Lecture 5 Intertemporal Labor Supply (continued)

References

Raj Chetty: "A New Method of Estimating Risk Aversion", *AER,* December 2006.

Tomas Havranek "Measuring Intertemporal Substitution: The Importance of Method Choices and Selective Reporting" JEEA, 2015.

David Card. "Intertemporal Labor Supply: An Assessment". In Christopher Sims, editor, *Advances in Econometrics, Sixth World Congresss* (volume 2). Cambridge University Press, 1994

Luigi Pistaferri. "Anticipated and Unanticipated Wage Changes, Wage Risk, and Intertemporal Labor Supply." *JOLE* 21 (3), 2003. pp. 729-754.

Thomas MaCurdy. "An Empirical Model of Labor Supply in a Life Cycle Setting" *JPE* 89 (6) 1981, pp.1059-1085.

Joseph Altonji. "Intertemporal Substitution in Labor Supply: Evidence from Micro Data". *JPE* 94 (3, part 2) 1986, pp. S176-S215.

Ham, John C., and Kevin T. Reilly. 2002. "Testing Intertemporal Substitution, Implicit Contracts, and Hours Restriction Models of the Labor Market Using Micro Data." American Economic Review, 92(4): 905-927.

Henry Farber. "WhyYou Can't Find a Cab in the Rain and Other Labor Supply Lessons from Cab Drivers." NBER Working Paper 20604, October 2014. Published in QJE, 2015.

a. Recap; More on the Relation of Intertemporal and Static Responses

Last time we laid out the prototypical setup: consumption in period t is c_t, hours of work are h_t, the wage is w_t. Individuals have flow utility $u(c_t, h_t; a_t)$ that is concave in (c, h), and an intertemporal budget contraint:

$$A_{t+1} = (1 + r_t)(A_t + y_t + w_t h_t - c_t).$$

The Bellman equation is

$$V_t(A_t) = \max_{c_t, h_t} u(c_t, h_t; a_t) + \beta E_t[V_{t+1}((1 + r_t)(A_t + y_t + w_t h_t - c_t))]$$

After defining $\lambda_t \equiv V_t'(A_t)$ we get the f.o.c. (assuming an interior solution for h_t) :

$$\begin{aligned} u_c(c_t, h_t; a_t) &= \lambda_t \\ u_h(c_t, h_t; a_t) &= -w_t \lambda_t \end{aligned}$$

and the intertemporal optimum condition:

$$\lambda_t = \beta(1 + r_t)E_t[\lambda_{t+1}].$$

Aside on the Intertemporal Elasticity based on Consumption

Some macro models ignore labor supply and focus on the implications of this simple model for the allocation of consumption over time. To illustrate, consider the additively separable case where

$$u(c_t, h_t; a_t) = v(c_t) - \phi(h_t; a_t)$$

In this case, Hall (1978) noted that, setting $\beta = (1+\delta)^{-1}$, the intertemporal optimum condition becomes:

$$E_t[v'(c_{t+1})] = \frac{1+\delta}{1+r_t} v'(c_t).$$

If we further assume that $v(c_t) = \frac{\sigma}{\sigma-1} c_t^{(\sigma-1)/\sigma}$, then $v'(c_t) = c_t^{-1/\sigma}$ and we get:

$$E_t[(c_{t+1})^{-1/\sigma}] = \frac{1+\delta}{1+r_t}(c_t)^{-1/\sigma}.$$

Taking logs of both sides and assuming we can interchange E and log, we get:

$$E_t log(c_{t+1}) = log(c_t) + \sigma log(1+r_t) - \sigma log(1+\delta)$$

which means that

$$log(c_{t+1}) - log(c_t) = constant + \sigma log(1+r_t) + error$$

where the error is an expectation error, and should be orthogonal to everything dated t or earlier. Here σ is the "elasticity of intertemporal substitution in consumption" (IES in the terminology of Havranek, 2015). Notice that $-\frac{c\, v_{cc}}{v_c} = 1/\sigma$, so σ is the inverse coefficient of relative risk aversion.

Back to Labor Supply

Define the Frisch demands as the solutions to these f.o.c., given (w_t, λ_t) and the preference shocks:

$$\begin{aligned} c_t &= c^F(w_t, \lambda_t, a_t) \\ h_t &= h^F(w_t, \lambda_t, a_t) \end{aligned}$$

Let's log-linearize the Frisch demands:

$$\begin{aligned} \log h_t &= A_t + \eta \log w_t + \delta \log \lambda_t \\ \log c_t &= B_t + \theta \log w_t + \kappa \log \lambda_t. \end{aligned}$$

Differentiating the f.o.c. we get

$$\begin{pmatrix} dc \\ dh \end{pmatrix} = \begin{bmatrix} U_{cc} & U_{ch} \\ U_{hc} & U_{hh} \end{bmatrix}^{-1} \begin{bmatrix} 1 & 0 \\ -w & -\lambda \end{bmatrix} \begin{bmatrix} d\lambda \\ dw \end{bmatrix}$$

So

$$\begin{aligned}
\frac{\partial h^F}{\partial w} &= \frac{-\lambda U_{cc}}{\Delta} \\
\frac{\partial h^F}{\partial \lambda} &= \frac{-wU_{cc} - U_{hc}}{\Delta} \\
\frac{\partial c^F}{\partial \lambda} &= \frac{wU_{ch} + U_{hh}}{\Delta} \\
\frac{\partial c^F}{\partial w} &= \frac{\lambda U_{ch}}{\Delta}
\end{aligned}$$

where $\Delta = U_{cc}U_{hh} - U_{ch}^2 > 0$, since for an intertemporal planning problem we need concave utility Note that

$$w\frac{\partial h^F}{\partial w} - \lambda\frac{\partial h^F}{\partial \lambda} = \frac{\partial c^F}{\partial w}$$

and dividing by h we get

$$\frac{w}{h}\frac{\partial h^F}{\partial w} - \frac{\lambda}{h}\frac{\partial h^F}{\partial \lambda} = \frac{c}{wh}\frac{w}{c}\frac{\partial c^F}{\partial w}$$

or in terms of the elasticities of the log-linearized system,

$$\eta - \delta = \frac{c}{wh}\theta.$$

On average $c \approx wh$ (other than for trust-fund babies), so this says that $\eta - \delta \approx \theta$. In particular, in the separable case $U_{ch} = 0$ which implies:

$$\begin{aligned}
\eta &= \delta = \frac{U_h}{hU_{hh}} \\
\kappa &= \frac{U_c}{cU_{cc}} \\
\theta &= 0
\end{aligned}$$

Note that in the separable case $\kappa = -1/R$ where R is the coefficient of relative risk aversion. Many macro economists think that $U_{ch} > 0$, implying that $\theta > 0$ and $\eta > \delta$.

