Supplemental Appendix for Understanding Women's Wage Growth using Indirect Inference with Importance Sampling

Robert M. Sauer*and Christopher Taber†

March 16, 2021

Appendix A: Asymptotic Results

A.1 Consistency

We first show consistency of the estimator.

Define

$$\begin{aligned}\widehat{G}(\beta) \equiv& \frac{1}{N}\sum_{i=1}^{N} g(X_i, Y_i, \beta)\\ \widetilde{G}_h(\theta,\beta) \equiv& \frac{1}{S}\sum_{s=1}^{S} \frac{\ell(\Upsilon_{hs}; X_{hs}, \theta)}{\ell_0(\Upsilon_{hs}; X_{hs})} g(X_{hs}, Y_{hs}; \beta).\end{aligned}$$

We need the following assumptions:

Assumption A1. *$\widehat{G}(\beta)$ converges uniformly in probability to $G(\theta_0, \beta)$.*

The key aspect of this is that g is well behaved so that this convergence is uniform. Note that we are being general enough not to require the expressions to be differential in the underlying function θ but are assuming that the auxiliary model that we estimate on the actual data is simple.

The next are standard regularity assumptions as well as a condition for identification of $B(\theta)$.

Assumption A2. *Θ and $\mathcal{B}$ are compact.*

Assumption A3. *For each $\theta \in \Theta$, $B(\theta)$ is a singleton. F,g, and B are continuous.*

*Department of Economics, Royal Holloway, University of London. Email: robertmsauer@protonmail.com
†Department of Economics, University of Wisconsin. Email: taberchris@gmail.com

Presumably one could relax the assumption of point identification of $B(\theta)$ allowing this to be a set and modify the objective function so that the set $B(\widehat{\theta})$ is close to the set $\widehat{\beta}$. This seems straight forward, but we do not know of an empirical researcher that has done this, so we focus on the point identified case.

Next we have the identification assumption for θ :

Assumption A4. *If $\theta_1 \neq \theta_2$ then*

$$B(\theta_1) \neq B(\theta_2).$$

If this assumption were relaxed we would no longer obtain point identification, but would instead obtain set identification.

Assumption A5. *We can write $\Upsilon_{hs} = \left\{\Upsilon_{hs}^d, \Upsilon_{hs}^c\right\}$ where Υ_{hs}^d is discrete taking on values $\Upsilon_{(1)}^d, ..., \Upsilon_{(K_\Upsilon)}^d$ and Υ_{hs}^c is continuous. For every $\theta \in \Theta$, the support of Υ_{hs} generated by $\ell(\Upsilon_{hs}; X_{hs}, \theta)$ is a subset of or equal to the support of Υ_{hs} generated by $\ell_0(\Upsilon_{hs}; X_{hs})$.*

This assumption makes the likelihood function easy to write down. We could easily extend the results to accommodate other specific cases.

Assumption A6. *For each simulation $h = 1, .., H$,*

$$\frac{1}{S}\sum_{s=1}^{S} \frac{\ell(\Upsilon_{hs}; X_{hs}, \theta)}{\ell_0(\Upsilon_{hs}; X_{hs})} g(X_{hs}, Y_{hs}; \beta))$$

converges uniformly in probability over β and θ to

$$E_s\left(\frac{\ell(\Upsilon_{hs}; X_{hs}, \theta)}{\ell_0(\Upsilon_{hs}; X_{hs})} g(X_{hs}, Y_{hs}; \beta); \Xi_0, \ell_0\right)$$

where E_s represents the expected value when the data are generated from a simulation in which X_{hs} is drawn from Ξ_0 and Υ_{hs} is drawn from $\ell_0(\Upsilon; X_{hs})$.

There are really three separate aspects of this assumption. First that we are drawing Υ_{hs} from $\ell_0(\Upsilon; X_{hs})$ which is a fundamental part of the importance weight sampling approach. The second is that convergence is uniform which is standard. The third aspect reflects how X_{hs} is chosen. For calculating the asymptotic distribution we will need to put more structure on this, but here we just require that asymptotically it is drawn from the true distribution. There are many ways to do this, and we will discuss this in the next section.

Theorem 1. *Under Assumptions A1-A6, $\widehat{\theta}$ converges in probability to θ.*

Proof. We verify the four conditions for consistency from Newey and McFadden Theorem 2.1.

Following their notation we define

$$Q_0(\theta) \equiv -\left(B(\theta) - B(\theta_0)\right)' \Omega \left(B(\theta) - B(\theta_0)\right).$$

Their first assumption is that $Q_0(\theta)$ is maximized at θ_0. This follows since it is negative at any other value and zero when evaluated at θ_0 by Assumption A4.

Their second assumption is that Θ is compact which we assume directly in Assumption A2.

Their third assumption is that Q_0 is continuous which follows directly from Assumption A3.

Finally we need that

$$-\left(\widetilde{B}(\theta)-\widehat{\beta}\right)'\Omega\left(\widetilde{B}(\theta))-\widehat{\beta}\right)$$

converges uniformly to Q_0.

We know that given assumptions A1,A2, and A6 the standard argument for consistency of M-estimators gives

$$\widehat{\beta}\xrightarrow{p}B(\theta_0).$$

Thus, what remains is that we need to show that $\widetilde{B}(\theta)$ converges uniformly to $B(\theta)$.

First note that when Υ_{hs} is simulated from $\ell_0(\Upsilon_{hs};X_{hs})$

$$\begin{aligned}
E_s\left(\frac{\ell(\Upsilon_{hs};X_{hs},\theta)}{\ell_0(\Upsilon_{hs};X_{hs})}g(X_{hs},Y_{hs};\beta)\right)=&\int\int\sum_{j=1}^{K_\Upsilon}\frac{\ell(\Upsilon^d_{(j)},\Upsilon^c;X_{hs},\theta)}{\ell_0(\Upsilon^d_{(j)},\Upsilon^c;X_{hs})}g(X_{hs},Y_{hs};\beta)\ell_0(\Upsilon^d_{(j)},\Upsilon^c;X_{hs})d\Upsilon^c d\Xi_0(x)\\
=&\int\int g(X_{hs},Y_{hs};\beta)\ell(\Upsilon^d_{(j)},\Upsilon^c;X_{hs},\theta)d\Upsilon^c d\Xi_0(x)\\
=&G(\theta,\beta)
\end{aligned}$$

where E_s is the expected value from the simulator.

