Toward an Understanding of Reference-Dependent Labor Supply: Theory and Evidence from a Field Experiment

Steffen Andersen* Alec Brandon Uri Gneezy John A. List

October 2014

Abstract: Perhaps the most powerful form of framing arises through reference dependence, wherein choices are made recognizing the starting point or a goal. In labor economics, for example, a form of reference dependence, income targeting, has been argued to represent a serious challenge to traditional economic models. We design a field experiment linked tightly to three popular economic models of labor supply—two behavioral variants and one simple neoclassical model—to deepen our understanding of the positive implications of our major theories. Consistent with neoclassical theory and reference-dependent preferences with endogenous reference points, workers (vendors in open air markets) supply more hours when presented with an expected transitory increase in hourly wages. In contrast with the prediction of behavioral models, however, when vendors earn an unexpected windfall early in the day, their labor supply does not respond. A key feature of our market in terms of parsing the theories is that vendors do not post prices rather they haggle with customers. In this way, our data also speak to the possibility of reference-dependent preferences over other dimensions. Our investigation again yields results that are in line with neoclassical theory, as bargaining patterns are unaffected by the unexpected windfall.

* Stefano DellaVigna, Casey Mulligan, Zongjin Qian, and George Wu provided helpful comments and suggestions.
Andersen: Department of Economics, Copenhagen Business School, Porcelænshaven 16A, 1, DK-2000 Frederiksberg, Denmark (e-mail: sa.eco@cbs.dk);); Brandon: Department of Economics, University of Chicago, 1126 E. 59th St., Chicago, IL 60637 (e-mail: alec@uchicago.edu); Gneezy: Rady School of Management, University of California, San Diego, 9500 Gilman Drive, La Jolla, CA 92093 (e-mail: ugneezy@ucsd.edu); List: Department of Economics, University of Chicago, 1126 E. 59th St., Chicago, IL 60637, and NBER (e-mail: jlist@uchicago.edu).

Several fundamental postulates give rise to the predictions from neoclassical models. One of the most basic tenets—the basic independence assumption—has recently come under heavy scrutiny. Results across several experimental settings suggest that preferences are not independent of current entitlements (see, e.g., Knetsch, 1989; Kahneman et al., 1991; Englemann and Hollard, 2010). The most accepted theory explaining such preferences invokes psychological effects, and is broadly termed "prospect theory" (Kahneman and Tversky, 1979). Recent studies have shown the power of using such behavioral insights—from motivating workers in manufacturing plants to convincing teachers and students to exert more effort, incentive schemes have been used to leverage such preferences (see Hossain and List, 2012; Fryer et al., 2012; Levitt et al., 2012).

In a normative sense, such preferences call into question commonly held interpretations of indifference curves and cripple applied welfare analysis. For instance, cost/benefit analysis and damage resolution would need a new architecture without the basic independence assumption. From a positive perspective, a large disparity between Hicksian equivalent surplus and Hicksian compensating surplus essentially renders the invariance result of Coase invalid and calls into question how we calculate gains and losses of public policies, ranging from computing the correct gains to trade and who is affected by welfare and transfer programs.

A prime example of the interplay between prospect theory and positive economics is how to grapple with some of the key questions we face in labor economics. For example, how does labor supply respond to changes in taxes, welfare, and transfer programs? How important is inter-temporal substitution? Traditionally, when computing the inter-temporal elasticity of substitution (IES), scholars make use of synthetic cohorts or panels from large survey datasets and employ various regression models. A typical finding in this literature is an estimate of the IES between 0 and 0.4 suggesting a small labor supply response to changing the transitory returns to supplying labor (Ghez and Becker (1975), MaCurdy (1981), and Altonji (1986)). A potential concern with these methods is that the typical worker is not free to vary labor supply and hence the estimates are attenuated from the true, structural and policy-relevant, parameter of interest (Chetty et al. (2011)).[1]

[1] Parsing anticipated and unanticipated changes in wages is another challenge in this literature. Mulligan (1995) summarizes labor supply responses to events that the econometrician can confidently call anticipated, like World

As a result, researchers have turned to novel samples, using high-frequency labor supply data on workers free to choose their daily hours of work. But instead of clarifying the magnitude of the IES these studies have suggested that under certain circumstances the IES is actually negative. In the first study using high-frequency labor supply data, Camerer et al. (1997) finds that New York City taxi cab drivers work fewer hours when their hourly wage rate is higher, with estimates of the IES ranging from -0.2 to -0.5. Cab drivers make an attractive sample because their wages fluctuate from day-to-day, but within a day, their wages are relatively stable. This result is hard to square with neoclassical theory. To explain the observed behavior Camerer et al. (1997) informally propose a model of labor supply that borrows from prospect theory, where drivers set a daily income target beyond which their marginal utility of income decreases substantially.

Clearly the implications of income targeting are far reaching, but the empirical evidence presented in the literature thus far has been mixed. For example, Oettinger (1999), Farber (2005) and Goldberg (2014) find that labor supply responds positively to increases in wages for stadium vendors, a richer sample of New York City taxi cab drivers, and laborers in Malawai, while Chou (2002), Goette et al. (2004), and Fehr and Goette (2007) report evidence consistent with reference-dependent preferences for taxi cab drivers in Singapore and bike messengers.

Koszegi and Rabin (2006), KR henceforth, develop a novel theory of reference-dependent preferences that helps to synthesize the literature by proposing an explicit mechanism for the generation of reference points in the labor supply decision.[2] In the KR model, each worker wakes up each morning and endogenously generates a reference point via rational expectations of their daily wages and hours worked. This approach is appealing compared to a model of income targeting where reference points cannot change in response to expected shifts in supply or demand. Because reference points in KR are generated via expectations, the theory establishes a

War II, the Trans-Alaskan pipeline construction, or the Exxon Valdez cleanup. Mulligan (1995) shows that the IES is well above unity for such samples.

[2] The model of reference-dependent preferences developed in KR is quite general. It also reconciles disparate findings on puzzles like the endowment effect. Empirical research has found support for the role of reference points in explaining the endowment effect (e.g., List (2003; 2004), Knetsch and Wong (2009), Engelmann and Hollard (2010)) and against it (e.g., Heffetz and List (Forthcoming)).

clear set of predictions for labor supply in response to expected and unexpected changes in income.[3]

In the first empirical test of KR in the context of labor supply decisions, Crawford and Meng (2011) reanalyze the naturally occurring data from Farber (2005, 2008) to test a model where cab drivers target both daily income and hours worked via rational expectations. They find that KR explains labor supply behavior well, whereas a neoclassical model has difficulties.[4] This represented an important contribution, in that it showcases that in a naturally-occurring economic environment an important behavioral model can capture the essence of the labor supply decision. Since currently the dominant paradigm underlying policy estimates is the neoclassical variant, such a result highlights that with reference dependent preferences the usual elasticity estimates can be severely biased.

We complement this research by extending it in several dimensions. We begin by formalizing three major theories of labor supply—two behavioral and their neoclassical counterpart. Through the lens of the models, one can see that the previous empirical work using naturally-occurring data relies on an identification strategy that requires strong assumptions. For instance, in the cab literature, identification is not possible without assumptions on the determinants of driver expectations. We demonstrate the importance of this assumption in the cab literature by investigating the sensitivity of key parameters in Crawford and Meng to perturbations in expectations and realized earnings. This exercise, proposed in Gentzkow and Shapiro (2014), allows us to explore the relative importance of identifying assumptions for parameter estimates.

[3] In particular, an increase in expected wages on a given day should increase labor supply (consistent with the findings of Oettinger (1999), Fehr and Goette (2007), Goldberg (2014)), and an unexpected increase in wages should decrease labor supplied (consistent with the findings of Camerer et al. (1997), Chou (2002), and Goette et al. (2004)).

[4] There is also increasing evidence of expectation based reference points explaining other types of behavior in naturally occurring data. Pope and Schweitzer (2011) find that professional golfers behave consistent with KR's predictions: They make a higher percentage of putts when a miss would put them in the losses domain when the reference point is a function of the average score on a whole (as opposed to par). In a similar vein, using expectations of a win from gambling markets, Card and Dahl (2011) find that when a football team has an unexpected loss it explains violent behavior well. In the lab, Abeler et al. (2011) test KR in a real effort task. The lab lends itself well to exogenously varying expectations and Abeler et al. (2011) find that reference points, generated via rational expectations, influence effort provision as predicted by KR.

Our findings induce us to design a field experiment that is tightly linked to the theory, where the exogeneity is achieved by design rather than assumption.

We visit a well-functioning market in India and use a field experiment to construct a panel dataset of labor supply for 335 vendors that are free to choose their hours of work. We randomly assign vendors to receive large shocks to their earnings in order to test explicitly the conflicting predictions of a KR, an income targeting, and a neoclassical model of labor supply. Our sample allows us to generate data that speaks directly to the earlier literature that utilized naturally occurring data while enjoying a level of control on the exogenous shocks that allows us to speak clearly to the theory. In doing so, this study provides a framework to isolate the impact of expected and unexpected changes in daily income on labor supply over both the hours and effort dimensions by simply varying the timing of information available to vendors.

We show that expected increases in wages lead to modest, but statistically significant increases in hours worked, with an implied labor supply elasticity between 0.01 and 0.03. We also show that when vendors are shocked with an unexpected windfall their hours of labor supply do not respond: vendors who receive an unexpected windfall work a similar number of hours as control vendors, and their probability of quitting as the day progresses is also indistinguishable from the behavior of control vendors. Furthermore, as prices are determined by haggling, the bargaining data allows us to compare vendor effort between control and treatment and before and after treatment. Across all comparisons, treatment and control vendors are statistically indistinguishable. In sum, the evidence suggests that a simple neoclassical model of labor supply captures the behavior of vendors in our market.

The remainder of the study is organized as follows: Section II reviews the theoretical content of the competing models of labor supply. Section III links the theory to the experimental design. Section IV summarizes the evidence generated by our field experiment, first reviewing the behavioral response to an expected wage increase and then analyzing the response to an unexpected wage increase. Section V concludes with a brief discussion.

II. Theory

To motivate our experimental design we present a discussion of the labor and effort supply decisions for vendors in our sample. The decision problem for vendors each day is to choose their hours of labor supply, h,[5] according to that day's marginal return to labor supplied, w, which we can think of as the vendor's hourly wage. Wages change day-to-day with vendor's learning a day's wage rate when they wake up in the morning. Vendor total income endowment, y, is also affected by unanticipated changes to their endowment, s_0, that are uncorrelated with w, like a large overpayment from a customer or a loss due to a theft, giving us $y = wh + s_0$.[6]

With this framework established, we examine predictions of three models of labor supply. First, we discuss a simple neoclassical labor supply model and characterize the response of labor supply to anticipated changes in wages and to an unanticipated income shock. Next we consider the same comparative statics in a model where workers bracket their utility on a day-by-day basis with gain-loss utility over daily earnings in relation to an exogenously determined reference point—we refer to this model as an income targeting model as it is motivated by the informal model of Camerer et al. (1997). Last, we treat reference points as endogenous distributions of rational expectations for hours and income as presented in KR.

