, vol. 101(3), pp. 471–5.

Chou, Y.K. (2002). 'Testing alternative models of labour supply: Evidence from taxi drivers in Singapore', *Singapore Economic Review*, vol. 47(1), pp. 17–47.

Contandriopoulos, D. and Perroux, M. (2013). 'Fee increases and target income hypothesis: Data from Quebec on physicians' compensation and service volumes', *Healthcare Policy*, vol. 9(2), pp. 30–5.

Crawford, V.P. and Meng, J. (2011). 'New York City cabdrivers' labor supply revisited: Reference-dependent preferences with rational-expectations targets for hours and income', *American Economic Review*, vol. 101(5), pp. 1912–32.

DellaVigna, S. (2009). 'Psychology and economics: Evidence from the field', *Journal of Economic Literature*, vol. 47(2), pp. 315–72.

DellaVigna, S., Linder, A., Reizer, B. and Schmider, J.F. (2017). 'Reference-dependent job search: Evidence from Hungary', *Quarterly Journal of Economics*, vol. 132(4), pp. 1969–2018.

Dodini, S. (2022). 'Making reference-dependent preferences: Evidence from door-to-door sales', Working paper, Norwegian School of Economics.

Doran, K. (2014). 'Are long-term wage elasticities of labor supply more negative than short-term ones', *Economics Letters*, vol. 122(2), pp. 208–10.

Dupas, P., Robinson, J. and Saavedra, S. (2020). 'The daily grind: Cash needs and labor supply', *Journal of Economic Behavior and Organization*, vol. 177, pp. 399–414.

Engelmann, D. and Hollard, G. (2010). 'Reconsidering the effect of market experience on the "endowment effect"', *Econometrica*, vol. 78(6), pp. 2005–19.

Farber, H.S. (2005). 'Is tomorrow another day? The labor supply of New York City cabdrivers', *Journal of Political Economy*, vol. 113(1), pp. 46–82.

Farber, H.S. (2008). 'Reference-dependent preferences and labor supply: The case of New York City taxi drivers', *American Economic Review*, vol. 98(3), pp. 1069–82.

Farber, H.S. (2015). 'Why you can't find a taxi in the rain and other labor supply lessons from cab drivers', *Quarterly Journal of Economics*, vol. 130(4), pp. 1975–2026.

Fehr, E. and Goette, L. (2007). 'Do workers work more if wages are high? Evidence from a randomized field experiment', *American Economic Review*, vol. 97(1), pp. 298–317.

Ferraro, P.J. and Tracy, J.D. (2022). 'A reassessment of the potential for loss-framed incentive contracts to increase productivity: A meta-analysis and a real-effort experiment', *Experimental Economics*, vol. 25, pp. 1441–66.

Fisher, J. (2022). 'The cost of labor supply biases', Working paper, London School of Economics.

Friedman, M. (1957). *A Theory of the Consumption Function*, Princeton, NJ: Princeton University Press.

Fryer, R.G., Levitt, S.D., List, J.A. and Sadoff, S. (2022). 'Enhancing the efficacy of teacher incentives through loss aversion: A field experiment', *American Economic Journal: Economic Policy*, vol. 14(4), pp. 269–99.

Gill, D. and Prowse, V. (2012). 'A structural analysis of disappointment aversion in a real effort competition', *American Economic Review*, vol. 102(1), pp. 469–503.

Giné, X., Martinez-Bravo, M. and Vidal-Fernández, M. (2017). 'Are labor supply decisions consistent with neoclassical preferences? Evidence from Indian boat owners', *Journal of Economic Behavior and Organization*, vol. 142, pp. 331–47.

Gneezy, U., Leonard, K.L. and List, J.A. (2009). 'Gender differences in competition: Evidence from a matrilineal and a patriarchal society', *Econometrica*, vol. 77(5), pp. 1637–64.

Goette, L., Huffman, D. and Fehr, E. (2004). 'Loss aversion and labor supply', *Journal of the European Economic Association*, vol. 2(2–3), pp. 216–28.

Goldberg, J. (2016). 'Kwacha gonna do? Experimental evidence about labor supply in rural Malawi', *American Economic Journal: Applied Economics*, vol. 8(1), pp. 129–49.