For any $\varepsilon>0$, the following three inequalities hold with probability approaching 1.

$$\begin{aligned}
\sup_{\theta\in\Theta}\Bigg[&\frac{1}{H}\sum_{h=1}^{H}F\left(\frac{1}{S}\sum_{s=1}^{S}\frac{\ell(\Upsilon_{hs};X_{hs},\theta)}{\ell_0(\Upsilon_{hs};X_{hs})}g(X_{hs},Y_{hs};\widetilde{B}(\theta)),\widetilde{B}(\theta)\right)\\
&-\frac{1}{H}\sum_{h=1}^{H}F\left(\frac{1}{S}\sum_{s=1}^{S}\frac{\ell(\Upsilon_{hs};X_{hs},\theta)}{\ell_0(\Upsilon_{hs};X_{hs})}g(X_{hs},Y_{hs};B(\theta)),B\left(\theta\right)\right)\Bigg]<\frac{\varepsilon}{3}
\end{aligned}$$

$$\sup_{\theta\in\Theta}\left[\frac{1}{H}\sum_{h=1}^{H}argmin_\beta F\left(\frac{1}{S}\sum_{s=1}^{S}\frac{\ell(\Upsilon_{hs};X_{hs},\theta)}{\ell_0(\Upsilon_{hs};X_{hs})}g(X_{hs},Y_{hs};\widetilde{B}(\theta)),\widetilde{B}(\theta)\right)-F\left(G(\Xi_0,\theta,\widetilde{B}(\theta)),\widetilde{B}(\theta)\right)\right]<\frac{\varepsilon}{3}$$

and

$$\sup_{\theta\in\Theta}\left[F\left(G(\theta,B(\theta))\right)-\frac{1}{H}\sum_{h=1}^{H}argmin_\beta F\left(\frac{1}{S}\sum_{s=1}^{S}\frac{\ell(\Upsilon_{hs};X_{hs},\theta)}{\ell_0(\Upsilon_{hs};X_{hs})}g(X_{hs},Y_{hs};B(\theta)),B(\theta)\right)\right]<\frac{\varepsilon}{3}.$$

The first one comes from the fact that $\widetilde{B}(\theta)$ maximizes the objective function and the second two come from assumption A6.

So with probability approaching one

$$
\begin{aligned}
&\sup_{\theta\in\Theta}\left[F\left(G(\theta,B(\theta))\right),B(\theta)\right)-F\left(G(\theta,\widetilde{B}(\theta))),\widetilde{B}(\theta)\right)\Big] \\
\leq &\sup_{\theta\in\Theta}\left[F\left(G(\theta,B(\theta)),B(\theta)\right)-\frac{1}{H}\sum_{h=1}^{H}argmin_{\beta}F\left(\frac{1}{S}\sum_{s=1}^{S}\frac{\ell(\Upsilon_{hs};X_{hs},\theta)}{\ell_0(\Upsilon_{hs};X_{hs})}g(X_{hs},Y_{hs};B(\theta)),B(\theta)\right)\right] \\
&+\sup_{\theta\in\Theta}\left[\frac{1}{H}\sum_{h=1}^{H}argmin_{\beta}F\left(\frac{1}{S}\sum_{s=1}^{S}\frac{\ell(\Upsilon_{hs};X_{hs},\theta)}{\ell_0(\Upsilon_{hs};X_{hs})}g(X_{hs},Y_{hs};B(\theta)),B(\theta)\right)\right. \\
&\qquad\left.-\frac{1}{H}\sum_{h=1}^{H}argmin_{\beta}F\left(\frac{1}{S}\sum_{s=1}^{S}\frac{\ell(\Upsilon_{hs};X_{hs},\theta)}{\ell_0(\Upsilon_{hs};X_{hs})}g(X_{hs},Y_{hs};\widetilde{B}(\theta)),\widetilde{B}(\theta)\right)\right] \\
&+\sup_{\theta\in\Theta}\left[\frac{1}{H}\sum_{h=1}^{H}argmin_{\beta}F\left(\frac{1}{S}\sum_{s=1}^{S}\frac{\ell(\Upsilon_{hs};X_{hs},\theta)}{\ell_0(\Upsilon_{hs};X_{hs})}g(X_{hs},Y_{hs};\widetilde{B}(\theta)),\widetilde{B}(\theta)\right)-F\left(G(\theta,\widetilde{B}(\theta)),\widetilde{B}(\theta)\right)\right] \\
<&\varepsilon.
\end{aligned}
$$

Since F and G are continuous and Θ and $\mathcal{B}$ are both compact, for any δ define

$$
\varepsilon^*(\delta)\equiv\inf_{\theta\in\Theta,\beta\in\mathcal{B}|\|\beta-B(\theta)\|\geq\delta}F\left(G(\theta,B(\theta)),B(\theta)\right)-F\left(G(\theta,\beta),\beta\right).
$$

Then choose $\varepsilon=\varepsilon^*(\delta)$. That means with probability approaching 1, $\sup_{\theta\in\Theta}\inf_{\beta\in\mathcal{B}|\|\beta-B(\theta_0)\|\geq\delta}\sup_{\theta\in\Theta}\Big[F(G($
$\varepsilon^*(\delta)$ so with probability approaching 1, $\sup_{\theta\in\Theta}\left\|\widetilde{B}(\theta)-B(\theta)\right\|<\delta$.