Neoclassical labor supply: Consider the preferences of a neoclassical vendor on a given day, t, whose utility is additively separable into utility $u(\cdot)$ from earnings, income y and $v(\cdot)$ from hours worked h. The neoclassical vendor chooses h according to:

(1) $$V = u(y) - v(h)$$

where $u(\cdot)$ is increasing and concave and $v(\cdot)$ increasing and convex. Totally differentiating the first order condition with respect to w and re-arranging yields the following comparative static:

$$\frac{dh}{dw} = \frac{u''(y)wh + u'(y)}{v''(h) - u''(y)w^2}$$

[5] h can also be thought of as worker effort, e, which we discuss later.

[6] We can also think of s_0 as a vendor's expected discounted lifetime earnings.

which is positive if $u'(y) > -u''(y)wh$. This condition conveys the mathematical content behind the belief that labor supply curves slope upwards in response to transitory changes in wages. In particular, provided vendors have sufficient expected discounted lifetime earnings (s_0), $u(\cdot)$ shouldn't display much curvature at y and the inequality should hold.[7]

The same procedure for a change in savings yields:

$$\frac{dh}{ds_0} = \frac{u''(y)w}{v''(h) - u''(y)w^2}$$

which will be ≈ 0 for $u(\cdot)$ without dramatic curvature locally. Again, this captures the sense in which a vendor that receives a shock to non-wage income will only change their behavior through a wealth effect, which ought to be small over small changes in lifetime income. These comparative static predictions are summarized in the first row of table 1.

Income targeting labor supply: Under the informal framework suggested by Camerer et al. (1997)[8] vendors choose hours worked in order to maximize a functional form along the lines of:

$$V = \begin{cases} \lambda[u(y) - y_r] - v(h), & y \le y_r \\ [u(y) - y_r] - v(h), & y > y_r \end{cases} \tag{2}$$

where y_r is an exogenous daily income utility target and $\lambda > 1$ captures the aversion vendors have towards failing to meet their daily target. Income targeting vendors who are below their daily target earn λw from each hour they work.[9] Then, when a vendor crosses their daily target, the returns to an hour of work drops to w, which captures the sense in which income targeting vendors will work more on days when their wage is low and less when their wage is high ($dh/dw < 0$). In an income-targeting model an income shock impacts hours worked by moving

[7] This condition is not an artifact of our static modeling of the decision problem. A similar restriction emerges in a dynamic model.

[8] We follow the model suggested by Camerer et al. (1997, pp. 425-426): "One possible explanation for the negative hours elasticities is that cabdrivers take a one-day horizon, and set a target (or a target range) and quit when the target is reached. This decision rule can be modeled by a marginal utility of income declining substantially around the average daily income level."

[9] To see this, note that for vendors with $y \le y_r$, their marginal rate of substitution is equal to $w\lambda$, not w:

$$\frac{v'(h)}{u'(y)} = \lambda w.$$

vendors who are in the loss domain into the gain domain, reducing their effective wage rate from λw to w, which induces vendors to work fewer hours ($dh/ds_0 < 0$).

Figure 1: Panel A shows how labor supply interacts with wages in an income-targeting model using simulated data.[10] The curve labeled No Overpayment plots labor supply when there is no shock to earnings. The sharp discontinuity around the daily income target is what induces negative labor supply elasticities over a large region of wage changes. The curve labeled Overpayment shows labor supply for an income-targeting vendor on a day where they earn a large income shock. The income shock induces income-targeting vendors into the gains domain, causing them to work weakly fewer hours at a given wage rate than they would if they had not earned the overpayment.

To explore these comparative statics further, we suppose a vendor wakes up on a given day and observes a wage of w_1. If there is no overpayment, then the vendor will supply labor along the associated supply curve and choose $h(w_1)$ for that day. If the wage is increased to w_2, then the vendor will instead choose $h(w_2)$ on the same supply curve, where $h(w_1) > h(w_2)$ for $w_1 < w_2$ if the wages place the vendor above and below the day's reference income. To see the comparative static on an overpayment, compare the supply of labor within a wage, w_1, when there is no overpayment, which is $h(w_1)$, to the supply of labor when there is an overpayment, which is $\tilde{h}(w_1)$. As the figure shows, provided the vendor is in the income-loss domain, $h(w_1) > \tilde{h}(w_1)$, the negative relationship between hours worked and an overpayment in an income-targeting model arises.

Labor supply in KR: The KR model has reference points generated via rational expectations of daily income and hours worked.[11] Thus, if a change to wages is expected, vendors will adjust their reference points and labor supply will behave according to the neoclassical model's

---

[10] In particular, we assume the following preference parameterization: $u(y) = \frac{y^{1-\sigma}-1}{1-\sigma}$ and $v(h) = \gamma\frac{\epsilon}{1+\epsilon}h^{1+\epsilon/\epsilon}$ with the following choices of parameter values: $\sigma = 0.01, s_0 = 10{,}000, \gamma = 0.005, \epsilon = 0.3$, and $\lambda = 3$ with $y_r = 100$. The income shock is 500. For the region where vendors are above target income if maximizing their loss utility and below the target income if maximizing their gains utility we set hours according to gains utility.

[11] By rational expectations we mean that KR vendors would have a plan for hours worked for each possible value of w. Each plan is made according to a vendor's neoclassical preferences (1). Vendor expectations for hours worked and income are then the cumulative distribution of each of these plans weighted by how likely the plan for each wage will occur.

predictions. However, when unexpected changes occur within-day, reference points are not updated and gain-loss utility augments the vendors maximization problem.

More formally, the KR vendor maximize an augmented version of (2) where vendors now assign probabilistic beliefs to every possible outcome, ex-ante, creating a reference lottery.[12] Under this specification, gain-loss utility arises from a vendor's comparison of a day's outcome in relation to her distribution of weighted reference points, weighing losses more heavily than gains, the intuition being that for a given realized daily income, y, vendors will feel a mixture of gains and losses in relation to all the incomes they were expecting, weighted by the probability with which they were expecting each possible day's income:

$$V = u(y) - v(h) + \int_{-\infty}^{y} [y - Y_r] dF(Y_r) - \lambda \int_{y}^{\infty} [Y_r - y]\, dF(Y_r) + \int_{h}^{\infty} [H_r - h] dG(H_r) - \lambda \int_{-\infty}^{h} [h - H_r] dG(H_r) \tag{3}$$

where $\lambda > 1$ is the weight of losses[13] and $F(\cdot)$ and $G(\cdot)$ are cumulative distribution functions that capture the probabilistic beliefs of daily income (the random variable, Y_r) and labor supplied (H_r), respectively. A KR vendor chooses h to maximize (2), yielding two distinct predictions. Unlike an income targeting model, KR predicts positive labor supply elasticities in response to expected changes in wages, w, $dh/dw > 0$, but with respect to an unpredicted shock to income, s_0, KR predicts a negative elasticity, $dh/ds_0 < 0$.

[12] As KR explains it, the idea of a reference lottery is that if you are drawing a dollar amount from 0 to 100 and you pull a 50, you will feel gains in comparison to the possible draws below 50 and you will feel losses in relation to the possible draws above 50.

[13] We neglect the typical weight of gain-loss utility for the sake of parsimony. Mathematically, we can think of it being normalized to 1 and subsumed in our λ term.

Figure 1: Panel B shows how labor supply evolves for the KR vendor.[14] The curve labeled Baseline plots labor supply for a typical distribution of wages on a day where there is no shock to income.[15] Comparing labor supplied for a given wage rate, for example w_1, with and without an overpayment shows how unexpected changes in earnings can induce vendors to quit early. The curve labeled Overpayment plots hours worked for each possible wage rate, which is below the hours worked when there is no shock to earnings. To see this, compare the labor supplied by the supply function when there is no overpayment, $h(\cdot)$, and the supply function when there is an overpayment, $\tilde{h}(\cdot)$, at w_1. The comparative static is negative because $h(w_1) > \tilde{h}(w_1)$. To show the comparative static on an expected change in daily wages, we provide the curve labeled Expected wage increase, which plots labor supplied when vendors earn an expected hourly wage in addition to their stochastic wage, w_2.[16] The comparison between the Baseline curve and Expected wage increase curve at w_1 and w_2 show how an expected wage increase in KR leads to a positive labor supply response.

Design Suggested by Theory: Figure 2 illustrates how expected and unexpected changes in income interact with reference points. The figure plots leisure (24 hours - hours worked) against income with each unit of leisure that is sold in the labor market yielding income as a function of wages. The figure also features vendor reference points, H_r and Y_r, which can be thought of as the weighted average of each of the vendor's reference points.

To see the importance of expectations in KR, we consider a vendor on day t that chooses labor along $\overline{AB}$ and is informed that they will earn a supplemental wage on days $t+1$ and $t+2$ for every hour they work. This intervention changes w, wages, and shifts $\overline{AB}$ up to $\overline{AC}$ for days $t+1$ and $t+2$. Because information is revealed about w on day t, reference points in KR also shift on days $t+1$ and $t+2$ from H_r and Y_r, up to $H_r{}'$ and $Y_r{}'$, leading to a prediction consistent

[14] We assume the same preferences for the KR vendor as we did the income-targeting vendor with wages distributed according to: $w \sim \Gamma(4, 0.2)$ and the overpayment, $s \sim Bernoulli(0.001)$. Baseline plots labor supplied when the overpayment is 0. Overpayment plots labor supplied when the overpayment is realized, $s = 500$.

[15] The labor supply curve is upward sloping for KR vendors, but does not have to be for unexpected realized wage rates. See KR for a discussion.

[16] In particular, the expected wage increase line plots labor supplied for $\widetilde{w} = w + 10,\ w \sim \Gamma(4, 0.2)$, and the overpayment, $s \sim Bernoulli(0.001)$, where the overpayment does not occur.

with the neoclassical prediction of an upward sloping labor supply for KR. A vendor that generates references points via lagged earnings, on the other hand, would not move from H_r and Y_r to H_r' and Y_r' and labor supply would be negative in response to the perfectly expected increase in wages.