It useful to relate δ and κ to the more familiar income effects in static labor supply models. To do this consider the static labor supply problem with the same preferences

$$\max_{c,h} U(c,h) \quad s.t. \quad c = y + wh$$

The f.o.c. for this problem are

$$\begin{aligned}
U_c(c,h) - \lambda &= 0 \\
U_h(c,h) + \lambda w &= 0 \\
-c + wh + y &= 0.
\end{aligned}$$

Differentiating these we get

$$\begin{bmatrix} dc \\ dh \\ d\lambda \end{bmatrix} = \begin{bmatrix} U_{cc} & U_{ch} & -1 \\ U_{hc} & U_{hh} & w \\ -1 & w & 0 \end{bmatrix}^{-1} \begin{bmatrix} 0 & 0 \\ -\lambda & 0 \\ -h & -1 \end{bmatrix} \begin{bmatrix} dw \\ dy \end{bmatrix}$$

and we can show

$$\frac{\partial c}{\partial y} = \frac{-U_{ch}w - U_{hh}}{\Delta'}$$
$$\frac{\partial h}{\partial y} = \frac{U_{cc}w + U_{ch}}{\Delta'}$$

where Δ' is the determinant of the bordered Hessian. Note that the numerators of these expressions are the same as the numerators of $\frac{\partial c^F}{\partial \lambda}$ and $\frac{\partial h^F}{\partial \lambda}$ respectively (with a sign change). Thus:

$$\frac{\frac{\partial h^F}{\partial \lambda}}{\frac{\partial c^F}{\partial \lambda}} = \frac{\frac{\partial h}{\partial y}}{\frac{\partial c}{\partial y}}.$$

This is useful because we know $\frac{\partial c}{\partial y} = 1 + mpe$ and $w\frac{\partial h}{\partial y} = mpe$, where mpe is the marginal propensity to earn out of non-labor income, and is thought to be a number like -0.1 or so. Thus

$$\frac{w\frac{\partial h^F}{\partial \lambda}}{\frac{\partial c^F}{\partial \lambda}} = \frac{mpe}{1 + mpe}.$$

Converting to elasticities, we get:

$$\frac{wh\frac{\lambda}{h}\frac{\partial h^F}{\partial \lambda}}{c\frac{\lambda}{c}\frac{\partial c^F}{\partial \lambda}} = \frac{wh}{c}\frac{\delta}{\kappa} = \frac{mpe}{1 + mpe}$$

implying that

$$\frac{\delta}{\kappa} = \frac{c}{wh}\frac{mpe}{1 + mpe}.$$

This says that the ratio of the elasticities of labor supply and consumption with respect to λ is (roughly) the same as the ratio of mpe to $(1 + mpe)$. if $mpe = -0.1$ then $\frac{mpe}{1+mpe} \approx -0.1$.

In the additively separable case we can use this ratio to think about a likely magnitude for η. Specifically, as shown above in the separable case $\kappa = \frac{U_c}{cU_{cc}}$ is the negative of the inverse of the coefficient of relative risk aversion defined over gambles. Define

$$R = -c\frac{U_{cc}}{U_c}.$$

Then $\kappa = -1/R$. So if $R = 1$ and $mpe = -0.1$, then $\delta \approx 0.1$, implying $\eta = 0.1$ (since in the separable case $\eta = \delta$).

As noted above, in the macro consumption literature (which ignores labor supply and focuses on the optimal allocation of consumption over the lifecycle) it is standard to estimate $\kappa = -IES$ by relating the change in consumption between periods to the real interest rate. Some commentators have argued that a plausible estimate based on this approach is $\kappa = -1$, though a recent meta-analysis by Havranek (JEEA, 2015) suggests that there is *selective reporting* in the literature, and that the corrected measure of central tendency is smaller.

To get a larger value for η when we think that κ is not too big we need to have a big value for θ (since $\eta \approx \delta + \theta$). This means that researchers who believe in big-η need to specify non-separable preferences, with $U_{ch} > 0$. One way that people have tried to estimate U_{ch} is to look at consumption expenditures of people around the point of retirement, though this has problems because people can do home production or change the way they shop when they retire.

The relationship between the coefficient of relative risk aversion and the Frisch elasticity of labor supply is developed nicely in Chetty's 2006 *AER* paper. Richard Rogerson and Johanna Wallenius (AER, 2013) is an example of recent work that tries to look at retirement behavior and learn something about intertemporal labor supply.

b. Reduced Form Evidence on Intertemporal Labor Supply Elasticities

Let's return to the log-linearized labor supply and consumption equations:

$$\begin{aligned} \log h_t &= A_t + \eta \log w_t + \delta \log \lambda_t \\ \log c_t &= B_t + \theta \log w_t + \kappa \log \lambda_t. \end{aligned}$$

Focusing on hours, we can difference over time:

$$\Delta \log h_t = \log h_t - \log h_{t-1} = \Delta A_t + \eta \Delta \log w_t + \delta(\log \lambda_t - \log \lambda_{t-1}).$$

Next, use the fact that $\lambda_{t-1} = \beta(1 + r_{t-1})E_{t-1}[\lambda_t]$. Thus

$$\log \lambda_{t-1} = \log[\beta(1 + r_{t-1})] + \log E_{t-1}[\lambda_t]$$

Now define

$$\phi_t = \log E_{t-1}[\lambda_t] - E_{t-1} \log[\lambda_t]$$

and define the innovation in the log marginal utility of income as ξ_t where:

$$\log \lambda_t = E_{t-1} \log[\lambda_t] + \xi_t.$$

Combining all these terms we get

$$\log \lambda_{t-1} = \log[\beta(1 + r_{t-1})] + \log \lambda_t - \xi_t + \phi_t.$$

If we write $\beta = (1 + \rho)^{-1}$ and approximate $\log[(1 + r_{t-1})/(1 + \rho)] = r_{t-1} - \rho$ we get a very useful expression for the evolution of the log marginal utility of income:

$$\log \lambda_t = \log \lambda_{t-1} - (r_{t-1} - \rho) - \phi_t + \xi_t$$

So

$$\Delta \log h_t = \Delta A_t + \eta \Delta \log w_t + \delta \xi_t - \delta(r_{t-1} - \rho) - \delta \phi_t. \quad (1)$$

And following the same steps:

$$\Delta \log c_t = \Delta B_t + \theta \Delta \log w_t + \kappa \xi_t - \kappa(r_{t-1} - \rho) - \kappa \phi_t. \quad (2)$$