The fact that $\widetilde{B}(\theta)$ converges uniformly to $B(\theta)$ and that $\widehat{\beta}\xrightarrow{p}B(\theta_0)$ means that $\left(\widetilde{B}(\theta)-\widehat{\beta}\right)'\Omega\left(\widetilde{B}(\theta)-\widehat{\beta}\right)$ converges uniformly in probability to $-\left(B(\theta)-B(\theta_0)\right)'\Omega\left(B(\theta)-B(\theta_0)\right)$.

Thus we have verified all of the conditions of Newey and McFadden Theorem 2.1. ■

A.2 Asymptotic Distribution

We now explicitly define $G_j,\widehat{G}_j$, and $\widetilde{G}_j$ to be the j^{th} element of $G,\widehat{G}$, and $\widetilde{G}_h$ respectively. We first assume the following regularity conditions. These are weak assumptions that are standard and will hold in typical applications. The first is standard.

Assumption A7. *$B(\theta)$ is differentiable with*

$$
B_\theta\equiv\frac{dB(\theta_0)}{d\theta'}
$$

and $B_\theta'\Omega B_\theta$ is of full rank and θ_0 is an interior point.

The second is an assumption about stochastic equicontinuity. In small samples our estimator is potentially discontinuous in θ but it converges to a smooth function.

Assumption A8. *For and δ_N,*

$$
\sup_{\|\theta-\theta_0\|\leq\delta_N}\frac{\sqrt{N}\left\|\widetilde{B}(\theta)-\widetilde{B}(\theta_0)-B(\theta)+B(\theta_0)\right\|}{1+\sqrt{N}\|\theta-\theta_0\|}\xrightarrow{p}0.
$$

Finally we use,

Assumption A9. *$\widehat{G}$ and $\widetilde{G}$ are differentiable in β. Letting the notation $\overset{U_p}{\to}$ denote uniform convergence in probability,*

$$\frac{\partial \widehat{G}_j(\beta)}{\partial \beta} \overset{U_p}{\to} \frac{\partial G_j(\theta_0, \beta)}{\partial \beta}$$
$$\frac{\partial^2 \widehat{G}_j(\beta)}{\partial \beta \partial \beta'} \overset{U_p}{\to} \frac{\partial^2 G_j(\theta_0, \beta)}{\partial \beta \partial \beta'}$$
$$\frac{\partial \widetilde{G}_j(\theta_0, \beta)}{\partial \beta} \overset{U_p}{\to} \frac{\partial G_j(\beta)}{\partial \beta}$$
$$\frac{\partial^2 \widetilde{G}_j(\theta_0, \beta)}{\partial \beta \partial \beta'} \overset{U_p}{\to} \frac{\partial^2 G_j(\beta)}{\partial \beta \partial \beta'}$$

and all of these objects are continuous in their arguments.

The differentiability rules out some interesting cases like quantile regression. Extending this to allow for more complicated cases should be straight forward but our main goal is to provide the formula for the asymptotic variance in the typical case rather than the most general case.

Let $\beta_0 = B(\theta_0)$, we also need

Assumption A10. *F is two time (totally) continuously differential and define*

$$F_{\beta\beta} \equiv \frac{d^2 F(G(\theta_0, \beta_0), \beta_0)}{d\beta d\beta'}$$

and assume $F_{\beta\beta}$ is of full rank.

We assume that the X_{hs} are composed of actual values that we see in the data. Let M_{hi} be the total number of times X_i is used for each simulated data set h. This can pick up two important cases. In one case we let each simulated data set be the same size as the actual data ($S = N$) and each value of X_i is used once so $X_{hs} = X_s$. In this case $M_{hi} = 1$ for every i. The other case is one in which X_{hs} is drawn from the empirical distribution. In this case M_{hi} is a random variable taking integer values with expected value S/N. Of course it can cover other cases as well, for example if $S = 2N$ and each observable is used twice.

To simplify notation let Υ_{him} denote the m^{th} simulation using observation i for sample h. Define

$$\widetilde{g}_{hi}(\beta) \equiv \frac{N}{S} \sum_{m=1}^{M_{hi}} \frac{\ell(\Upsilon_{him}; X_i, \theta_0)}{\ell_0(\Upsilon_{him}; X_i)} g(X_i, y_\Upsilon(\Upsilon_{him}, X_i; \theta_0), \beta)$$

Notice that this means that

$$\begin{aligned}\widetilde{G}_h(\theta_0, \beta) =& \frac{1}{S} \sum_{s=1}^{S} \frac{\ell(\Upsilon_{hs}; X_{hs}, \theta)}{\ell_0(\Upsilon_{hs}; X_{hs})} g(X_{hs}, Y_{hs}, \beta) \\ =& \frac{1}{N} \sum_{i=1}^{N} \widetilde{g}_{hi}(\beta).\end{aligned}$$

Theorem 2. *Under Assumptions A1-A10, $\sqrt{N}\left(\widehat{\theta}-\theta\right)$ converges in distribution to a normal random variable with expected value 0 and variance*

$$\left[\frac{\partial B(\theta_0)'}{\partial\theta}\Omega\frac{\partial B(\theta_0)}{\partial\theta'}\right]^{-1}\frac{\partial B(\theta_0)'}{\partial\theta}\Omega F_{\beta\beta}^{-1}VF_{\beta\beta}^{-1}\Omega\tfrac{\partial B(\theta_0)}{\partial\theta}\left[\frac{\partial B(\theta_0)'}{\partial\theta}\Omega\frac{\partial B(\theta_0)}{\partial\theta'}\right]^{-1}.$$

Proof. We follow Newey and McFadden Theorem 7.2.
Our estimator satisfies their basic conditions to apply the theorem. Assumption (i) holds since $B(\theta_0)-B(\theta_0)=0$. Assumption A7 guarantees that (ii) and (iii) hold, and A8 guarantee that (v) holds.
To prove the result we need to derive the asymptotic distribution of $\sqrt{N}\left(\widetilde{B}(\theta_0)-\widehat{\beta}\right)$.
Consider $\widehat{\beta}$. Then the first order condition comes from totally differentiating the objective function

$$0=\frac{dF\left(\widehat{G}\left(\widehat{\beta}\right),\widehat{\beta}\right)}{d\beta}.$$