When no information is given prior to a random shock of a day's income, s, reference points will not be able to shift and gain-loss domains will play an important role in both KR and an income targeting model. Consider again the vendor that accrues hours worked and income along line $\overline{AB}$. At the start of the day, this vendor's gain-loss utility is mostly in hours-gains/income-losses so the marginal cost of each hour worked is cheap compared to the marginal utility gained from escaping the feeling of income losses. Line $\overline{AB}$ passes right through the intersection of H_r and Y_r but consider an unexpected deviation from $\overline{AB}$ at point D where a vendor earns a windfall of s pushing them up to point E.[17] As the vendor continues to work after the windfall, they now move along line $\overline{EC}$ where gain-loss utility from income is no longer heavily weighted because the vendor is well beyond Y_r. But as the hours accrue and the vendor gets closer and closer to H_r, the weight of hours-losses starts to mount, motivating the vendor to quit early. This can be gleaned by considering the vendor still on $\overline{AB}$ whose disutility from approaching H_r is counteracted by the utility from increased earnings and approaching Y_r. This yields a negative within-day elasticity with respect to unexpected changes in income.

Expectations and identification in past studies: One of the key results in Crawford and Meng is the relationship between a driver's likelihood of quitting after a given fare, [18] cumulative hours worked, cumulative earnings, and gain/loss status. In particular, on days that a driver's wage is above expectations, cumulative hours worked is an important predictor of quitting for the day.[19] However when drivers observe a wage below their expectations, the opposite is observed. Cumulative income instead of hours worked predicts that a driver will quit. As Crawford and

[17] Note that $\overline{AB}$ and $\overline{EC}$ are parallel. This is because we assume that the windfall earned at point D contains no information that would change w in future hours of the day.

[18] These marginal effects are from a Probit model reported in Columns 2 and 3 in Panel B of Table 2 in Crawford and Meng.

[19] To overcome endogeneity concerns, Crawford and Meng cut the sample by whether a driver's earnings during the first hour of work is above or below expectations for that hour.

Meng explain, "Such a reversal is inconsistent with a neoclassical model, in which the targets are irrelevant; but it is gracefully explained by a reference-dependent model."

Crawford and Meng use lagged average earnings and hours worked for each driver/day-of-week combination. They concede that this proxy is noisy but find that their results are "robust to variations in the specification of the targets." The stability of their results to alternative definitions of expectations is key because it suggests that their reported labor supply dynamics are not driven by their simple approximation of the data generating process for expectations.

As an alternative to their robustness exercise we calculate the (local) sensitivity of key parameters in Crawford and Meng[20] to the assumption that (i) lagged earnings measure expectations and (ii) first hour earnings measure realized earnings. This exercise, proposed by Gentzkow and Shapiro (2014), allows us to compare the sensitivity of parameters with respect to these two assumptions (holding all else equal). This exercise is informative because (ii) is an administrative measure of earnings. That is, parameters ought be sensitive to perturbations to (ii), whereas Crawford and Meng present suggestive evidence that their estimates are robust to perturbations to (i).

Figure 3 presents results from our analysis, with each plotted value representing the expected change in the marginal effect in response to a standard deviation change in expected or realized earnings holding all else equal.[21] The top frame of Figure 3 shows that when drivers are in the gains domain, identification is sensitive to both assumptions with the marginal effect responding slightly more to perturbations in first hour earnings than expectations. This trend reverses in the bottom frame. When drivers are in the losses domain the marginal effect on quitting with respect to changes in cumulative total hours worked and income is more sensitive to changes in expected earnings than realized earnings.

---

[20] We downloaded Crawford and Meng's analysis from their online supplementary materials. We focus on the marginal effects with respect to cumulative total hours worked and cumulative income that are presented in Table 2, Panel B, Columns 3 and 4 in Crawford and Meng. See Gentzkow and Shapiro (2014) for details on conducting this calculation.

[21] Perhaps an easier way to view our investigation here is that Crawford and Meng's estimates are via GMM with two additional moment conditions in their estimates in Table 2: (i) Expected earnings minus lagged average earnings is 0 on average and (ii) First hour earnings minus wage is 0 on average. Then results in Figure 3 show the extent to which parameter estimates are sensitive to model misspecification.

Put differently, consider the following question: To what extent is the assumption that lagged earnings measures driver expectations affecting estimates in Crawford and Meng? Figure 3 shows us that key parameters depend as much or more on assumption (i) as they do on assumption (ii). That is, while the reversal described by Crawford and Meng are robust to specific changes in the specification of expectations, a more formal investigation reveals that key parameters are quite sensitive to misspecification of expected earnings.[22]

Alternative dimensions of labor supply: In Figure 2, one of the key requirements for the predictions of KR is that after the overpayment at D, the slope of $\overline{EC}$ is equal to the slope of $\overline{AB}$. If workers expected their wage to increase after the overpayment then the slope of $\overline{EC}$ would be steeper and the change in the gain-loss utility from the overpayment (the second and third lines of equation (2)) could be counteracted by the neoclassical component of utility in KR (the first line in equation (2)). One way around this difficulty is to design an overpayment that is orthogonal to future earnings that day. Another would be to explore an alternative dimension of labor supply that does not rely on intuiting the worker's expectations about future wages.

We present a model of effort provision for a worker at a single-worker firm who collects a portion of their firm's profits and where the production function for that firm is only a function of bargaining effort.[23] Using a simple theory of the firm suggests that worker effort can be directly observed in our setting. To see this, we start with the typical equilibrium condition for wages at a firm:

(3) $$MPL = w$$

where the marginal product of labor is simply the first derivative of the firm's production function, $f(e)$, where $f(\cdot)$ is strictly increasing and strictly concave in effort, e. As a result, (3) simplifies to:

$$f'(e_i) = w_i = \gamma(p_i - c)$$

---

[22] We only conduct a sensitivity analysis on Table 2 but extending this to Crawford and Meng's structural model would be straightforward. Difficulties in estimating their structural model cause us to focus our efforts on Table 2.

[23] At this point we are introducing a feature of the market that we later study. Some other industries, such as taxi cab drivers, do not have a similar margin for effort that is easily detectable in the same way.

where γ is the fraction of profits the worker takes home in wages from transaction i when the worker exerts effort level e_i. The profits from transaction i are simply the difference in price that results from bargaining and the per-unit cost of the good being sold. Thus, a simple theory of the firm tells us exactly how to identify effort in a market with single-worker firms and no posted prices—just observe the final negotiated price. That is:

(4) $$e_i = f^{-1'}[\gamma(p_i - c)].$$

By observing p_i we can observe effort for a given transaction.

But how do the three models discussed above deal with quality of labor supply (effort), as opposed to quantity (hours worked)? Conceptually, the three models work identically to the description above.[24] A neoclassical model has a worker that wakes up each morning and observes the returns to effort on that day, chooses an aggregate level of effort to provide on that day, and then smooths the effort supplied over all the transactions that occur over the course of the day. The neoclassical model therefore predicts no change in effort supplied in response to an overpayment.[25] The two behavioral models, on the other hand, predict reduced effort in response to an overpayment.[26]

In summary, the three major models of labor supply yield differing sets of predictions in response to expected and unexpected changes in wages and earnings, as summarized in Table 1. Neoclassical theory and KR predicts increasing labor supply to increases in expected wages, while an income-targeting model would predict decreasing. With respect to unexpected changes in earnings, neoclassical theory predicts no change in labor supply or effort provision but KR and a status quo model predict negative supply elasticities for both quantity and quality of labor.

---

[24] KR also suggests that effort and hours worked are interchangeable in their model of labor supply.

[25] A condition that is easy to satisfy provided the overpayment occurs without any bargaining by the customer.

[26] Note that we are modeling the quantity of labor supplied as fixed and the effort provided in those hours as the relevant margin (see also Abeler et al. (2011)). A joint model would be less tractable and proves unnecessary to interpret the data from our field experiment.

III. From Theory to Experimental Design

To test the central predictions of these three theories of labor supply we conduct two field experiments in the Spring of 2006 (Session 1) and the Fall of 2007 (Session 2) in an Indian open air market, the Burra Bazaar in Shillong, North East. The market is the city's main poultry, meat, vegetable, produce, and merchandise market, and has a large number of vendors who are geographically organized by the wares they sell. Vendors in our experiment work as independent contractors. They either own their shop or get paid a percentage of daily profits for working. Only single-vendor shops that sold non-perishables were selected for observation, as vendors that sold perishables would face a different maximization problem than discussed above.

The market opens at sunrise, which was 8:00am during Session 1 and 6:00am during Session 2, and it closes at sunset, around 9:00pm. While these constraints place a ceiling on hours of labor supplied, actual hours are well below the ceiling, giving us a generous margin to examine changes in labor supplied. Also, importantly, vendors in Shillong are not trading their time between multiple labor markets, such as agricultural work, allowing us to confidently call the hours worked we observe, each worker's total hours of labor supplied that day.

We conduct two separate experiments in the market; (i) The Market Survey experiment, where we vary expected wages for vendors and observe labor supply, (ii) The Betel Nut Experiment where we observe labor supply as well as bargaining effort after an unexpected earnings shock. Table 2 summarizes several key variables of vendors in our two experiments. In particular, the data confirm that vendors in both experiments set their own hours, as within-vendor variance is nearly as high as between-vendor variance in both experiments.

In the Market Survey Experiment we highlight the various margins of hours of labor supply for our sample. In particular, vendors are free to participate in supplying labor on any given day, with 98% of the vendor days supplying some labor. Also, vendors showed up to the market between 9:00 am and 10:00 am and left between 6:00 pm and 7:00 pm, with a small percentage taking breaks during the day.[27] These four margins and the substantial within-vendor variance for

---

[27] For the sake of simplicity, we report average arrival time even though the market's opening time varied between the two sessions, as it does not change within vendor variance in arrival time to the market.

each outcome highlight the numerous ways that vendors can take on extra hours of labor each day if offered an expected transitory increase in wages.

In the Betel Nut Experiment we observe both hours of labor supplied as well as bargaining effort. To quantify bargaining effort, we send local confederates to vendors to negotiate for a uniform good sold by all vendors—betel nuts of two varieties. It is critical to our identification strategy that vendors actually vary prices and the second half of Table 2 confirms our claim that vendors do not use posted prices. In the bargaining task we observe the vendor's initial offer and final price (in INR). Again, there is both within- and between-vendor variance over these outcomes, suggesting that vendors are free to shirk if they pass their daily income-target early in the day but high adjustment costs could cause them to supply their typical number of hours worked.[28]

To observe hours of labor supply and effort in bargaining we hired twenty-two local agents over the two periods. Before the experiment, an extensive mapping of the market was conducted and each shop was carefully marked out. All agents were carefully trained in identifying the shops such that purchases and hours worked could be cleanly observed. Figure 4 summarizes the dates of each experiment, the number of vendors in each cell, the number of days of observation, the dynamics of treatment assignment, and the size of the treatment incentives compared to approximate control market earnings. We discuss the details of treatment for each experiment separately below.