Estimation Based on Equation (1)

When wages are uncertain equation (1) cannot be estimated by OLS because $\Delta \log w_t$ is correlated with ξ_t. For example, Pistaferri (2003) writes an approximating model (also used by MaCurdy, 1981) of the form

$$\delta \xi_t = \delta \left(\log \lambda_t - E_{t-1} \log[\lambda_t] \right) \approx \sum_{j=0}^{T-t} \gamma_j \left(E_t \log[w_{t+j}] - E_{t-1} \log[w_{t+j}] \right)$$

(The coefficients γ_j are negative). If wages follow an AR(1) process

$$\log w_t = \lambda \log w_{t-1} + \zeta_t$$

then

$$E_t \log[w_{t+j}] - E_{t-1} \log[w_{t+j}] = \lambda^j \zeta_t$$

and

$$\delta \xi_t = \sum_{j=0}^{T-t} \gamma_j \lambda^j \zeta_t = \Gamma(t, T, \lambda) \zeta_t.$$

The update to $\log \lambda_t$ is some coefficient Γ (which depends on (t, T, λ)) times the wage innovation ζ_t. Note that the$\gamma_j' s$ should also depend on age and current wealth, which introduces even more heterogeneity into the coefficient Γ.

One approach to estimation is to find instruments that predict wage growth, that are orthogonal to the surprise component in wages (and therefore in $\log \lambda_t$) and that also do not enter in ΔA_t (the preference shock). MaCurdy (1981) used experience: according to the simplest Mincer model

$$\log w_t = b_1 x(t) + b_2 x(t)^2 + ...$$

where $x(t)$ is experience at time t. Since $x(t) = x(t-1) + 1$ predicted wage growth is a simple linear function of experience in year $t-1$. In fact experience works as a predictor of wage growth, though the first stage is often weak. Estimates of η based on this approach tend to be small - on the order of 0.1 to 0.3 (see MaCurdy's orginal analysis and Altonji (1986, Table 2) for a variety of estimates based on the MaCurdy approach).** A concern is that experience may have some direct effect on preferences. This, coupled with the a priori belief that η must be relatively large, has led to ongoing interest in other approaches.

**An interesting feature of Altonji's paper is that he reports the first stage equations, so you can judge the power of the instruments, though his paper was written before the weak instruments critique was well understood (and before the cluster option made it easy to account for serial correlation within the data for each person over time).

Altonji (1886) tried using consumption as a proxy for the (unobserved) marginal utility of wealth, which seems quite plausible. This is easiest to understand in the within-period separable case: then the system of interest is

$$\begin{aligned} \log h_t &= A_t + \eta \log w_t + \eta \log \lambda_t + e_{1t} \\ \log c_t &= B_t + \kappa \log \lambda_t + e_{2t} \end{aligned}$$

where I have added measurement errors e_{1t} and e_{2t}. This implies

$$\log h_t = (A_t - \frac{\eta}{\kappa} B_t) + \eta \log w_t + \frac{\eta}{\kappa} \log c_t + e_{1t} - \frac{\eta}{\kappa} e_{2t}.$$

If the$e'_{jt}s$ are really measurement errors the only remaining problems with this specification are that $\log c_t$ is correlated with e_{2t} and any unobserved components of B_t, and that $\log w_t$ is measured with error. [It is also possible, as in a static labor supply model, that the unobserved parts of A_t are correlated with $\log w_t$]. The main advantage of Altonji's approach is that we don't have to first difference – so there is a lot of variation left and many potential instruments for $\log w_t$ and $\log c_t$. Altonji (1986) used a second measure of wages (collected at the interview in the PSID, and representing the point-in-time wage on the job at the time of the interview), the mean wage observed in other years, and various demographic factors (e.g. spouse's education, parental education/income). His estimates (Altonji (1986, Table 4)) for η are between 0.1 and 0.2.

The main concern with Altonji's approach is that preferences may not be separable. If

$$\log c_t = B_t + \theta \log w_t + \kappa \log \lambda_t + e_{2t}$$

then solving for $\log \lambda_t$ and substituting into the hours equation (with a coefficient δ for $\log \lambda_t$ that is potentially different from η) leads to an hours model:

$$\log h_t = (A_t - \frac{\eta}{\kappa} B_t) + (\eta - \theta \frac{\delta}{\kappa}) \log w_t + \frac{\delta}{\kappa} \log c_t + e_{1t} - \frac{\delta}{\kappa} e_{2t}.$$

Notice that the coefficient on $\log w_t$ in this case is

$$\eta - \theta \frac{\delta}{\kappa} \approx \eta - \theta \frac{mpe}{1 + mpe}$$

using the result presented earlier that$\frac{\delta}{\kappa} \approx \frac{mpe}{1+mpe}$. Assuming $mpe \approx -0.1$, this implies that the estimate obtained using Altonji's procedure is an estimate of

$\eta + 0.11\theta$. Assuming $\theta \geq 0$ we get an upward-biased estimate for η, though the magnitude of the bias is arguably small.

Pistaferri (2003) presents an interesting addition to this literature, using information on wage growth expectations that is collected in the Bank of Italy's Survey of Household Income and Wealth (SHIW). Pistaferri assumes that individual wages follow a random walk:

$$\log w_t = \log w_{t-1} + \zeta_t$$

and adopts the assumption (presented above) that the (scaled) innovation in the log marginal utility of wealth follows

$$\delta\xi_t = \delta\left(\log \lambda_t - E_{t-1}\log[\lambda_t]\right) = \sum_{j=0}^{T-t} \gamma_j \left(E_t \log[w_{t+j}] - E_{t-1}\log[w_{t+j}]\right).$$

With the unit root assumption $E_t \log[w_{t+j}] - E_{t-1}\log[w_{t+j}] = \zeta_t$ and

$$\delta\xi_t = \sum_{j=0}^{T-t} \gamma_j \zeta_t = \Gamma\zeta_t.$$

With this substitution, equation (1) becomes:

$$\begin{aligned} \Delta \log h_t &= \Delta A_t + \eta \Delta \log w_t + \delta\xi_t - \delta(r_{t-1} - \rho) - \delta\phi_t \\ &= \Delta A_t + \eta E_{t-1}[\Delta \log w_t] + (\eta + \Gamma)\zeta_t - \delta(r_{t-1} - \rho) - \delta\phi_t \end{aligned} \tag{3}$$