Let $\beta_0=B\left(\theta_0\right)$. With the mean value theorem we get

$$0=\left[\frac{d^2F\left(\widehat{G}\left(\overline{\beta}\right),\overline{\beta}\right)}{d\beta d\beta'}\right]\left(\widehat{\beta}-\beta_0\right)+\frac{dF\left(\widehat{G}\left(\beta_0\right),\beta_0\right)}{d\beta}.$$

Let G_j and $\widehat{G}_j$ be the j^{th}elements of G and $\widehat{G}$ respectively, then

$$\begin{aligned}
\frac{d^2F\left(\widehat{G}\left(\overline{\beta}\right),\overline{\beta}\right)}{d\beta d\beta'} =&\sum_{j=1}^{K_g}\frac{\partial F\left(\widehat{G}\left(\overline{\beta}\right)\right)}{\partial G_j}\frac{\partial^2\widehat{G}_j\left(\overline{\beta}\right)}{\partial\beta\partial\beta'}\\
&+\frac{\partial\widehat{G}\left(\overline{\beta}\right)'}{\partial\beta}\left(\frac{\partial^2F\left(\widehat{G}\left(\overline{\beta}\right),\overline{\beta}\right)}{\partial G\partial G'}\frac{\partial\widehat{G}\left(\overline{\beta}\right)}{\partial\beta'}+\frac{\partial^2F\left(\widehat{G}\left(\overline{\beta}\right),\overline{\beta}\right)}{\partial G\partial\beta'}\right)\\
&+\frac{\partial F\left(\widehat{G}\left(\overline{\beta}\right),\overline{\beta}\right)}{\partial\beta\partial G'}\frac{\partial\widehat{G}\left(\overline{\beta}\right)}{\partial\beta'}+\frac{\partial^2F\left(\widehat{G}\left(\overline{\beta}\right),\overline{\beta}\right)}{\partial\beta\partial\beta'}\\
\xrightarrow{U_p}&\sum_{j=1}^{K_g}\frac{\partial F\left(G\left(\beta_0\right)\right)}{\partial G_j}\frac{\partial^2G_j\left(\beta_0\right)}{\partial\beta\partial\beta'}\\
&+\frac{\partial G\left(\beta_0\right)}{\partial\beta}\left(\frac{\partial^2F\left(G\left(\beta_0\right),\beta_0\right)}{\partial G\partial G'}\frac{\partial G\left(\beta_0\right)}{\partial\beta'}+\frac{\partial^2F\left(G\left(\beta_0\right),\beta_0\right)}{\partial G\partial\beta'}\right)\\
&+\frac{\partial F\left(G\left(\beta_0\right),\beta_0\right)}{\partial\beta\partial G'}\frac{\partial G\left(\beta_0\right)}{\partial\beta'}+\frac{\partial^2F\left(G\left(\beta_0\right),\beta_0\right)}{\partial\beta\partial\beta'}\\
=&F_{\beta\beta}
\end{aligned}$$

and using the fact that β_0 solves

$$
\begin{aligned}
0 = & \frac{dF\left(G\left(\beta_0\right), \beta_0\right)}{d\beta} \\
= & \frac{\partial G\left(\beta_0\right)}{\partial \beta} \frac{\partial F\left(G\left(\beta_0\right), \beta_0\right)}{\partial G} + \frac{\partial F\left(G\left(\beta_0\right), \beta_0\right)}{\partial \beta}
\end{aligned}
$$

then adding and subtracting terms including the term in the above expression and using the mean value theorem

$$
\begin{aligned}
& \sqrt{N} \frac{dF\left(\widehat{G}\left(\beta_0\right), \beta_0\right)}{d\beta} \\
= & \sqrt{N}\left(\frac{\partial \widehat{G}\left(\beta_0\right)'}{\partial \beta} \frac{\partial F\left(\widehat{G}\left(\beta_0\right), \beta_0\right)}{\partial G} + \frac{\partial F\left(\widehat{G}\left(\beta_0\right), \beta_0\right)}{\partial \beta}\right) \\
= & \sqrt{N}\left(\frac{\partial \widehat{G}\left(\beta_0\right)}{\partial \beta}' \frac{\partial F\left(\widehat{G}\left(\beta_0\right), \beta_0\right)}{\partial G} - \frac{\partial \widehat{G}\left(\beta_0\right)}{\partial \beta}' \frac{\partial F\left(G\left(\beta_0\right), \beta_0\right)}{\partial G}\right) \\
& + \sqrt{N}\left(\frac{\partial \widehat{G}\left(\beta_0\right)'}{\partial \beta} \frac{\partial F\left(G\left(\beta_0\right), \beta_0\right)}{\partial G} - \frac{\partial G\left(\beta_0\right)'}{\partial \beta} \frac{\partial F\left(G\left(\beta_0\right), \beta_0\right)}{\partial G}\right) \\
& + \sqrt{N}\left(\frac{\partial F\left(\widehat{G}\left(\beta_0\right), \beta_0\right)}{\partial \beta} - \frac{\partial F\left(G\left(\beta_0\right), \beta_0\right)}{\partial \beta}\right) \\
= & \sqrt{N}\left(\frac{\partial G\left(\beta_0\right)'}{\partial \beta} \frac{\partial^2 F\left(G\left(\beta_0\right), \beta_0\right)}{\partial G \partial G'} \frac{1}{N} \sum_{i=1}^{N}\left(g\left(X_i, Y_i, \beta_0\right) - G\left(\beta_0\right)\right)\right) \\
& + \sqrt{N}\left(\frac{1}{N} \sum_{i=1}^{N}\left(\frac{\partial g\left(X_i, Y_i, \beta_0\right)'}{\partial \beta} - \frac{\partial G\left(\beta_0\right)'}{\partial \beta}\right) \frac{\partial F\left(G\left(\beta_0\right), \beta_0\right)}{\partial G}\right) \\
& + \sqrt{N}\left(\frac{\partial^2 F\left(G\left(\beta_0\right), \beta_0\right)}{\partial \beta \partial G'} \frac{1}{N} \sum_{i=1}^{N}\left(g\left(X_i, Y_i, \beta_0\right) - G\left(\beta_0\right)\right)\right) + o_p(1) \\
= & \frac{1}{\sqrt{N}} \sum_{i=1}^{N}\left[\left(\frac{\partial G\left(\beta_0\right)}{\partial \beta}' \frac{\partial^2 F\left(G\left(\beta_0\right), \beta_0\right)}{\partial G \partial G'} + \frac{\partial^2 F\left(G\left(\beta_0\right), \beta_0\right)}{\partial \beta \partial G'}\right)\left(g\left(X_i, Y_i, \beta_0\right) - G\left(\beta_0\right)\right)\right. \\
& \left. + \left(\frac{\partial g\left(X_i, Y_i, \beta_0\right)'}{\partial \beta} - \frac{\partial G\left(\beta_0\right)'}{\partial \beta}\right) \frac{\partial F\left(G\left(\beta_0\right), \beta_0\right)}{\partial G}\right] + o_p(1) \\
= & \frac{1}{\sqrt{N}} \sum_{i=1}^{N} \vartheta_i + o_p(1).
\end{aligned}
$$