Treatments in the Market Survey Experiment: The Market Survey Experiment consists of transitory expected increases in the wage for vendors. A total of 250 shops were selected over two periods of time—90 in the Spring of 2006 and 160 in the Fall of 2007—with the 250 shops being loosely geographically clustered into 14 geographic groupings where vendors sold similar types of goods in each cluster. After being identified, vendors were monitored by our experimenters. On day t they were only monitored for hours of labor supplied to obtain a pre-treatment baseline. At the end of day t a portion of the vendors were randomly selected as

[28] Vendors sell one or two types of betel nuts that vary slightly in price. We present data on both in Table 2. The only difference between the two types of betel nuts is price. Both types were purchased from all but 15 of the 85 vendors in the Betel Nut Experiment.

treatment vendors. The survey rewarded vendors for each hour that they recorded the number of customers that approached the shop over the course of the next two days. Vendors were told that an agent would come by their shop each hour of day $t + 1$ and $t + 2$ to collect their tally of visitors to their shop. When a vendor was done for the day they would be paid INR 10, INR 30, or INR 60 for each hour that they recorded this information.[29] Because vendors agree to participate in the market study at the end of day t, expectations going into the next day are for a higher wage, however payment was conducted at the end of days $t + 1$ and $t + 2$, as opposed to each hour that our monitors visited. The first five columns of Table 3 break down the number of vendors and days of observation for each treatment in the Market Survey Experiment.

Under the control header is also the approximate hourly wage of a vendor based on our discussions with locals—INR 10. The unit of randomization differed between sessions of the Market Survey Experiment. In Session 1, six clusters of vendors were identified and then each cluster had a treatment assignment. Then, within each cluster, vendors were randomized into treatment (the level of which was constant within cluster) or control. Treatment vendors were then randomized into a two-day period that the treatment would occur. In Session 2 eight clusters of vendors were identified. Then four of those clusters had treatment (which was constant in level again) and control status randomized; however the days of the treatment were constant within cluster. The other four clusters in Session 2 were assigned to be either all treatment or control, with treatment level also constant within cluster.

Treatment in the Betel Nut Experiment: In the second experiment we shock vendors with an unexpected overpayment for a good, allowing us to cleanly transfer an overprize, s, early in the day. In response to the overprize we observe both hours of labor supplied and vendor bargaining effort for treatment and control. The experiment was run over 3 days according to the following timeline: On day t we only observed bargaining behavior across different vendors, on $t + 1$ a random sample of vendors were selected to receive approximately a week of profit (INR 500) from a westerner. This transfer happens through the purchase of betel nuts, a mild narcotic, and a huge overprize is made to the shopkeeper without any bargaining by the westerner. After quickly delivering the shock to each vendor, the westerner leaves the market immediately. In the time

[29] Two vendors in session 1 declined participation in the INR 10 treatment. We omit these vendors from our sample.

spent in this market, western tourists were spotted occasionally, but only rarely, making an overpayment by a westerner the ideal method of transferring an overprize, as it is unlikely to convey information on future wages or bargaining effort that day. On day $t + 1$ and $t + 2$ we observe bargaining effort again, as well as hours of labor supplied.[30]

The commodity chosen was betel nuts, since this was a commonly found non-perishable good, allowing us to locate scores of different vendors. Our confederates were instructed to bargain at pre-selected shops that were carefully mapped out. Over the three days, both treated and untreated shops were approached several times by different agents. Each time a fixed amount of betel nuts were purchased and bargaining behavior was recorded. Two types of betel nuts are sold in the market, one type having been suspended in water for about a month. The agents were instructed to purchase 20 betel nuts on each purchase. The average price elicited beforehand was approximately 20 Rupees so agents were given 30 Rupees for each purchase. Any money retained after the purchase was the agent's to keep — bargaining was therefore incentivized for our confederates (List, 2004). For each purchase we registered: Initial offer by the vendor, final price agreed upon, time taken on bargaining, and the type of betel nut.

Figure 4 summarizes the details of this experiment. We note that the typical daily earnings of a vendor tend to be around INR 80, a total of 85 vendors were randomized into treatment or control, and the hours of labor supply was observed for two days while bargaining was observed for all three.

IV. Experimental Results

Market Survey Experiment: Figure 5 shows a time series of unconditional average hours worked in the Market Survey Experiment, which we discuss now to motivate our identification strategy below. The first set of dates is from Session 1, which was run in the Spring of 2006 and the second set of dates is from Session 2 in the Fall of 2007. The top panel plots the vendors that

[30] Importantly, neither experiment change daily earnings via a mechanism that would induce workers to reciprocate with increased hours or effort (see, e.g., Akerlof (1982), Fehr et al. (1993), Gneezy and List (2006), and Kube et al. (2012)) nor do subjects know they are participating in an experiment designed to observe labor supply responses to changes in earnings.

earned a high hourly supplemental wage (INR 30 and INR 60) against control vendors while the bottom panel plots the low wage increase vendors (INR 10) against the same set of control vendors. The numbers next to points on the time-series are included to indicate that treatment was applied on that day and whether it was the first or second day of treatment, with a 1 denoting the first day of treatment and a 2 denoting the second day of treatment.

Figure 5 allows us to compare the impact of treatment status on labor supplied both within treatment arms and between treatment and control. Immediately a number of trends jump out from Session 1. In particular, consistent with both KR and neoclassical theory, treatment status appears to lead to increases of between 0.2 to 0.6 hours of labor supplied, with the biggest increases on the second day of treatment status. This observation is independent of any possible day fixed effect for Session 1, as we randomize the start of treatment in Session 1 and both waves of treatment see an increase in labor supplied on the second day of treatment status. While treatment vendors were told at the end of day t that their wages would be increased on days $t+1$ and $t+2$, the supplemental wage may not have been entirely credible until the end of the day $t+1$ when vendors were actually paid for the first time by our agents. Thus, the time-series suggests that the purest test of the impact on hours worked of an expected increase in wages would be a comparison of pre-treatment hours worked to hours worked in day $t+2$.

Figure 5 also shows that, in general, hours of labor supplied decreased for vendors after treatment status, especially in Session 1. This is consistent with vendors trading current work for future leisure on treatment days, but to fully test this prediction would require observing vendors over a longer time-span so we focus on the comparison of pre-treatment hours of labor supplied to labor supplied on day $t+2$ because using post-treatment variation in hours worked would fail the strict exogeneity assumption of any panel regression model.

One other feature of the data that Figure 5 highlights is the increased variability in hours worked in Session 2. Recalling Table 3, Session 2 featured a larger number of vendors over a shorter period of time than Session 1. Furthermore, the variability of hours worked by geographic cluster was greater in Session 2, making inference from a comparison of means in Figure 5 difficult. Nonetheless, Figure 5 does show that the increase in labor supplied on day $t+2$ for treatment vendors holds for all treatments but INR 10 in Session 2.

To more precisely test the predictions of the three theories, we estimate the following linear model:

$$q_{igt} = \alpha_g + T'_{it}\beta + \epsilon_{igt} \tag{5}$$

Where q_{igt} represents the quantity of labor supplied by vendor i in cluster g on day t. T is a vector with dummies for the three levels of treatment status and control. β is a vector with our coefficients of interest. Augmenting the analysis we also include fixed effects for cluster (α_g).[31] Furthermore, we estimate (5) over two samples. In column (1) we use both days of treatment. In column (2) estimates are over just Day 2 of treatment. The robust standard errors reported in Table 3 are clustered at the lowest level of randomization for a geographic cluster of vendors.

Table 3 shows that across both specifications the coefficient for INR 60 is significant at traditional levels with the inclusion of day 1 of treatment substantially attenuating the estimate. For example, the coefficient of INR 60 in column (1) tells us that an expected transitory wage increase of INR 60 led to 0.238 more hours of labor supply. In column (2) that estimate increases to 0.618 more hours. Estimates of the coefficient on INR 30 are also positive, but only significant in the statistical sense when day 1 of treatment is dropped in column (2). Across both specifications we cannot detect whether the impact of a supplemental wage of INR 10 is different from zero. These estimates show that an expected increase in wages leads to an increase in hours worked. This outcome is in line with predictions of a neoclassical model of labor supply and KR, however it is not consonant with a model of labor supply according to an income-targeting model with exogenous income targets.

While our estimates are statistically significant, economically they are quite small. A back of the envelope calculation of the implied elasticity suggests an intertemporal substitution elasticity of labor supply between 0.01 to 0.03 for our sample.[32] To place this elasticity in relation to other estimates we briefly review estimates of this parameter. Card (1994, pp. 63) notes that, "[t]aken

[31] Estimation with vendor specific fixed effects does not change the analysis. Estimation with day*cluster fixed effects is not feasible because their inclusion drops the four clusters that had no variation in treatment and day.

[32] As reported in Table 4, the average hours worked for vendors in control is 8.343 hours and treatment vendors increase their hours worked by 0.6 hours in response to a 600% increase in wages, giving an elasticity of labor supplied of 0.01. The largest elasticity implied by Table 4 is given by the INR 30 treatment in column 2, where wages were increased by 300% and hours worked increased by 0.7 hours, yielding a labor supply elasticity of 0.03.

together, the literature suggests that the elasticity of intertemporal substitution is surely no higher than 0.5, and probably no higher than 0.2." Subsequent research has sought to estimate this parameter using quasi-experiments (e.g., Iceland eliminated its income tax in 1987, Bianchi et al. (2001)), finding elasticities between 0.2 and 0.4 but these elasticities are still an order of magnitude too small to explain the changes in labor supply over the business cycle. Chetty et al. (2011) argues that adjustment costs and employer constraints on changing labor supply attenuates the micro estimates of these elasticities. Fehr and Goette (2007) report evidence consistent with this claim from a field experiment using bike messengers in Zurich with high-frequency data. However, we have argued that our sample faces no such constraints and the labor supply response in our sample is still quite small.

To explore why the response in our sample is so small, Figure 5 pools treatment vendors and compares the densities of hours worked, start time, and end time for treatment and control vendors.[33] If our vendors face frictions that prevent them from supplying labor right when the market opens or up until the market closes then the comparison of treatment and control distributions of start or end time should appear identical. Figure 5 suggests that there are no such frictions at work, as treatment vendors exploit all three margins by showing up earlier, leaving later, and taking off the day less often.[34] This leads us to attribute the differences between our findings and recent work to the fact that our estimates are driven off of changes over a very short time-span, even though it is unclear why an intertemporal substitution elasticity ought to be driven by a longer time-span. Furthermore, the magnitude of the transitory increase in wages in our experiment are so large that the returns to supplying more labor ought to overwhelm any possible adjustment costs.

Betel Nut Experiment: Whereas neoclassical theory and KR have identical predictions for the market survey experiment, the Betel Nut experiment pits the two theories against each other. Neoclassical theory predicts no response to labor supply, while KR predicts that vendors would quit early or supply less effort over the hours of labor they supply. Table 4 presents summary

---

[33] Figure 4 compares vendors on day 2 of treatment to control vendors that were also in clusters with treatment vendors. Different specifications yield the same responses.