(which is Pistaferri's equation (13)). In the SHIW people are asked directly their expected rate of growth of earnings over the next year. Letting $y_t = w_t h_t$, this means that we observe $E_{t-1}[\Delta \log y_t]$ in the year $t-1$ survey. This means we have to translate the labor supply model into a model of hours and earnings. Using the definition of earnings we get:

$$E_{t-1}[\Delta \log w_t] = E_{t-1}[\Delta \log y_t] - E_{t-1}[\Delta \log h_t]$$

and taking expectations of (3) and substituting we get

$$E_{t-1}[\Delta \log h_t] = \frac{1}{1+\eta}\{\Delta A_t + \eta E_{t-1}[\Delta \log y_t] - \delta(r_{t-1} - \rho) - \delta\phi_t\}.$$

Finally, if we define

$$\psi_t = \Delta \log y_t - E_{t-1}[\Delta \log y_t]$$

as the innovation in log earnings, and use the fact that $\Delta \log h_t - E_{t-1}[\Delta \log h_t] = (\eta + \Gamma)\zeta_t$ (from equation (3)) we get

$$\begin{aligned} \zeta_t &= \psi_t - (\eta + \Gamma)\zeta_t \Rightarrow \zeta_t = \frac{\psi_t}{1 + \eta + \Gamma} \\ &\Rightarrow \Delta \log h_t - E_{t-1}[\Delta \log h_t] = \frac{(\eta + \Gamma)}{1 + \eta + \Gamma}\psi_t. \end{aligned}$$

Thus we can write the labor supply equation in terms of expected earnings changes and the innovation in earnings as:

$$\Delta \log h_t = \frac{1}{1+\eta}\Delta A_t + \frac{\eta}{1+\eta}E_{t-1}[\Delta \log y_t] - \frac{\delta}{1+\eta}(r_{t-1}-\rho) - \frac{\delta}{1+\eta}\phi_t + \frac{(\eta+\Gamma)}{1+\eta+\Gamma}\psi_t \tag{4}$$

(As a final step, Pistaferri solves for ϕ_t in terms of the variance in the earnings forecast, under the assumption of log-normality, but we will leave that aside). Notice that if we observe expected and realized earnings then we can estimate this model taking $E_{t-1}[\Delta \log y_t]$ and ψ_t as observed variables. This procedure will yield estimates for η and Γ. Morever, Pistaferri uses the relatively short time period in his panel to get variation in the real interest rate, providing an estimate of δ. His estimates are

$$\begin{aligned} \eta &= 0.70\ (0.09) \\ \Gamma &= -0.20\ (0.09) \\ \delta &= 0.59\ \ (0.29) \end{aligned}$$

which look pretty large in magnitude. As discussed in his paper, one (plausible) explanation for this is that the true wage process is more like:

$$\begin{aligned} \log w_t &= z_t + \epsilon_t \quad \text{where} \\ z_t &= z_{t-1} + \zeta_t \end{aligned}$$

and ϵ_t and ζ_t are i.i.d. This says wages are a combination of a component with a unit root (the permanent wage component) and a serially uncorrelated component (the transitory component), and implies that

$$\log w_t = \log w_{t-1} + \zeta_t + \epsilon_t - \epsilon_{t-1}$$

which is an ARIMA(0,1,1) model. Now the innovation in log wages is $\zeta_t + \epsilon_t$, but only the permanent part is expected to persist, so holding constant the (observed) innovation in current wages (or, in Pistaferri's case, earnings) the apparent response in labor supply is bigger than it would be if the entire wage innovation persisted (which is what is being assumed in equation (3). Preferences are being credited for a labor supply response that is due in part to the temporary nature of the wage innovation, so there is an upward bias in the estimate of η.

Some simple evidence on the right statistical model for wages is presented in Card (1994): there I used data on a sample of male household heads from the PSID observed continuously over an 8-year period to fit a model of the form

$$\log w_{it} = \omega_i + v_t + u_{it} + \mu_{it}$$

where

$$u_{it} = \alpha u_{it-1} + \xi_{it}$$

and ξ_{it} and μ_{it} are mutually uncorrelated, the innovations in the AR(1) component are uncorrelated (but allowed to have differenent variances in different years), and ξ_{it} and μ_{it} are uncorrelated with the random effect. Pistaferri (effectively) assumes $\alpha = 1$ and $var(\mu_{it}) = 0$. The estimates are reported in Table 2.3 of my paper and show that: (1) such a model fits relatively well; (2) $\alpha \approx 0.9$; (3) about 50% of the variance in wages is attributed to ω_i, 16% to the transitory component μ_{it} and 34% to the serially correlated component u_{it}. Arguably, Pistaferri's assumption of a pure random walk model for wages is too restrictive.

Extensive margin

Many labor supply estimates ignore the extensive margin – workers who don't work for a year are dropped. This is potentially important for understanding aggregate movements in hours because:

(a) there are substantial numbers of people who move in/out of employment

(b) the elasticity of participation w.r.t. wages can be relatively high, even if η is small.

A simple approach to this problem is to go back to the first order conditions defining the Frisch labor supply/consumption choices, and define a reservation wage in each period (or more generally a selection equation determining whether the individual works in period t. For an example of this see J. Kimmel and T. Kniesner, New Evidence on Labor Supply: Employment vs Hours Elasticities by Sex and Marital Status." Journal of Monetary Econ 42 (1998).

Manoli and Weber (2011) is a very recent attempt to look at one of the important extensive margins: variation in the length of time people work. This paper uses an RD design to study the effects of a benefit that is paid to workers who retire after certain tenure milestones. Since workers start jobs at different ages, there is a smooth distribution of people across the tenure distribution at different ages, and Manoli and Weber find strong evidence that some workers appear to delay retirement to get the benefits. However, the implied responsiveness is relatively small (elasticities on the order of 0.3 or smaller). An earlier paper by Krueger and Pischke (Journal of Labor Economics, 1992) looked at the effect of a revision in the indexing formula for Social Security, which sharply lowered the benefits to people born in 1917-1921 relative to those born 1915-1916 (who got very high benefits as a result of an error in the indexing formula). As shown in their figures:

(1) people born in 1918-20 suffered a sharp drop in benefits to earliest possible retirement (age 62)

(2) people born in 1914-1916 had unusually high incentives to delay retirement to age 68

(3) BUT LFPR's trended pretty smoothly down across these cohorts

(A recent paper by Alex Gelber re-investigates the notch and concludes that there was some labor supply effect). The extensive margin (EM) is an area of active research interest. One (serious) difficulty with studying EM responses is that wages are only observed for workers. So it becomes necessary to impute shadow wages (or make other assumptions) to correlate changes in participation with changes in wages.