Next we derive the asymptotic distribution for $\widetilde{B}\left(\theta_0\right)$. This follows an analogous but slightly more complicated derivation. First define

$$
\widetilde{B}_h\left(\theta\right) \equiv argmin_\beta F\left(G(\theta, \beta)\right)
$$

then the first order condition and mean value theorem gives for each $h = 1, ..., H$

$$0 = \frac{d^2 F\left(\widetilde{G}_h\left(\theta_0, \overline{\beta}\right), \overline{\beta}\right)}{d\beta d\beta'}\left(\widetilde{B}_h\left(\theta_0\right) - \beta_0\right) + \frac{dF\left(\widetilde{G}_h\left(\theta_0, \beta_0\right), \beta_0\right)}{d\beta}$$

where

$$\begin{aligned}
\frac{d^2 F\left(\widetilde{G}_h\left(\theta_0, \overline{\beta}\right), \overline{\beta}\right)}{d\beta d\beta'} =& \sum_{j=1}^{K_g} \frac{\partial F\left(\widetilde{G}_h\left(\theta_0, \overline{\beta}\right), \overline{\beta}\right)}{\partial G_j'} \frac{\partial^2 \widetilde{G}_{hj}\left(\theta_0, \overline{\beta}\right)}{\partial\beta\partial\beta'} \\
&+ \frac{\partial \widetilde{G}_h\left(\theta_0, \overline{\beta}\right)'}{\partial\beta}\left(\frac{\partial^2 F\left(\widetilde{G}_h\left(\theta_0, \overline{\beta}\right), \overline{\beta}\right)}{\partial G\partial G'} \frac{\partial \widetilde{G}_h\left(\theta_0, \overline{\beta}\right)}{\partial\beta'} + \frac{\partial^2 F\left(\widetilde{G}_h\left(\theta_0, \overline{\beta}\right), \overline{\beta}\right)}{\partial G\partial\beta'}\right) \\
&+ \frac{\partial F\left(\widetilde{G}_h\left(\theta_0, \overline{\beta}\right), \overline{\beta}\right)}{\partial\beta\partial G'} \frac{\partial \widetilde{G}_h\left(\theta_0, \overline{\beta}\right)}{\partial\beta'} + \frac{\partial^2 F\left(\widetilde{G}_h\left(\theta_0, \overline{\beta}\right), \overline{\beta}\right)}{\partial\beta\partial\beta'} \\
\overset{U_p}{\rightarrow}& \sum_{j=1}^{K_g} \frac{\partial F\left(G\left(\beta_0\right)\right)}{\partial G_j'} \frac{\partial^2 G_j\left(\beta_0\right)}{\partial\beta\partial\beta'} \\
&+ \frac{\partial G\left(\beta_0\right)'}{\partial\beta}\left(\frac{\partial^2 F\left(G\left(\beta_0\right), \beta_0\right)}{\partial G\partial G'} \frac{\partial G\left(\beta_0\right)}{\partial\beta'} + \frac{\partial^2 F\left(G\left(\beta_0\right), \beta_0\right)}{\partial G\partial\beta'}\right) \\
&+ \frac{\partial F\left(G\left(\beta_0\right), \beta_0\right)}{\partial\beta\partial G'} \frac{\partial G'\left(\beta_0\right)}{\partial\beta} + \frac{\partial^2 F\left(G\left(\beta_0\right), \beta_0\right)}{\partial\beta\partial\beta'} \\
=& F_{\beta\beta}
\end{aligned}$$

where $\widetilde{G}_{hj}$ is the j^{th}element of $\widetilde{G}_h$.