[34] Which can be seen by comparing the left tail of hours worked.

statistics from the treatment day of the Betel Nut Experiment. The first row reports the hours of labor supplied by treatment and control vendors conditional on supplying labor in the market that day. Table 4 shows that hours of labor supplied across the two samples are identical. The same trend emerges in the bargaining data, with treatment and control vendors making identical initial offers, agreeing to the same prices, and using similar amounts of time to negotiate.

Table 4 reports the total hours worked by treatment and control vendors. In line with neoclassical theory the point estimates are strikingly similar. However, Table 4 does not report one source of variation that we can exploit. In particular, the intervention occurred just after 10:00 am on the day of treatment, but vendors arrived at the market at various times. Thus, as the day progressed, it is possible that the cumulative hours worked by a vendor explains quitting for treatment vendors but not control. To test this we follow Crawford and Meng (2011) and estimate a probit model of the probability that a vendor quits in a given hour separately for treatment and control vendors:

(6) $$S_{it} = \alpha h_{it} + \epsilon_{it}$$

Where vendor i stops working at t if $S_{it} > 0$ where h_{it} measures cumulative hours worked at hour t for vendor i. Regardless of the model we are testing, cumulative hours worked should be positively related to the probability of quitting, but if the effect of cumulative hours worked weighs more heavily for treatment vendors that passed their income target early in the morning then we would have evidence consistent with KR.

Table 5 reports the marginal effects from a probit regression where, following Farber (2005) and Crawford and Meng (2011) we assess the marginal effects at the 8th hour of work. Both estimates in Table 6 suggest that a one hour increase in hours worked corresponds to 30-40 percentage-point increase in the probability of quitting, but consistent with neoclassical theory the effect of cumulative hours worked does not weigh more heavily on treatment vendors than control vendors with the two estimates strikingly similar.

Tables 4 and 5 show that an overpayment does not lead to the predicted change in behavior of KR and an income-targeting model of hours of labor supply, however, we've made a strong assumption in our interpretation of the hours worked data from the Betel Nut Experiment. In particular we assumed that a large overpayment early in the day will leave each vendor's expected wages unchanged. KR contends that reference points are fixed at the start of each day but the neoclassical utility component in KR is still free to respond to changes in expected wages. Thus, it's possible that these two components counteract one another and the results presented so far are the sum of these two effects.

However, as we showed in the alternative dimensions discussion in Section 1, effort exerted by vendors can be indirectly observed by final prices bargained in the marketplace and expected returns to effort should be unchanged in response to an instantaneous overpayment in the marketplace.

The bottom rows of Table 4 show that treatment and control vendors were equally effective at bargaining prices suggesting they supplied similar levels of effort in bargaining on the treatment day. Nonetheless, we can exploit the panel structure of the bargaining data to consider the possibility that the sample means in Table 4 are motivated by other sources of variation. Figure 6 presents a time series of mean final price standardized across Betel Nut type and confederate conducting the bargaining for each hour-day over the course of the experiment for treatment and non-treatment vendors. The number next to each dot in Figure 6 shows the quantity of transactions observed in that hour-day combination.

Comparing final price for treatment and non-treatment vendors before and after treatment, no substantial trends emerge. Before the treatment day, prices vary from hour-to-hour between treatment and control vendors but the number of transactions per hour is small and averaged over the entire day there is no significant difference. On treatment day, (2/28), the final price for treatment vendors start out lower (although not lower in a statistically significant sense) than control vendors. However this trend reverses later in the day. No trend emerges from the post-treatment days of observation either.

To more formally rule out the influence of other sources of variation we estimate the following linear model:

$$p_{ite} = \delta_t + \alpha_i + \beta T_{it} + \epsilon_{ite} \tag{7}$$

Where p_{ite} is the final price agreed upon for vendor i on day t with experimenter e. We run a number of specifications that include fixed effects for day, δ_t, and vendor, α_i, in order to test the robustness of treatment, T_{it}, on price. Furthermore, we include control variables for the type of betel nut bargained over.

Table 6 shows results from estimating (7). Moving from column to column of Table 6 the treatment status coefficient is consistently close to zero suggesting that receiving a large overpayment early in the day does not affect vendor effort in bargaining. In fact, as control variables are added, the sign of treatment actually becomes positive, although not large. Instead of reference points describing vendor effort, it appears that vendors treat the overpayment just as a neoclassical model would predict.

While it may be possible that vendors shirk over other dimensions of their effort, it is unclear how any such dimension could affect the vendor's profits, as we have generated data that speaks both to the effort a vendor spends on bargaining and the number of hours vendors spend performing this task. Previous research on reference points and effort provision in lab experiments (e.g., Abeler et al. (2011)) has similarly captured the profits to subjects by varying the payoff to quantity and quality of effort. Yet, we find that when such tasks are conducted in the field outside the scrutiny of the lab, there is little advantage of adding gain-loss utility to a simple neoclassical model of labor supply.

V. Concluding Remarks

Understanding the behavioral response to transitory changes in wages is critical for designing policies ranging from taxation, unemployment benefits, and intergenerational transfers programs like Social Security. As a result, a large literature in economics attempts to estimate the magnitude of the intertemporal elasticity substitution for labor supply, the response to transitory (as opposed to permanent) changes in the wage rate. A recent literature in behavioral economics has emerged questioning whether the neoclassical model should be relied on to generate such estimates. In particular, Farber (2008) notes:

> Evaluation of much government policy regarding tax and transfer programs depends on having reliable estimates of the sensitivity of labor supply to wage rates and income levels. To the extent that individuals' levels of labor supply are the result of optimization with reference dependent preferences, the usual estimates of wage and income elasticities are likely to be misleading.

In this study, we design two field experiments to test the role of reference-dependent preferences in the labor supply decision. We present a several competing models of labor supply and show how an ideal experiment would test the predictions of these competing models. Beyond providing empirical insights, theoretically, the results of such exercises are important, as questions of what models best explain behavior and whether life cycle labor supply models explain business-cycle labor supply behavior remain ill-understood.

We report that expected changes in wages lead to increases in labor supply—a finding consistent with neoclassical models of labor supply and KR but not with an income-targeting model. We then report evidence on the behavioral response to unexpected changes in wages. Consistent with a neoclassical model of labor supply we find that unexpected changes in wages have no impact on hours of labor supply or on effort provided, suggesting that reference utility does not drive behavior of vendors in our study.

Methodologically our study differs substantially from earlier work. By using a field experiment in a naturally-occurring market, we can control changes in earnings and also changes in expected earnings by simply divulging or withholding information. This added control of variation, however, does not come at the expense of using subjects unfamiliar with allocating their labor in the market studied. In particular, vendors mimic the features of those in studies that have used naturally occurring data in the sense that the work of vendors is to efficiently allocate their labor supply in the face of day-to-day changes in earnings.

Overall, we find that the predictions of the neoclassical model speak most clearly to observed behavior. This, of course, does not mean that continued behavioral modeling is fruitless. More work needs to be done to determine if our results are replicable; furthermore, much can be gained by deepening our understanding of the models that can explain behavioral responses to important changes in the economic environment.

References

Abeler, Johannes, Armin Falk, Lorenz Goette, and David Huffman, "Reference Points and Effort Provision," *American Economic Review*, April 2011, 101 (2), 470–492.

Akerlof, George A., "Labor Contracts as Partial Gift Exchange," *Quarterly Journal of Economics*, 1982, 97 (4), 543–569.

Altonji, Joseph G., "Intertemporal Substitution in Labor Supply: Evidence from Micro Data," *Journal of Political Economy*, June 1986, 94 (3), S176–S215.

Bianchi, Marco, Bjorn R. Gudmundsson, and Gylfi Zoega, "Iceland's Natural Experiment in Supply-Side Economics," *American Economic Review*, 2001, 91 (5), 1564–1579.

Camerer, Colin, Linda Babcock, George Loewenstein, and Richard Thaler, "Labor Supply of New York City Cabdrivers: One Day At A Time," *Quarterly Journal of Economics*, 1997, 112 (2), 407–441.

Card, David, "Intertemporal Labour Supply: An Assessment," in Christopher Sims, ed., *Advances in Econometrics: Sixth World Congress*, Volume II, Cambridge University Press 1994.

Card, David and Gordon B. Dahl, "Family Violence and Football: The Effect of Unexpected Emotional Cues on Violent Behavior," *Quarterly Journal of Economics*, 2011, 126 (1), 103–143.

Chetty, Raj, John N. Friedman, Tore Olsen, and Luigi Pistaferri, "Adjustment Costs, Firm Responses, and Micro vs. Macro Labor Supply Elasticities: Evidence From Danish Tax Records," *Quarterly Journal of Economics*, 2011, 126 (2), 749–804.

Chou, Yuan K., "Testing Alternative Models of Labor Supply: Evidence from Taxi-Drivers in Singapore," Working Paper, 2002.

Crawford, Vincent P. and Juanjuan Meng, "New York City Cabdrivers' Labor Supply Revisited: Reference-Dependent Preferences with Rational-Expectations Targets for Hours and Income," *American Economic Review*, 2011, 101 (5), 1912–1932.

Englemann, Dirk and Guillaume Hollard, "Reconsidering the Effect of Market Experience on the 'Endowment Effect'," *Econometrica*, 2010, 78 (6): 2005-2019.

Farber, Henry S., "Is Tomorrow Another Day? The Labor Supply of New York City Cab-drivers," *Journal of Political Economy*, 2005, 113 (1), 46–82.

Farber, Henry S., "Reference-Dependent Preferences and Labor Supply: The Case of New York City Taxi Drivers," *American Economic Review*, 2008, 98 (3), 1069–1082.

Fehr, Ernst and Lorenz Goette, "Do Workers Work More if Wages Are High? Evidence from a

Randomized Field Experiment," *American Economic Review*, 2007, 97 (1), 298–317.

Fehr, Ernst, Georg Kirchsteiger, and Arno Riedel, "Does Fairness Prevent Market Clearing? An Experimental Investigation," *Quarterly Journal of Economics*, 1993, 108 (2), 437–459.

Fryer, Roland G., Steven D. Levitt, John List, and Sally Sadoff, "Enhancing the Efficacy of Teacher Incentives through Loss Aversion: A Field Experiment," NBER Working Paper No. 18237, 2012.

Gentzkow, Matthew and Jesse M. Shapiro, "Measuring the Sensitivity of Parameter Estimates to Sample Statistics," Working Paper, 2014.

Ghez, Gilbert R. and Gary S. Becker, *Allocation of Time and Goods over the Life Cycle*, National Bureau of Economic Research, 1975.

Gneezy, Uri and John A. List, "Putting Behavioral Economics To Work: Testing For Gift Exchange In Labor Markets Using Field Experiments," *Econometrica*, 2006, 74 (5), 1365– 1384.