Is Labor Supply Really a Worker Choice?

Ham and Reilly (AER 2002) ask whether information from the demand side affect hours choices, controlling for wages and other factors. This could arise in a contract setting where workers agree to work for some wage and allow the employer to specify hours. In a simple neoclassical model of the labor market the two sides of the market are separated by the wage:

$$\begin{aligned} h &= h^d(w,x) \\ &= h^s(w,y) \end{aligned}$$

where h^d and h^s are the demand and supply functions for hours (by some group of workers), and x and y represent demand and supply shocks. In this class of models, any effect of demand shocks works through w: the two sides of the market both make independent decisions, taking w as given. Thus, a test of the standard model is to fit the supply function and include x directly in the estimating equation. This requires that there be instruments for w in addition to the demand shock variables - so HR's test one interpretation of their test is that they are testing whether one set of demand shock variables affect supply, when wages are instrumented with other variables.

Formally, H-R consider two specifications. Their first set of models use first differenced labor supply models of the type described above:

$$\Delta \log h_t = \Delta A_t + \eta \Delta \log w_t + \delta \xi_t - \delta(r_{t-1} - \rho) - \delta \phi_t. \tag{5}$$

The include an extra set of explanatory variables representing the changes in the unemployment rates for the industry and occupation that the agent was working in in the base year $(\Delta UR^{ind}, \Delta UR^{occ})$. These are treated as potentially endogenous because they may reflect the news shocks incorporated in the innovation in the log marginal utility of consumption, ξ_t. They also present models with future wage changes $(\Delta \log w_{t+1})$ included on the right hand side, as a potential way to incorporate non-separable preferences (basically, if people forsee high wages ahead they may work more or less this period) See Table 1 of their paper.

H-R's second specification builds on Altonji's idea of controlling directly for consumption. Here the baseline specification is:

$$\log h_t = (A_t - \frac{\eta}{\kappa} B_t) + (\eta - \theta \frac{\delta}{\kappa}) \log w_t + \frac{\delta}{\kappa} \log c_t + e_{1t} - \frac{\delta}{\kappa} e_{2t}.$$

In this case they augment the model with (UR^{ind}, UR^{occ}), and include specifications with future wages. See tables 2 and 3, which use PSID and CES data.

Their key finding is that predictable movements in ΔUR^{ind} and ΔUR^{occ} (or in the levels of UR^{ind}, UR^{occ}), have a lot of explanatory power. They interpret this as evidence that wages are not translating all the necessary information about the state of the demand side to the worker.

Labor Supply in the Very Short Term

Farber's 2014 "taxi" paper looks at the detailed trip-level data from NYC cabs, focusing on how taxi driver's labor supply in a given hour is affected by their wage. One motivation for the paper is the idea – proposed in a well known paper by Camerer, Babcock, Lowenstein, and Thaler (QJE, 1997) – that people do not choose hours according to standard labor supply models: instead, they have a target level of earnings in mind, and if they get to that level of earnings, they stop working. This gives rise to a very perverse -1 elasticity of labor supply in the short run! The target earnings idea comes from the Tversky and Kahneman (1991) idea of loss aversion. Farber implements a loss aversion component in the labor supply choice by assuming utility in a given day is:

$$\begin{aligned} U(y,h) &= (1+\alpha)(Y-T) - \frac{\theta}{1+\upsilon}h^{1+\nu} \; Y < T \\ &= (1-\alpha)(Y-T) - \frac{\theta}{1+\upsilon}h^{1+\nu} \; Y \geq T \end{aligned}$$

where T is an earnings target (the reference point). Note that if $\alpha > 0$ the individual's MU of income drops from $1+\alpha$ to $1-\alpha$) as earnings pass through the reference point. If $\alpha = 0$ the individual has a constant MU of income – which is appropriate for a short run labor supply problem. The $\pm\alpha(T-T)$ part of U is called the "gain loss" utility component, and is assumed to be added to the "regular" utility function $Y - \frac{\theta}{1+\upsilon}h^{1+\nu}$, which is called "consumption utility". Driver's indifference curves have a kink, as shown in the figure at the end. The MRS is

$$\begin{aligned} MRS(Y,h) &= \frac{\theta h^{\nu}}{1+\alpha} \quad Y < T \\ &= \frac{\theta h^{\nu}}{1-\alpha} \quad Y \geq T \end{aligned}$$

The choice of hours given a parametric wage w is as follows:

$$\begin{aligned} h &= \left(\frac{(1+\alpha)w}{\theta}\right)^{1/\nu} \quad w < w^* \\ &= T/w \quad w^* < w < w^{**} \\ &= \left(\frac{(1-\alpha)w}{\theta}\right)^{1/\nu} \quad w > w^{**} \end{aligned}$$

where:

$$\begin{aligned} w^* &= \left(\frac{\theta}{1+\alpha}\right)^{\frac{1}{1+\nu}} T^{\frac{\nu}{1+\nu}} \\ w^{**} &= \left(\frac{\theta}{1-\alpha}\right)^{\frac{1}{1+\nu}} T^{\frac{\nu}{1+\nu}} \end{aligned}$$

Note that if $\alpha = 0$ the driver has hours function:

$$h = \left(\frac{w}{\theta}\right)^{1/\nu}$$

with (intertemporal) elasticity $\frac{1}{\nu}$. But if $\alpha > 0$ there is a range of variation in w such that hours are decreasing in the wage (with elasticity -1). Next, Farber invokes the "rational reference point" assumption of Koszegi and Rabin (2006), which he translates into the assumption that $log\, T = E[log(wh)]$ when hours are determined by consumption utility only:

$$log\, T = \frac{1+\nu}{\nu} E[log\, w] - \frac{1}{\nu}\theta.$$

Farber interprets this as the reference point for earnings during an hour with a given expected wage (so the reference point is higher on a typically high-wage hour, like Friday at 6:00 pm). This is very important point that comes from the KR idea: you only get reference point behavior from *unanticipated variation.*

This leads to a really nice pair of expressions for the logs of w^* and w^{**} :

$$\begin{aligned} log\, w^* &= E[log\, w] - \frac{1}{1+\nu} log(1+\alpha) \\ log\, w^{**} &= E[log\, w] - \frac{1}{1+\nu} log(1-\alpha) \end{aligned}$$

which means that reference dependence behavior only arises when the wage is in an interval of from $\frac{1}{1+\nu} log(1+\alpha)$ below the mean to $\frac{1}{1+\nu} log(\frac{1}{1-\alpha})$ above the mean. Farber notes that in the behavioral literature, people believe that $\frac{1+\alpha}{1-\alpha}$ is a number like 1.5 to 2.5 (this is the so-called coefficient of loss aversion). In this case, α is in the interval $[0.2, 0.43]$. Assuming $1/\nu = 0.5$ this means that the range around the mean is something like 6% below the mean to 7% above (if $\alpha = .2$), or a wider range of 12% below the mean to 19% above the mean (if $\alpha = .43$). Farber argues this means that most wage variation is in the "reference dependence range", because (as he shows) the wage is quite predictable.