Analogously to above

$$
\begin{aligned}
&\sqrt{N}\frac{dF\left(\widetilde{G}_h\left(\theta_0,\beta_0\right),\beta_0\right)}{d\beta}\\
=&\sqrt{N}\left(\frac{\partial\widetilde{G}_h\left(\theta_0,\beta_0\right)}{\partial\beta}\frac{\partial F\left(\widetilde{G}_h\left(\theta_0,\beta_0\right),\beta_0\right)}{\partial G}+\frac{\partial F\left(\widetilde{G}_h\left(\theta_0,\beta_0\right),\beta_0\right)}{\partial\beta}\right)\\
=&\sqrt{N}\left(\frac{\partial\widetilde{G}_h\left(\theta_0,\beta_0\right)}{\partial\beta}\frac{\partial F\left(\widetilde{G}_h\left(\theta_0,\beta_0\right),\beta_0\right)}{\partial G}-\frac{\partial\widetilde{G}_h\left(\theta_0,\beta_0\right)}{\partial\beta}\frac{\partial F\left(G\left(\beta_0\right),\beta_0\right)}{\partial G}\right)\\
&+\sqrt{N}\left(\frac{\partial\widetilde{G}_h\left(\theta_0,\beta_0\right)}{\partial\beta}\frac{\partial F\left(G\left(\beta_0\right),\beta_0\right)}{\partial G}-\frac{\partial G\left(\beta_0\right)}{\partial\beta}\frac{\partial F\left(G\left(\beta_0\right),\beta_0\right)}{\partial G}\right)\\
&+\sqrt{N}\left(\frac{\partial F\left(\widetilde{G}_h\left(\theta_0,\beta_0\right),\beta_0\right)}{\partial\beta}-\frac{\partial F\left(G\left(\beta_0\right),\beta_0\right)}{\partial\beta}\right)\\
=&\sqrt{N}\left(\frac{\partial G\left(\beta_0\right)}{\partial\beta'}\frac{\partial F\left(G\left(\beta_0\right),B\left(\theta_0\right)\right)}{\partial G\partial G'}\frac{1}{N}\sum_{i=1}^{N}\left(\widetilde{g}_{hi}\left(\beta_0\right)-G\left(\beta_0\right)\right)\right)\\
&+\sqrt{N}\left(\frac{1}{N}\sum_{i=1}^{N}\left(\frac{\partial\widetilde{g}_{hi}\left(\beta_0\right)}{\partial\beta}-\frac{\partial G\left(\beta_0\right)}{\partial\beta}\right)\frac{\partial F\left(G\left(\beta_0\right),\beta_0\right)}{\partial G}\right)\\
&+\sqrt{N}\left(\frac{\partial F\left(G\left(\beta_0\right),\beta_0\right)}{\partial\beta\partial G'}\frac{1}{N}\sum_{i=1}^{N}\left(\widetilde{g}_{hi}\left(\beta_0\right)-G\left(\beta_0\right)\right)\right)+o_p(1)\\
=&\frac{1}{\sqrt{N}}\sum_{i=1}^{N}\left[\left(\frac{\partial G\left(\beta_0\right)}{\partial\beta'}\frac{\partial F\left(G\left(\beta_0\right),\beta_0\right)}{\partial G\partial G'}+\frac{\partial F\left(G\left(\beta_0\right),\beta_0\right)}{\partial\beta\partial G'}\right)\left(\widetilde{g}_{hi}\left(\beta_0\right)-G\left(\beta_0\right)\right)\right.\\
&\left.+\left(\frac{\partial\widetilde{g}_{hi}\left(\beta_0\right)}{\partial\beta}-\frac{\partial G\left(\beta_0\right)}{\partial\beta}\right)\frac{\partial F\left(G\left(\beta_0\right),\beta_0\right)}{\partial G}\right]+o_p(1)\\
=&\frac{1}{\sqrt{N}}\sum_{i=1}^{N}\widetilde{\vartheta}_{hi}+o_p(1).
\end{aligned}
$$

And so

$$
\begin{aligned}
\sqrt{N}\left[\widehat{B}(\theta_0)-\widehat{\beta}\right]=&F_{\beta\beta}^{-1}\frac{1}{\sqrt{N}}\sum_{i=1}^{N}\left(\left[\frac{1}{H}\sum_{h=1}^{H}\widetilde{\vartheta}_{hi}\right]-\vartheta_i\right)+o_p(1)\\
\sim&N\left(0,F_{\beta\beta}^{-1}VF_{\beta\beta}^{-1}\right).
\end{aligned}
$$

Then

$$
\sqrt{N}\left(\widehat{\theta}-\theta_0\right)\xrightarrow{d}N\left(0,\left[\frac{\partial B(\theta_0)'}{\partial\theta}\Omega\frac{\partial B(\theta_0)}{\partial\theta'}\right]^{-1}\frac{\partial B(\theta_0)'}{\partial\theta}\Omega F_{\beta\beta}^{-1}VF_{\beta\beta}^{-1}\Omega\frac{\partial B(\theta_0)}{\partial\theta}\left[\frac{\partial B(\theta_0)'}{\partial\theta}\Omega\frac{\partial B(\theta_0)}{\partial\theta'}\right]^{-1}\right).
$$

■

Appendix B: Data

As mentioned in the text, we use white women from the last four panels of the Survey of Income and Program Participation. We first measure potential experience in months and use anyone from 1 month to 35 years of potential experience. The variable used in the data is annualized. The SIPP is asked every four months. We only use data from the month of the interview.
We detail construction of the variables

- Potential Experience: For older workers we don't know exactly when they graduated school. We assume that they graduate in June of the year they turn a)16 if education is less than 12, b) 18 if their education is exactly 12, c) 20 if their education is more than 12 but less than 16, and d) 22 if their education is larger than 22. Their potential experience is their current age minus the age when they graduated school.

- Employment: We define employment to be 1 for individuals who work some during the survey month.

- Education: We take the maximum of the education variable in each wave which is completed education. We convert to numeric variables as, 0 if less than first grade, 2.5 if education is first through fifth grade, 5.5 if it is fifth or sixth, 7.5 if it is seventh or eighth, the numeric grade completed through high school, 12 if high school or equivalent, 13 if a vocational certificate, 13.5 if a vocational associate degree, 14 if an academic vocational degree, 16 if a four year graduate, 17 if a masters degree, and 18 if professional degree or higher.

- Log wage: Wage is constructed as the hourly rate of pay for people who are paid by the hour and monthly earnings divided by (weeks worked$\times$usual hours per week). If one worked every week of the month we use 4.3 as the number of weeks. It is deflated to 2008 dollars using the personal consumption expenditures price index. We drop observations with a real wage below 1$ or above 300$.

- Married/Divorced. Each wave women are asked about their marital status. We simplify this to three possibilities. A person is coded as currently married if they are married at the time of the survey and not separated (whether the spouse is present or not). We define a person to be divorced if they were previously married, but are no longer. This can be either due to divorce, separation, or widowhood. Finally, never married women represent the third category.