Goette, Lorenz, David Huffman, and Ernst Fehr, "Loss Aversion and Labor Supply," *Journal of the European Economic Association*, 2004, 2 (2-3), 216–228.

Goldberg, Jessica, "Kwacha Gonna Do? Experimental Evidence about Labor Supply in Rural Malawi," Working Paper, 2014.

Heffetz, Ori and John A. List, "Is the Endowment Effect an Expectations Effect," Forthcoming, *Journal of the European Economic Association*.

Hossain, Tanjim and John A. List, "The Behavioralist Visits The Factory: Increasing Productivity Using Simple Framing Manipulations," *Management Science*, 2012, 58 (12): 21-51.

Kahneman, Daniel and Amos Tversky, "Prospect Theory: An Analysis of Decision under Risk," *Econometrica*, March 1979, 47 (2), 263–292.

Kahneman, Daniel, Jack L. Knetsch, and Richard H. Thaler, "Anomalies: The Endowment Effect, Loss Aversion, and Status Quo Bias," *Journal of Economic Perspectives*, 1991, 5 (1), 193–2006.

Knetsch, Jack L., "The Endowment Effect and Evidence of Nonreversible Indifference Curves," *American Economic Review*, 1989, 79 (5), 1277–1284.

Knetsch, Jack L. and Wei-Kang Wong, "The endowment effect and the reference state: Evidence and manipulations," *Journal of Economic Behavior & Organization*, 2009, 71 (2), 407-413.

Koszegi, Botond and Matthew Rabin, "A Model of Reference-Dependent Preferences," *Quarterly Journal of Economics*, 2006, 121 (4), 1133–1165.

Kube, Sebastian, Clemens Puppe, and Michel Andre Marechal, “The Currency of Reciprocity – Gift Exchange in the Workplace,” *American Economic Review*, 2012, 102 (4), 1655-1662.

Levitt, Steven D., John A. List, Susanne Neckermann, Sally Sadoff, "The Behavioralist Goes to School: Leveraging Behavioral Economics to Improve Educational Performance," NBER Working Paper No. 18165, 2012.

List, John A., “Does Market Experience Eliminate Market Anomalies?” *Quarterly Journal of Economics*, 2003, 118 (1), 41-72.

List, John A., “Neoclassical Theory Versus Prospect Theory: Evidence from the Marketplace,” *Econometrica*, 2004, 72 (2), 615-625.

MaCurdy, Thomas E., “An Empirical Model of Labor Supply in a Life-Cycle Setting,” *Journal of Political Economy*, 1981, 89 (6), 1059–1085.

Mulligan, Casey B., “The Intertemporal Substitution of Work What Does the Evidence Say?,” Working Paper, 1995.

Oettinger, Gerald S., “An Empirical Analysis of the Daily Labor Supply of Stadium Vendors,” *Journal of Political Economy*, 1999, 107 (2), 360–392.

Pope, Devin G. and Maurice E. Schweitzer, “Is Tiger Woods Loss Averse? Persistent Bias in the Face of Experience, Competition, and High Stakes,” *American Economic Review*, 2011, 101 (1), 129–157.

Table 1: Comparative Statics of Labor Supply Models

	$\frac{dh}{dw}$	$\frac{dh}{ds_0}$
Standard Neoclassical Model	> 0	≈ 0
Income Targeting Model	< 0	< 0
Reference Dependent Model (KR)	> 0	< 0

Note: Each cell contains the sign of comparative static predictions of three separate models of daily labor supply. Labor supply, h, responds to changes in wage rate, w, and to unexpected changes in wages consonant with a large tip or overpayment that is uncorrelated with future income, s_0.

Table 2: Summary Statistics

		Standard Deviation	
	Mean	Within	Between
Market Survey Experiment			
Hours	8.472	1.069	1.405
Participate	.976	.096	.116
Opening Time	9.622	.586	.764
Break	.025	.117	.091
Closing Time	18.33	.402	.521
Betel Nut Experiment			
Hours	8.733	1.028	1.508
Betel Nut Variety 1			
Initial Offer (INR)	18.333	3.232	4.163
Final Price (INR)	17.674	3.222	3.975
Betel Nut Variety 2			
Initial Offer (INR)	21.111	2.958	3.267
Final Price (INR)	19.829	3.007	2.749

Note: Summary statistics are reported for the two experiments. Standard deviations are reported for within vendor and between vendors. For the Market Survey Experiment the statistics reported are daily hours worked, probability of supply any labor, start time (in local 24-hour clock time), end time, and probability of taking a break during the course of the work day. For the Betel Nut Experiment the statistics reported are hours worked, initial price offer in bargaining task (in INR), and final price in bargaining (in INR).

Table 3: Labor Supply Regressions

	(1) Trt Days 1 & 2 Hours	(2) Trt Day 2 Hours
INR 10	0.026	0.231
	(0.251)	(0.245)
INR 30	0.134	0.366*
	(0.194)	(0.210)
INR 60	0.233**	0.609***
	(0.112)	(0.144)
Area Control	Yes	Yes
R^2	0.170	0.186
N	982	781

Note: Dependent variable is the number of hours of labor supplied in columns. Average hours worked by vendors in control in this sample was 8.343 hours. The unit of observation is vendor*day. Column (1) uses pre-treatment observations and both days 1 and 2 of treatment to estimate coefficients. Column (3) uses pre-treatment and day 2 of treatment to estimate coefficients. Post-treatment days are omitted for vendors assigned to both treatment and control for all columns. Robust standard errors accounting for relevant unit of clustering are reported in parentheses. Variance-covariance matrix estimated accounts for lowest level of randomization. For groups of vendors where treatment was varied within group, clustering unit is vendor. For groups where treatment did not vary within group, clustering unit is the group. Area Control fixed effects control for the geographic cluster of vendors.
$*p \leq 0.10, **\,p \leq 0.05, ***p \leq 0.01$

Table 4: Betel Nut Experiment Results—Summary Statistics

	Treatment Day	
Variable	Treated	Untreated
Total Hours—Working	8.9	8.9
	(.266)	(.118)
Initial Offer (INR)		
Betel Nut Variety 1	17.9	17.8
	(.983)	(.882)
Betel Nut Variety 2	21.4	21
	(1.42)	(1.006)
Final Price (INR)		
Betel Variety 1	17.1	17.4
	(.9)	(.838)
Betel Variety 2	20	19.9
	(1.323)	(.891)

Note: Summary statistics are reported for the treatment day of the Betel Nut Experiment. Statistics are reported for total hours worked conditional on working on treatment day and initial offer and final price (in INR) in the bargaining experiment for both types of Betel Nuts. Standard errors are in parentheses below.

Table 5: Marginal Effects On The Probability of Quitting—Split By Treatment Status

Variable	X^*	(1) Treated	(2) Untreated
Total Hours	8	.39	.374
		(.058)	(.072)
N		488	305
Log likelihood		-58.08	-38.77

Note: The sample is broken down by vendors in treatment and control. The number of observations is the number of hours worked by vendors on the day of treatment (2/28/2006). Marginal effects are evaluated at eight total hours worked, in keeping with Farber (2005) and Crawford and Meng (2011). Robust standard errors are reported under each marginal effect.

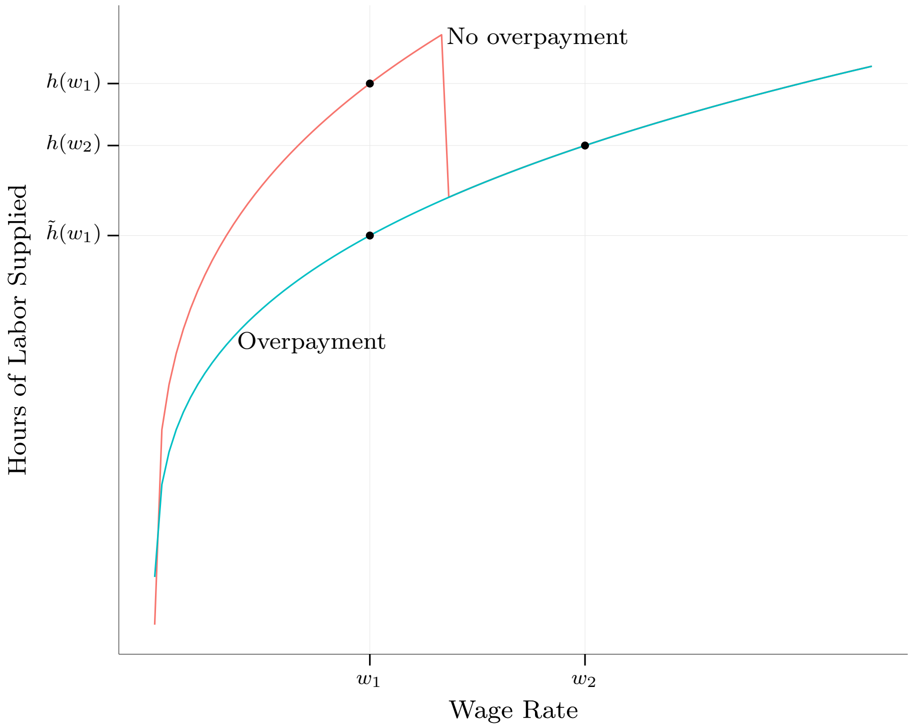
Table 6: Effort Provision: Bargaining Regressions

	(1) Final INR	(2) Final INR	(3) Final INR
Treatment	-0.325	1.081	0.910
	(0.765)	(1.243)	(1.096)
Constant	17.732***	16.756***	17.161***
	(0.498)	(0.619)	(0.781)
Betel Type Control	Yes	Yes	Yes
Vendor Control		Yes	Yes
Day Control		Yes	Yes
Confederate Control			Yes
R^2	0.059	0.421	0.529
N	258	258	258