Farber presents some very nice labor supply estimates based on how many hours a driver works per shift (there are day shifts, night shifts and some other miscellaneous kinds). Table 4 presents OLS models, Table 5 presents IV estimates where the instrument is the mean earnings per hour of other drivers in the same shift (which has a lot of predictive power). His IV estimates range from .4 to .9 when he includes fixed effects. These may seem large, but at the daily level we expect pretty large elasticities relative to the year.

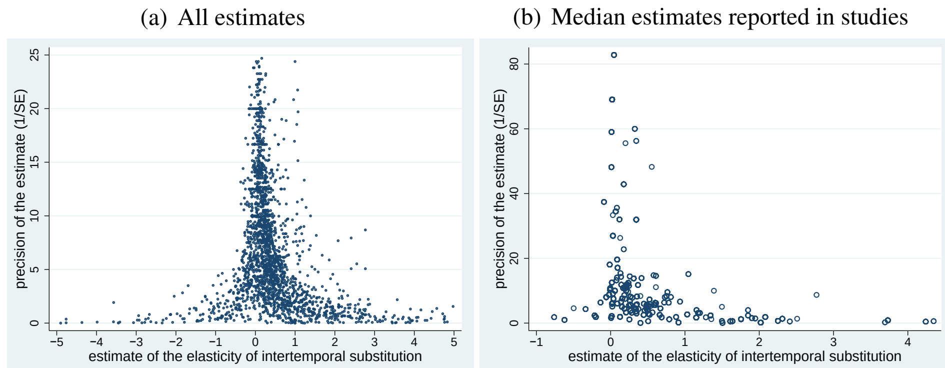


FIGURE 1. Negative estimates are underreported. In the absence of selective reporting the funnel should be symmetrical around the most precise estimates. I exclude estimates with extreme magnitude or precision from the figure but include all in the regressions.

and interesting result, not from the preferences of editors and referees—although the priors may be formed based on what results are publishable.

The Online Appendix shows four additional robustness checks. First, I test whether my results change if I only consider estimates of the EIS published in finance journals. Different values of the elasticity are needed to explain different facts in economics and finance; perhaps the two streams of literature differ in the extent of selective reporting. Nevertheless, my results suggest that the estimates of the EIS reported in finance are very similar to those reported in economics. Second, some studies report asymmetric confidence intervals for the estimates, which means that the ratio of the point estimate to the standard error is not t-distributed. I follow the advice of Stanley (2001, p. 135) to "better err on the side of inclusion" in meta-analysis, compute approximate standard errors for the estimates (based on the simplifying assumption of normal distribution), and include the estimates. Exclusion of these estimates does not change my results. Third, I exclude the three studies from my sample that use the long-run risks model to estimate the EIS, but the results are again similar. Finally, my results do not change qualitatively if I exclude estimates with bootstrapped confidence intervals.

It is difficult to say at this point which of the two potential sources of selective reporting drives the results in Table 1. A graphical inspection of the data suggests that both sources play a role. Figure 1 shows the so-called funnel plot, which is often used in medical meta-analyses to detect selective reporting (Egger et al. 1997). The horizontal axis measures the magnitude of the estimate of the EIS, while the vertical axis measures the estimate's precision, the inverse of the standard error. The most precise estimates should be concentrated close to the underlying effect at the top of the figure, while the imprecise estimates at the bottom should be more dispersed. The t-distribution of the ratio of point estimates to their standard errors ensures that in the absence of selective reporting the figure is symmetrical, forming an inverted funnel.

Panel (a) of Figure 1 shows the funnel plot with all estimates of the EIS. The most precise estimates are positive but small. Researchers report negative estimates

TABLE 1

SIMULTANEOUS-EQUATION ESTIMATION OF FIRST-DIFFERENCED LABOR-SUPPLY-EQUATION ESTIMATES OF THE INTERTEMPORAL SUBSTITUTION ELASTICITY

Estimation Procedure	D(Log Wage)	D(Log Earning)*	Intercept	Average of Year Dummies
2SLS	.23 (2.42)	. . .	−.009 (4.02)	. . .
3SLS	.14 (1.97)	. . .	−.008 (4.26)	. . .
2SLS	. . .	.35 (2.22)	−.006 (4.26)	. . .
3SLS	. . .	.25 (2.63)	−.006 (5.18)	. . .
2SLS	.15 (.98)	. . .	. . .	−.008
3SLS	.10 (.80)	. . .	. . .	−.008
2SLS	. . .	.45 (1.54)	. . .	−.007
3SLS	. . .	.30 (1.67)	. . .	−.007

NOTE.—Absolute values of t-statistics are in parentheses.

* The estimates and t-statistics reported in the "Log Earning" column are for δ; they are computed using the coefficient on earnings, denoted ψ, and its t-statistics. We have $\hat{\delta} = \hat{\psi}/(1 - \hat{\psi})$. To convert the t-statistics reported for ψ to those for δ requires division by the quantity $(d\delta/\partial\psi) = (1 + \delta)^2$ evaluated at $\delta = \hat{\delta}$.

higher estimate for δ in all cases, but these differences are small relative to their standard errors. The estimates of the intercepts indicate that the real rate of interest exceeds the rate of time preference on average by about 2–4 percentage points.

A comparison of these results with others in the literature is difficult since most studies use cross-section data for their empirical analysis where differences in lifetime wage paths are the primary source of wage variation across observations. As a consequence, they do not estimate the intertemporal substitution elasticity. Estimating equation (13) using cross-section data is complicated by the presence of individual effects, F_i. As discussed above, economic theory implies that F_i is correlated with a consumer's wages and all of his other characteristics. Therefore, it is not reasonable to assume that F_i is a "random factor" uncorrelated with explanatory variables. One, then, cannot directly use observations on individuals from a cross section to estimate the parameters of (13) even if one uses simultaneous equation estimation procedures. Such procedures implicitly treat individual effects as random variables, and this leads to inconsistent parameter estimates. To estimate the intertemporal elasticity using cross-section data, one requires a specification of the labor-supply equation where variation in wages reflects evolutionary wage change.