Hammarlund, C. (2018). 'A trip to reach the target? The labor supply of Swedish Baltic cod fishermen', *Journal of Behavioral and Experimental Economics*, vol. 75, pp. 1–11.
He, S., Qiu, L. and Cheng, X. (2022). 'Surge pricing and short-term wage elasticity of labor supply in real-time ridesharing markets', *MIS Quarterly*, vol. 46(1), pp. 193–228.
Hoffman, M., Gneezy, U. and List, J.A. (2011). 'Nurture affects gender differences in spatial abilities', *Proceedings of the National Academy of Arts and Sciences*, vol. 108(36), pp. 14786–8.
Hossain, T. and List, J.A. (2012). 'The behavioralist visits the factory: Increasing productivity using simple framing manipulations', *Management Science*, vol. 58(12), pp. 21–51.
Köszegi, B. and Rabin, M. (2006). 'A model of reference-dependent preferences', *Quarterly Journal of Economics*, vol. 121(4), pp. 1133–65.
Levitt, S.D., List, J.A., Neckermann, S. and Sadoff, S. (2016). 'The behavioralist goes to school: Leveraging behavioral economics to improve educational performance', *American Economic Journal: Economic Policy*, vol. 8(4), pp. 183–219.
List, J.A. (2003). 'Does market experience eliminate market anomalies?', *Quarterly Journal of Economics*, vol. 118(1), pp. 41–72.
List, J.A. (2004). 'Neoclassical theory versus prospect theory: Evidence from the marketplace', *Econometrica*, vol. 72(2), pp. 615–25.
List, J.A. (2011). 'Does market experience eliminate market anomalies? The case of exogenous market experience', *American Economic Review*, vol. 101(3), pp. 313–17.
List, J.A., Sadoff, S. and Wagner, M. (2011). 'So you want to run an experiment, now what? Some simple rules of thumb for optimal experimental design', *Experimental Economics*, vol. 14(4), pp. 439–57.
Lucas, R.E., Jr. and Rapping, L.A. (1969). 'Real wages, employment, and inflation', *Journal of Political Economy*, vol. 77(5), pp. 721–54.
MacDonald, D. and Mellizo, P. (2017). 'Reference dependent preferences and labor supply in historical perspective', *Journal of Behavioral and Experimental Economics*, vol. 69(C), pp. 117–24.
Martin, V. (2017). 'When to quit: Narrow bracketing and reference dependence in taxi drivers', *Journal of Economic Behavior and Organization*, vol. 144, pp. 166–87.
Morgul, E.F. and Ozbay, K. (2015). 'Revisiting labor supply of New York City taxi drivers: Empirical evidence from large-scale taxi data', Paper presented at Transportation Research Board 94th Annual Meeting, 1–15 January.
Nguyen, Q. and Leung, P. (2013). 'Revenue targeting in fisheries', *Environment and Development Economics*, vol. 18(5), pp. 559–75.
O'Donoghue, T. and Sprenger, C. (2018). 'Reference-dependent preferences', in (B.D. Bernheim, S. DellaVigna and D. Laibson, eds.), *Handbook of Behavioral Economics: Applications and Foundations*, vol. 1, pp. 1–77. Amsterdam: Elsevier.
Oettinger, G.S. (1999). 'An empirical analysis of the daily labor supply of stadium vendors', *Journal of Political Economy*, vol. 107(2), pp. 360–92.
Pierce, L., Rees-Jones, A. and Blank, C. (2020). 'The negative consequences of loss-framed performance incentives', Working Paper 26619, National Bureau of Economic Research.
Ran, T., Keithly, W.R. and Yue, C. (2014). 'Reference-dependent preferences in Gulf of Mexico shrimpers' fishing effort decision', *Journal of Agricultural and Resource Economics*, vol. 39(1), pp. 19–33.
Richards, T.J. (2020). 'Income targeting and farm labor supply', *American Journal of Agricultural Economics*, vol. 102(2), pp. 419–38.
Rizzo, J.A. and Blumenthal, D. (1996). 'Is the target income hypothesis an economic heresy?', *Medical Care Research and Review*, vol. 53(3), pp. 243–66.
Rizzo, J.A. and Zeckhauser, J.A. (2003). 'Reference incomes, loss aversion, and physician behavior', *Review of Economics and Statistics*, vol. 85(4), pp. 909–22.
Rizzo, J.A. and Zeckhauser, J.A. (2007). 'Pushing incomes to reference points: Why do male doctors earn more?', *Journal of Economic Behavior & Organization*, vol. 63(3), pp. 514–36.
Saia, A. (2022). 'Trouble underground: Demand shocks and the labor supply behavior of New York City taxi drivers', *Italian Economic Journal*, vol. 8(1), pp. 1–27.
Schmidt, M.A. (2018). 'The daily labor supply response to worker-specific earnings shocks', Working paper, University of Toronto.
Sheldon, M. (2016). 'Income targeting and the ridesharing market', Working paper, Uber Technologies.
Stafford, T.M. (2015). 'What do fishermen tell us that taxi drivers do not? An empirical investigation of labor supply', *Journal of Labor Economics*, vol. 33(3), pp. 683–710.
Thakral, N. and Tô, L.T. (2021). 'Daily labor supply and adaptive reference points', *American Economic Review*, vol. 111(8), pp. 2417–43.

The Economic Journal, 1–24 https://doi.org/10.1093/ej/ueaf051 © The Author(s) 2025. Published by Oxford University Press on behalf of Royal Economic Society. All rights reserved. For commercial re-use, please contact reprints@oup.com for reprints and translation rights for reprints. All other permissions can be obtained through our RightsLink service via the Permissions link on the article page on our site-for further information please contact journals.permissions@oup.com.
Advance Access Publication Date: 20 June 2025