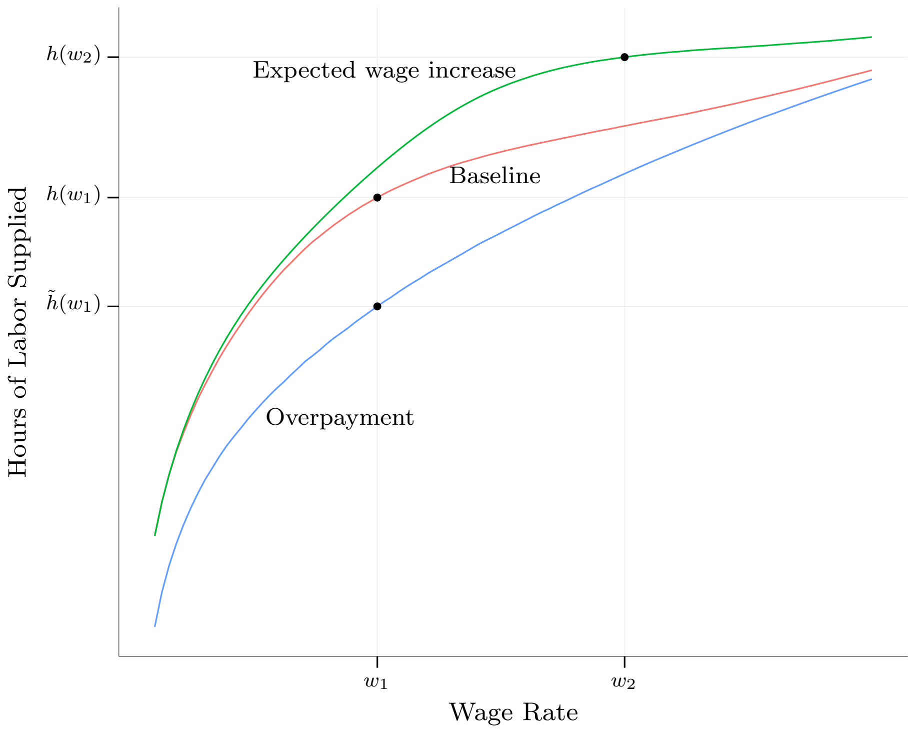
Note: Dependent variable is a dimension of effort provision elicited in negotiations with treated and untreated shops over three days of observation. The unit of observation is vendor*day*negotiation. The outcome variable in all three columns is final price in INR. Robust standard errors clustered at the vendor level are reported. Betel Type Control means that the Betel Nut type that was negotiated over was controlled for. Vendor Control means vendor fixed effects were estimated, while Day Control means a fixed effect for each day was estimated. Confederate Control estimates a fixed effect for the six confederates that were bargaining over the three days of observation.

$*p \leq 0.10, **p \leq 0.05, ***p \leq 0.01$

Figure 1: Simulated Labor Supply Response
Panel A: Simulated Labor Supply for Income Targeting Model



Panel B: Simulated Labor Supply for KR



Note: We assume the following preference parameterization: $u(y) = \frac{y^{1-\sigma}-1}{1-\sigma}$ and $v(h) = \gamma\frac{\epsilon}{1+\epsilon}h^{\frac{1+\epsilon}{\epsilon}}$ with the following parameter values: $\sigma = 0.01, s_0 = 10,000, \gamma = 0.005, \epsilon = 0.3, \lambda = 3$, and $y_r = 100$. When there is an income shock s_0 increases to $10,500$. For KR the baseline wage is drawn from $w \sim \Gamma(4, 0.2)$ and $P\{\text{Overpayment occurs}\} = 0.001$. For KR with an expected increase in the realized wage is simply, $\tilde{w} = w + 10, w \sim \Gamma(4, 0.2)$. Conceptually, the wage and overpayment is realized as soon as workers show up at the market that day and they pick their hours of labor supply accordingly.

Figure 2: Budget Constraint and Reference Point Dynamics with Expected Wage Change and Income Shock

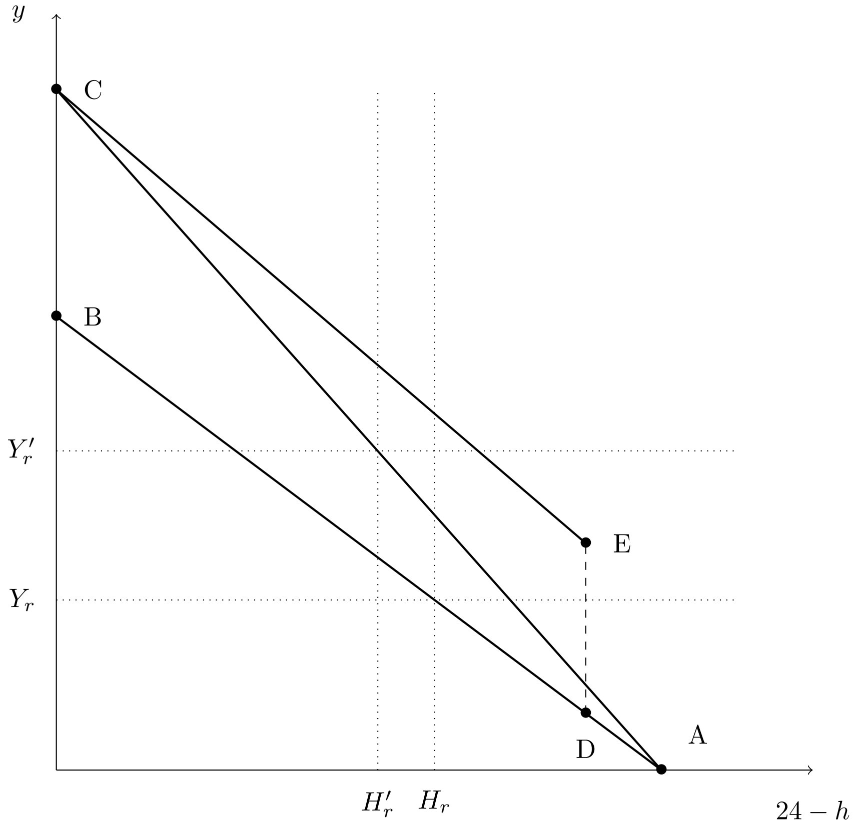
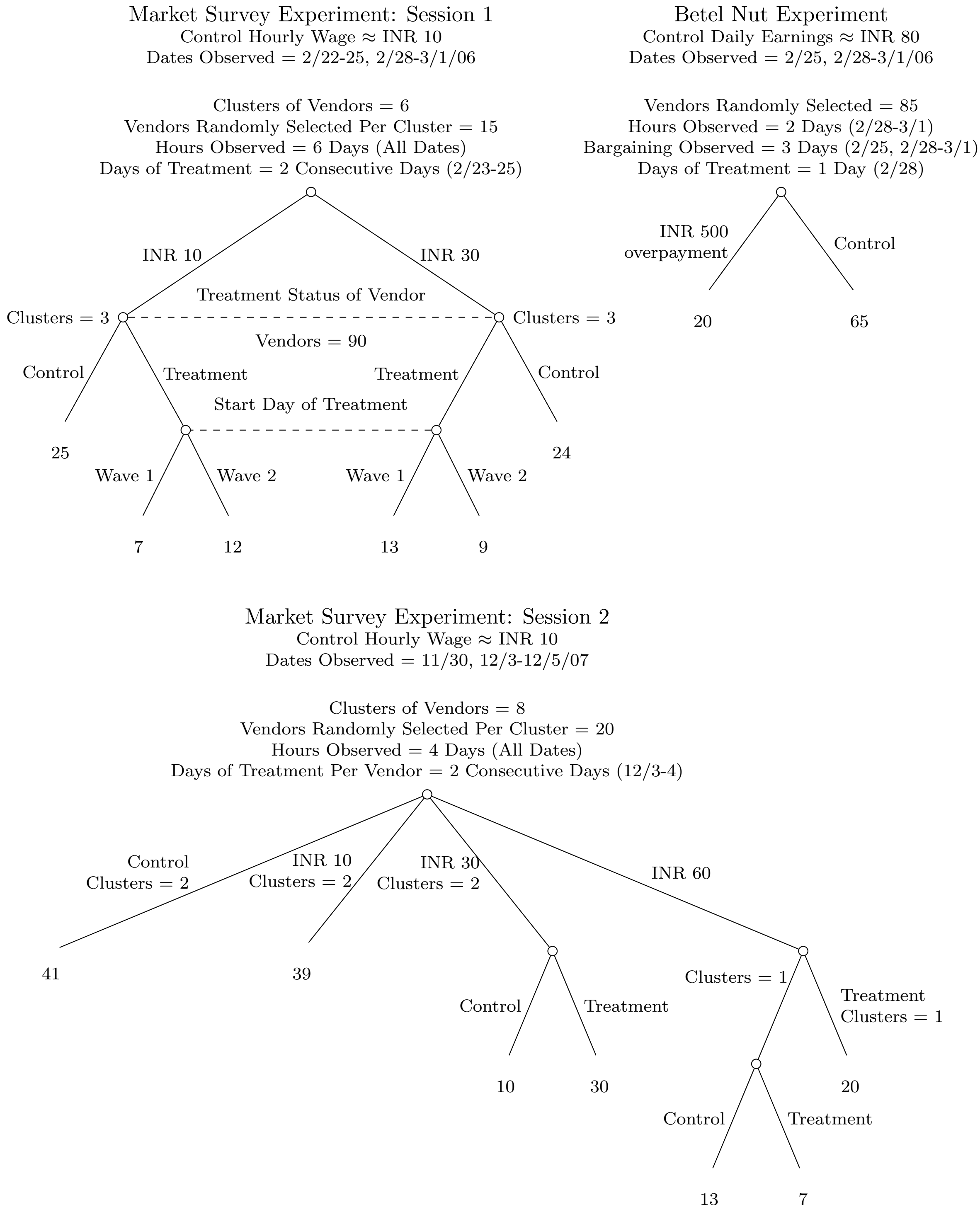


Figure 3: Sensitivity of Identification in Crawford and Meng (2011)