TABLE 2

First-Difference Equations for Labor Supply, Instrumental Variables for Dw_t^* and Dw_t^{**}: Human Capital, Family Background, Year Dummies (Dependent Variable = Dn_t^*)

Explanatory Variable	Full Sample (1)	(2)	(3)	(4)	Sample with Data on Dw_t^*, Dw_t^{**} (5)	(6)	(7)	(8)	(9)	(10)
Constant	−.0147	−.0025	−.0180	.0035	−.0092	−.0173	−.0491	−.0165	−.0439	−.0471
	(.1258)	(.0180)	(.0154)	(.0338)	(.0054)	(.0223)	(.0484)	(.0052)	(.0238)	(.108)
Dw_t^*	.2817	.2317	.4782	.2666	.0834	.1001	.3057	...	...	...
	(.1258)	(.1434)	(.3255)	(.4155)	(.1587)	(.1661)	(.654)			
Dw_t^{**}	...	...	...	...	...	...	...	.4508	.5425	.2937
								(.1893)	(.2079)	(1.78)
Age	...	−.0002	...	−.0004	...	.0002	−.0001	...	.0006	.0000
		(.0004)		(.0005)		(.0005)	(.00005)		(.0005)	(.001)
Year dummies?	no	no	yes	yes	no	no	yes	no	no	yes
Standard error	.237	.2293	.2720	.2347	.2066	.2084	.2333	.2074	.2114	.2013
F-ratio	5.01	2.90	1.48	1.88	.28	.21	3.32	5.67	3.43	4.43
R^2	.0005	.0006	.0019	.0026	.0001	.0001	.0099	.0014	.0017	.0132
Observations	10,036	10,036	10,036	10,036	3,996	3,996	3,996	3,996	3,996	3,996

Note.—Standard errors are in parentheses. The first-stage equations for Dw_t^* and Dw_t^{**} are reported in table A1.

TABLE 4

LABOR SUPPLY ESTIMATES USING FOOD CONSUMPTION AS A PROXY FOR λ_t (See Eq. [14])

	ESTIMATION METHOD			
	OLS (1)	OLS (Reduced Form)* (2)	IV (3)	IV (4)
Intercept	7.528	7.416	8.386	7.995
	(.155)	(.157)	(.644)	(.359)
$w_t^{*\dagger}$	−.1126	...	.1721	.0943
	(.014)		(.119)	(.057)
$c_t^{*\ddagger}$	.0788	...	−.5341	−.2972
	(.015)		(.386)	(.202)
w_t^{**}	...	−.019	...	...
		(.025)		
w_i^{**}	...	−.031	...	...
		(.032)		
Black	−.021	−.0028	−.0449	−.0315
	(.015)	(.015)	(.036)	(.025)
Health	−.079	−.0683	−.0702	−.0703
	(.015)	(.015)	(.020)	(.0168)
Age	.008	.0089	.0220	.0177
	(.007)	(.007)	(.013)	(.0101)
Age^2	−.0001	−.0001	−.0003	−.0239
	(.0001)	(.0001)	(.0002)	(.0116)
Size of family unit	−.0187	−.0060	.0424	.0192
	(.006)	(.006)	(.037)	(.0194)
Number of children	.0025	.0020	−.0171	−.0088
under 6	(.008)	(.008)	(.018)	(.0120)
Number of children	.0110	.0074	.0424	.0029
in family	(.006)	(.006)	(.011)	(.0079)
SMSA	−.0079	−.0198	.0026	−.0033
	(.015)	(.014)	(.025)	(.0197)
City ≥ 500,000	.0100	.0021	.0384	.0250
	(.017)	(.016)	(.034)	(.0246)
South	.0198	.0527	−.0116	.0038
	(.019)	(.021)	(.042)	(.0300)
West	.0284	.0417	.0025	.0060
	(.022)	(.024)	(.035)	(.0287)
North Central	.0327	.0542	−.0151	.0053
	(.019)	(.024)	(.045)	(.0299)
Year dummies?	yes	yes	yes	yes
Standard error	.217	.220	.292	.250
R^2	.0809	.0544	.0258	.034

NOTE.—Observations = 4,367. The standard errors (in parentheses) have been corrected for correlation over time in the observations on a given individual (see n. 36).

* The reduced-form equation also contains wife's education and dummy variables for whether the individual's parents were average or rich while he was growing up. The coefficients on these variables are .0112, .0086, and .0044. The standard errors are .0028, .0136, and .0224.

† Treated as endogenous in cols. 3 and 4. The instrumental variables for col. 3 are w_t^{**}, w_i^{**} plus the other variables in the labor supply equation. The instrumental variables for col. 4 are w_t^{**} and w_i^{**} only (see table A1, cols. 9 and 10).

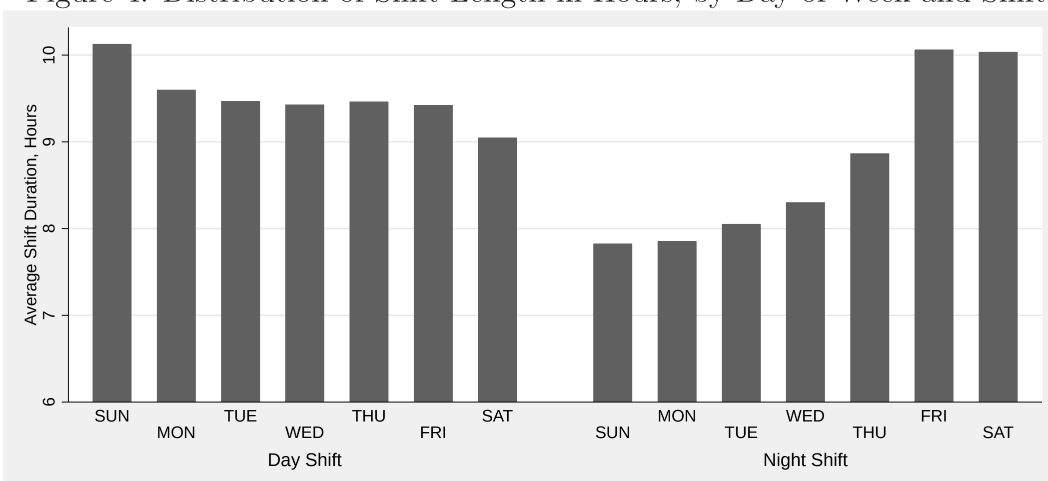
‡ Treated as endogenous in cols. 3 and 4. The first-stage equation is in table A1, col. 11.