- Children and Fertility: We use a number of different variables here that come from two different underlying sources. One is the household roster. For anyone in the household for whom the main respondent is listed as their mother, we keep track of the year and month in which they were born. The other source is topical module wave 2 which asks about fertility. From this, the variable tmomchl tells us the total number of children and we also collect the year and month of birth for the oldest and youngest child. From these we construct a number of different variables:

– The Number of Kids<7,Number of Kids <18, and number of kids age XX dummies are constructed based on the date of the survey and the age of kids from the household roster.

– The "Any Kids", "Two Kids", and "Number Kids" variables in Table 3d are based on the number of kids from the topical module. The "Gave Birth" variable in Table 3d comes from the house hold roster looking at children born between survey dates.

– The "Total Kids > 18" variable used in the wage growth regression is total kids (from the wave 2 survey) minus the number of kids <18 (from the household roster as of wave 2).

Appendix C: Auxiliary Model and Model Fit

We have a total of 418 auxiliary parameters. In this appendix we describe them, state how much weight is given to each, and show both the data and the simulated values. In terms of weights, all of them start with the inverse of the variance of estimated parameters and are scaled by different amounts listed below.

We then discuss identification loosely in that we discuss which parameters of the auxiliary parameters are useful for estimating the parameters of the structural model

C.1 Auxliary Parameters

Log Wage Fixed Effect Regression We begin with a regression of log wages on a number of variables controlling for individual fixed effects. We use all wage observations for which we have data. The results are in the first column of Table C1 and the first panel of Figure C1. We also keep an estimate of the fixed effect from this regression which we will use further.

- 35 Experience dummies (all scale 1)
- Number of Children < 18 (scale 101)
- Number of Children < 7 (scale 101)
- Dummy variable for Married (scale 101)

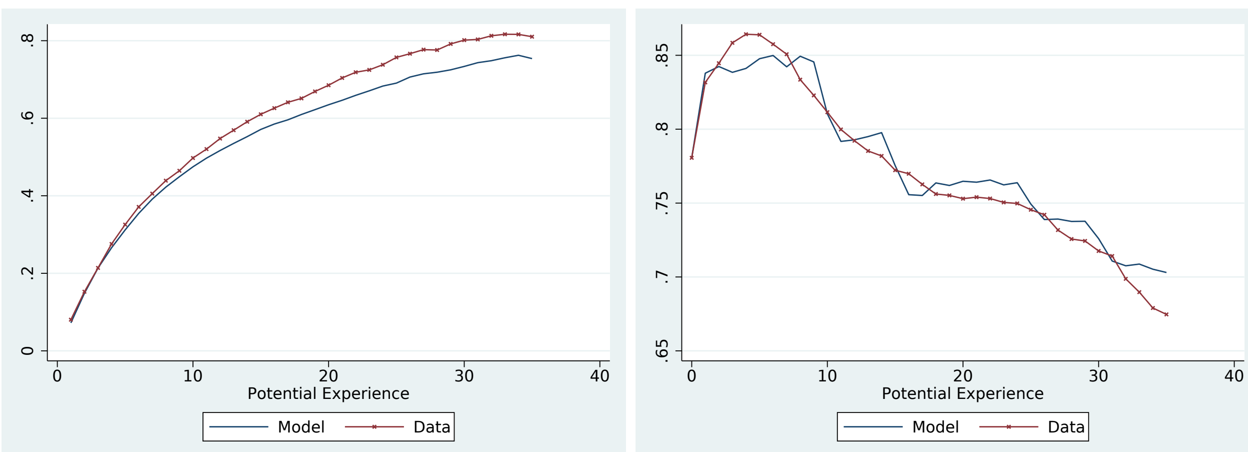
Table C1
Fit of Model:Fixed Effect Wages

Covariate	log(wage) with fixed effects Model	log(wage) with fixed effects Data	Fixed Effects Themselves Model	Fixed Effects Themselves Data
Education			0.114	0.114
				(0.001)
Married	0.019	0.019		
		(0.004)		
Number of Kids < 18	0.002	0.003		
		(0.003)		
Number of Kids < 7	0.001	0.001		
		(0.003)		

Figure C1: Fit Age Fixed Effects

(a) Log Wage Fixed Effects

(b) Work Fixed Effects



Regression of Wage Fixed Effect on Education We construct the estimates of the fixed effects from the previous regression and run a regression of them onto education (by person so each person gets equal weight regardless of the number of times we observe them in the data). The results are presented in the last to columns of Table C1.

- Coefficient on education (scale=101)
- Intercept from regression not matched (scale=0)

Work Fixed Effect Regression We next run a similar regression with individual fixed effects in which the dependent variable is a dummy variable for whether the respondent worked. We also save this variable for further results. The results can be seen in Table C2 and the second panel of Figure C1.

- 35 experience dummies (all scale 11)

- 8 dummy variables for number of kids ages 0-7 (0-6 scale 101, 7+ scale 100)
- a dummy variable for married (scale 101)

Note that our last column is greater than or equal to 7 as opposed to ages 7-18. The reason is that this is a short panel so we want identification of it to come entirely from the change in labor supply as kids move from 6 to 7 and older, not the change in labor supply as kids move from 18 to 19.

Table C2
Fit of Model:Fixed Effect Work

	Model	Data
Married	-0.028	-0.034
		(0.004)
Number of Kids age 0	-0.081	-0.092
		(0.005)
Number of Kids age 1	-0.052	-0.082
		(0.006)
Number of Kids age 2	-0.050	-0.061
		(0.006)
Number of Kids age 3	-0.051	-0.047
		(0.007)
Number of Kids age 4	-0.048	-0.039
		(0.007)
Number of Kids age 5	-0.049	-0.030
		(0.007)
Number of Kids age 6	-0.048	-0.022
		(0.008)
Number of Kids age ≥ 7	-0.004	-0.0004
		(0.007)