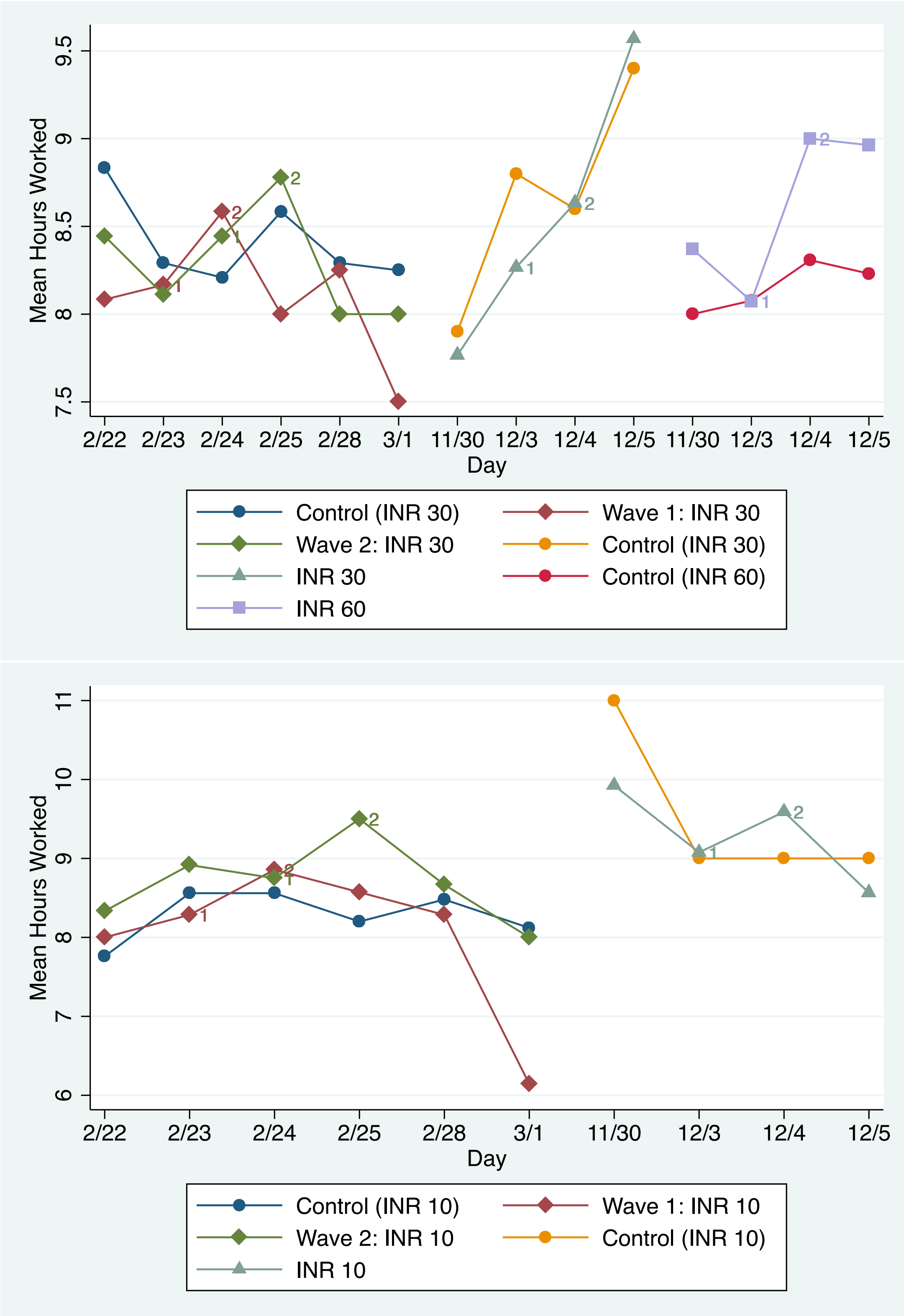
Note: Scaled sensitivity of marginal effects from Crawford and Meng (2011) Table 2, Panel B, Columns 3 and 4 to changes in expected first hour earnings and realized first hour earnings (see Gentzkow and Shapiro, 2014 for details on calculation).

Figure 4: Experimental Design



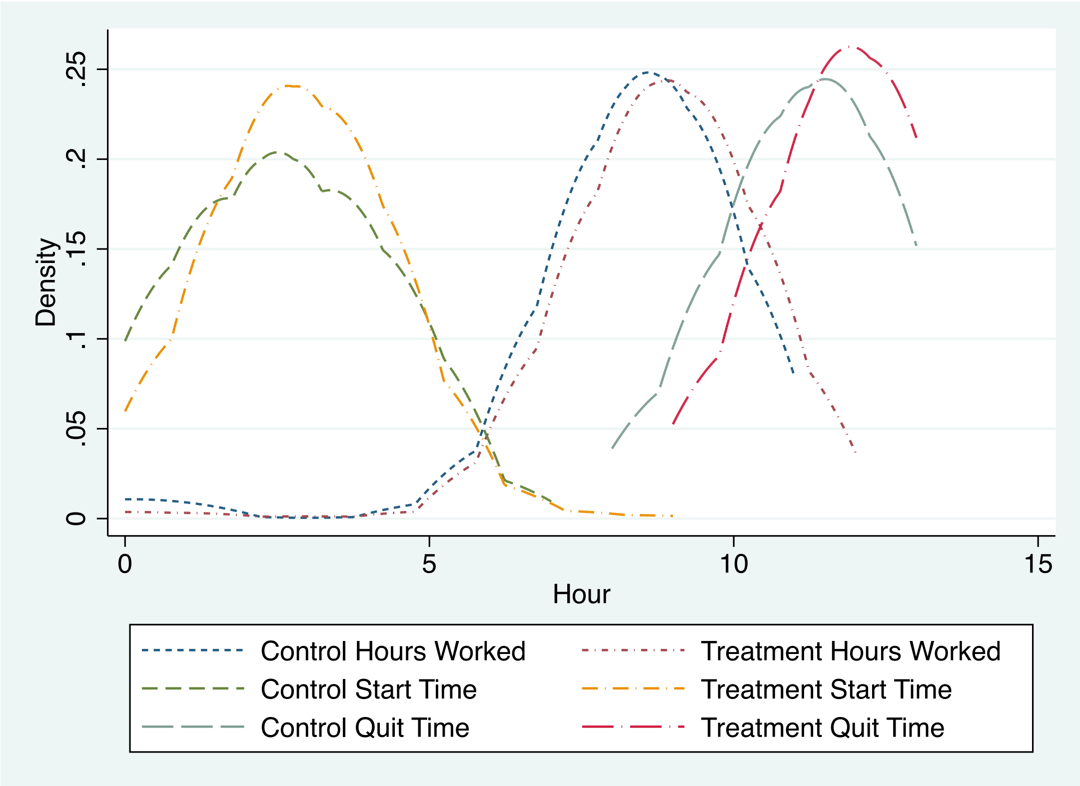
Note: Each tree summarizes the data collected on vendors in a given experiment and the manner in which treatment was assigned to individual vendors. The terminus of each edge represents the number of vendors. The label next to each edge represents the treatment randomly assigned. Labels above dotted lines indicate a change in the unit of randomization. The number of clusters is denoted to show how the sample is divided. Vendors in *Market Survey Experiments* were a randomly selected subset of 15 or 20 vendors (depending on the session) from a geographic cluster of vendors. In Session 1, clusters were then assigned a supplemental wage of INR 10 or INR 30 for two days, after which vendors within a cluster were randomized into treatment. Treatment vendors then had the start of treatment randomized. In Session 2 vendors in Control and INR 10 clusters received treatment. Vendors in clusters assigned INR 30 and INR 60 had treatment randomized within cluster but for one cluster of INR 60. No data was observed on 2/26-27 in Session 1 and 12/1-2 in Session 2 because they were "Market Days" where market conditions would not be comparable to the other days of the week. All treatment vendors in Session 2 were offered the expected wage increase on 12/3-4. The *Betel Nut Experiment* randomized 85 vendors into receiving an overpayment of INR 500 or not. Hours of labor supply were observed on 2/28 and 3/1. Confederates blind to treatment bargained with Betel Nut Experiment vendors on 2/25, 2/28-3/1. Treatment occurred on 2/28. Across sessions and experiments, no vendor was placed into multiple treatment arms. In the Market Survey Experiment: Session 1, one vendor received INR 30 in an INR 10 cluster. In Market Survey Experiment: Session 2, one vendor received Control in an INR 10 cluster. This detail is omitted for simplicity.

Figure 5: Time Series of Labor Supply by Treatment



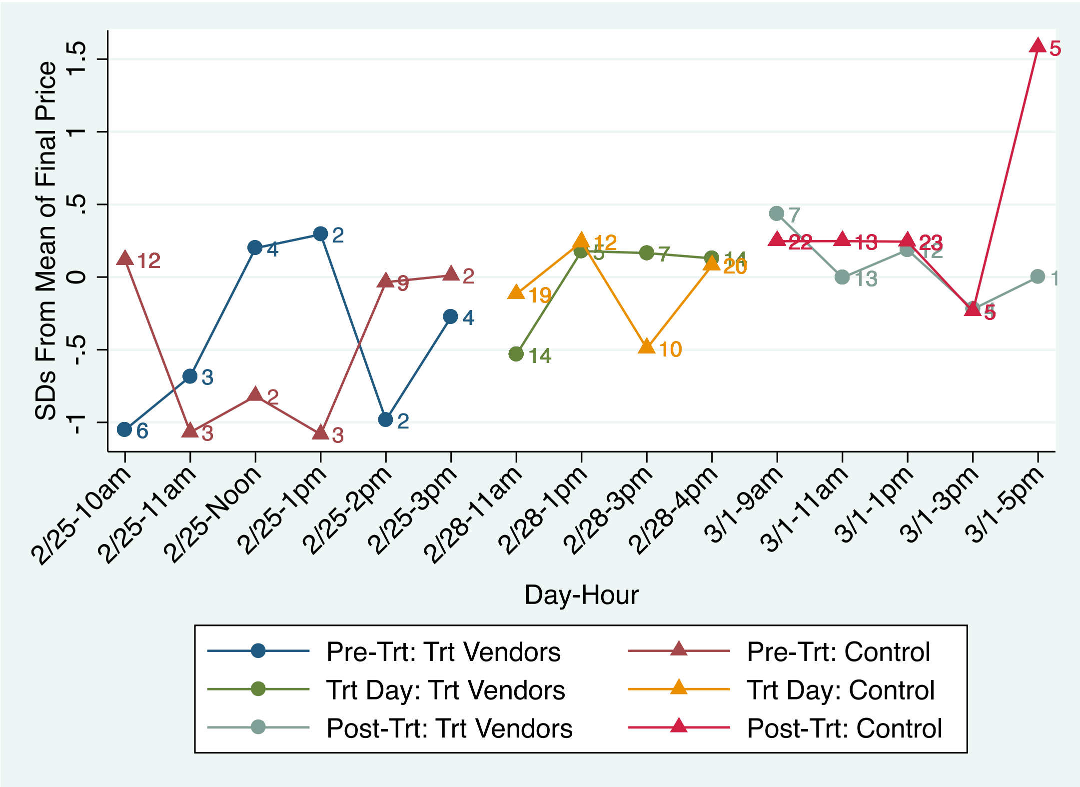
Note: The two time series show hours worked by day for session 1 and 2 of the Market Survey Experiment. The top panel shows hours worked for the INR 30 and INR 60 treatments. The bottom panel shows the hours worked for the INR 10 treatment. The number next to a day's mean conveys the day of treatment for that time series, 1 meaning that it was the first day of treatment status and 2 meaning it was the second day of treatment status. Values for the control group are calculated off of hours worked by vendors assigned to control in the same geographic clusters as vendors receiving treatment at that level.

Figure 6: Density Plot of Margins of Labor Supply



Note: The Kernel Density plots summarize the distribution of hours worked, start time, and quitting time for vendors in treatment on day 2 and compares to vendors in control that were in the same clusters as treatment vendors. Start and quit time has been normalized to hour 0—the opening time of the market, which varied over the two sessions of the Market Survey Experiment. Hours worked is the total number of hours of labor supplied.

Figure 7: Time Series of Final Price by Treatment



Note: The time series shows mean final price for each day-hour combination as negotiated by incentivized confederates over the course of the three days of observation standardized to mean zero and unit standard deviation for each Confederate*Betel Nut Type combination. The number next to each day-hour's mean conveys the number of transactions recorded in that day-hour.