Table 3: Mean Hours, Income, and Average Hourly Earnings. By Shift

Shift	Hours	Income	Wage	# Shifts
Day	9.48	248.41	26.39	2247417
	(3.97)	(106.03)	(5.75)	
Night	8.78	262.03	30.13	2137499
	(3.05)	(93.06)	(6.55)	
Unassigned	8.30	228.12	28.26	662427
	(5.33)	(138.46)	(8.47)	

Note: Day shifts start between 4AM and 9:59AM. Night Shifts start between 2PM and 7:59PM. Shifts starting at other times are Unassigned. Standard deviations are in parentheses.

Figure 4: Distribution of Shift Length in Hours, by Day of Week and Shift



Labor supply and earnings have distinct patterns over the week by shift. Figure 4 contains plots of the average shift length by day of week for day and night shifts.[23] Day shift drivers work longest on Sundays, with average hours declining from about 10 hours on Sunday to 9 hours on Saturday. In contrast night shift drivers work their shortest days on Sunday, with average hours increasing sharply from about 8 hours per shift early in the week to about 10 hours per shift on Friday and Saturday. Day shift drivers work longer hours than night shift drivers on all days but Friday and Saturday. Total income per shift, shown in figure 5, generally follows hours: Day shift drivers earn the most on Sunday and the least on Saturday, and night shift drivers earn the least on Sunday and the most on Friday and

[23] Some shifts span days of the week. I assign each shift to a particular day of the week based on the date of the first trip in the shift.

Table 4: Wage Elasticity, OLS Regression of Average Log Daily Hours, by Shift

Model	Controls	Driver F.E's	Elasticity All Shifts	Elasticity Day Shifts	Elasticity Night Shifts	Elasticity Other Shifts
(1)	No	No	0.0159	0.0485	-0.0017	0.0738
			(0.0154)	(0.0177)	(0.0169)	(0.0220)
(2)	Yes	No	-0.0034	0.0505	-0.0784	0.0606
			(0.0177)	(0.0203)	(0.0210)	(0.0240)
(3)	Yes	Yes	-0.1002	-0.0615	-0.1487	-0.0501
			(0.0089)	(0.0109)	(0.0077)	(0.0138)

Note: Each estimated elasticity is from a separate OLS regression. "Elasticity" is the estimated coefficient of log average hourly earnings from a regression of log shift duration. "Controls" include indicators for day of week (6), calendar week (51), year (4), the period subsequent to the September 4, 2012 fare increase (1), and major holiday (1). Estimated using sample of 5,047,343 shifts for 8,802 drivers from 2009-2013. Sample sizes are listed in table 3. Robust standard errors clustered by driver are in parentheses.

mated elasticity is negative and statistically significant but relatively small at -0.1. For day shifts, the estimated elasticities are small and positive but statistically significant in the first two specifications. When driver fixed effects are included, the estimated elasticity is again negative and statistically significant though small. The pattern for night shifts is that the elasticities are significantly negative when the controls are added. The estimated elasticity when driver fixed effects are included is more negative than for the day shift. The estimates for the unclassified (other) shifts are very close to those for day shifts.

While I do find some negative elasticities, none approach minus one as suggested by reference dependence. My elasticities are much smaller than those found using OLS by CBLT or Farber (2005), which may reflect a lower level of measurement error in my administrative data compared with the data transcribed from trip sheets used in the earlier work.

Although the administrative data may have less measurement error than data derived from the paper trip sheets, it is not error free. Simple consistency checks of the data show more than a few instances of trips ending before they start and new trips starting before the previous trip ends.[26] Additionally, as I mentioned earlier, my income data do not include tips, which surely vary across trips as a proportion of fares (Haggag and Paci, 2014). On this basis, it makes sense to estimate the model using an instrument for average hourly earnings.

In the spirit of CBLT, I use the average across other drivers of log average hourly earnings on the day each shift started. To avoid problems using an instrument derived from the

[26] I used some simple algorithms to adjust the data to eliminate these inconsistencies and serious outliers.

Table 5: Wage Elasticity, IV Regression of Average Log Daily Hours, by Shift

Model	Controls	Driver F.E's	Elasticity All Shifts	Elasticity Day Shifts	Elasticity Night Shifts	Elasticity Other Shifts
(1)	No	No	0.2288	0.0202	0.3484	0.2913
			(0.0101)	(0.0134)	(0.0117)	(0.0306)
(2)	Yes	No	0.5709	0.3683	0.6182	0.9383
			(0.0100)	(0.0119)	(0.0132)	(0.0329)
(3)	Yes	Yes	0.5890	0.3672	0.6344	0.8751
			(0.0099)	(0.0112)	(0.0124)	(0.0281)

Note: Each estimated elasticity is from a separate IV regression. The instrument for average hourly earnings is the average of average hourly earnings for a non-overlapping sample of drivers on the same day. "Elasticity" is the estimated coefficient of log average hourly earnings from a regression of log shift duration. "Controls" include indicators for day of week (6), calendar week (51), year (4), the period subsequent to the September 4, 2012 fare increase (1), and major holiday (1). Estimated using sample of 5,047,343 shifts for 8,802 drivers from 2009-2013. Sample sizes are listed in table 3. Robust standard errors clustered by driver are in parentheses.

dependent variable in the estimation sample, I use a non-overlapping randomly selected 2/15 subset of the drivers to generate the instruments.[27] The average of log average hourly earnings of shifts starting on date t in the non-overlapping sample ($\overline{\ell nW_t}$) serves as the instrument for the log average hourly earnings for driver i in my estimation sample for shifts that start on date t (ℓnW_{it}).[28]

The IV estimates of the labor supply elasticity are contained in Table 5. The results are striking in comparison with the OLS estimates in table 4. The estimated elasticities are substantially positive and strongly statistically significant. Adding the control variables raises the estimated elasticity for each sample, but controlling for driver fixed effects does not have much effect. The estimated elasticity on the day shift is about 0.36 while the elasticity on the night shift is about 0.62. The larger elasticity for the night shift is consistent with the observation that drivers on a night shift are more likely than drivers on a day shift to be able to adjust hours mid-shift in response to new information regarding earnings opportunities. Interestingly, the elasticity is even larger on unclassified shifts. It may be that these other shifts are less likely to be worked by lease drivers and more likely to be worked by owner-operators who have more flexibility in selecting hours.

[27] This sample contains 115,733,041 trips on 5,012,244 shifts for 8,768 drivers.

[28] While I do not present the first stage results, the instrument is very strong. The first-stage t-statistic on the instrument is generally greater than 100, and the coefficient on the instrument in the first stage is generally close to one.

Figure 8: Wage Elasticity of Labor Supply, IV Estimates by Experience