Within and Between Variance of Residual from Fixed Effect Regressions

- Let T_i be the number of wage observations we have from individual i. Construct residuals from the fixed effect regression above (where the fixed effect is included in the residual). Define these as ω_{it} and order them as $\omega_{i1}, ..., \omega_{iT_i}$. Define

$$\overline{\omega}_i \equiv \frac{1}{T_i} \sum_{t=1}^{T_i} \log(\omega_{it})$$
$$\overline{\omega} = \frac{\sum_{i=1}^{N} \sum_{t=1}^{T_i} \log(\omega_{it})}{\sum_{i=1}^{N} T_i}.$$

We can then decompose the total variance into

$$\frac{\sum_{i=1}^{N} \sum_{t=1}^{T_i} [\log(\omega_{it}) - \overline{\omega}]^2}{\sum_{i=1}^{N} T_i} = \frac{\sum_{i=1}^{N} \sum_{t=1}^{T_i} [\log(\omega_{it}) - \overline{\omega}_i]^2}{\sum_{i=1}^{N} T_i} + \frac{\sum_{i=1}^{N} T_i [\overline{\omega}_i - \overline{\omega}]^2}{\sum_{i=1}^{N} T_i}$$

where the first part is the within variance and the second is the between variance.

- The within variance has scale 11
- The between variance has scale 101

These results are shown in the first two rows of Table C3

Table C3
Fit of Model : Second Moments of Fixed Effects and Residuals

	Model	Data
Within Variance Log Wages	0.068	0.067
		(0.001)
Between Variance Log Wages	0.263	0.263
		(0.002)
Variance Log Work Fixed Effect	0.140	0.139
		(0.001)
Reg. Coef. Wage FE on Work FE	0.215	0.214
		(0.007)

Variance of Work Fixed Effect We take the sample variance of the work fixed effect. This has a scale of 101 and is shown in the third row of Table C3.

Regression of Wage Fixed Effect on Work Fixed Effect By person we run a regression of the fixed effect from wages on the work fixed effect. The intercept in this model gets no weight and the coefficient on the work fixed effect receives a scale of 101. The slope coefficient is shown in the last row of Table C3.

Fixed Effect Residual In the results that follow when we use either the wage or work fixed effect, we first purge it of the education effect by running it on education and taking a residual. When we use these fixed effects from this point forward, we mean this residualized value.

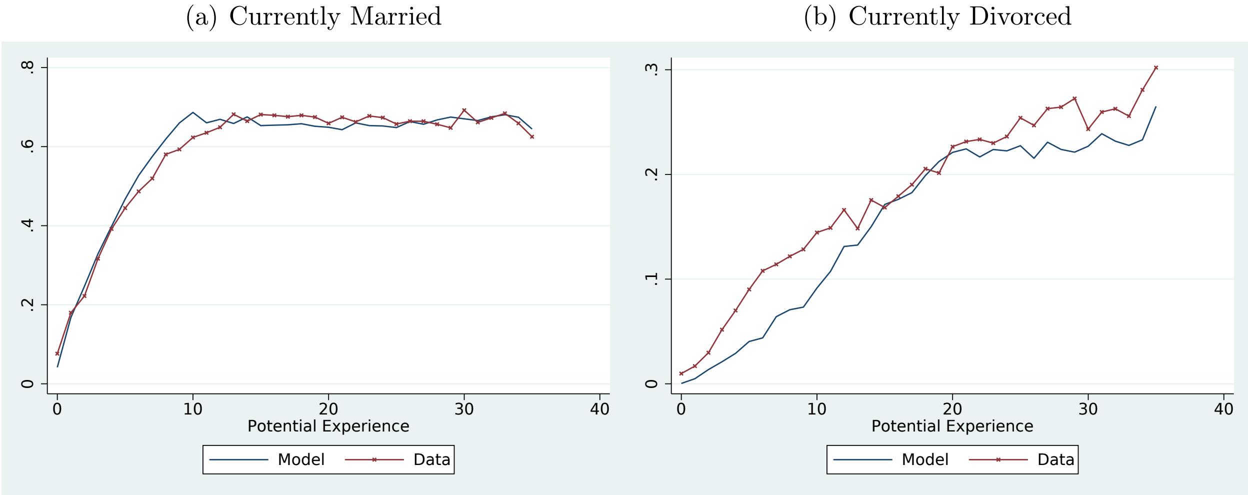
Linear probability model of being married/divorced in the first period observed For the first time we observe someone we run a regression of their marital status on age dummies. We run the same regression for the variable married and for divorced (with the third category never married). The scale on all parameters is 1. These results are shown in Figure C2.

Linear probability of marital status conditional on previous marital status We use all panels beyond the first. Our first regression is marriage in period t conditional on not being married in period t-1. The second is divorce in period t conditional on marriage in t-1.

They take identical specifications:

- quadradic in experience dummies (all scale=0)
- education (scale=101)

Figure C2: Fit Marital Status by Age



- employed period t-1 (scale=101)
- wage in period t-1 (scale=101)
- intercept (scale=0)

The results are presented in Table C4.

Table C4
Fit of Model : Getting Married/Divorced

Covariate	Get Married($\times$100) Model	Get Married($\times$100) Data	Get Divorced($\times$100) Model	Get Divorced($\times$100) Data
Education	-0.009	0.002	-0.062	-0.066
		(0.018)		(0.008)
Initial Work	-0.158	-0.157	1.036	0.916
		(0.221)		(0.121)
Initial Wage	0.158	0.155	-0.273	-0.292
		(0.084)		(0.041)
Pot. Exp. Quadradic	Yes	Yes	Yes	Yes

Marital and Work Status at Time of Birth For people who have a child in one period we calculate: a) the fraction that were married 1 period earlier and b) the fraction who were working 1 year earlier. Both of these moments get a scale of 101 and are presented in the first two rows of Table C5.

Table C5
Fit of Model : Fertility Related Moments

	Model	Data
Fraction Married When Giving Birth	0.739	0.733
		(0.007)
Working Before Giving Birth	0.693	0.690
		(0.008)
Age Difference Youngest/ Oldest	5.779	5.776
		(0.023)

Age difference between oldest and youngest child From the wave 2 records of children we take the difference in ages between the youngest and oldest child. This gets a scale of 101 and is shown in the last row of Table C5

Regression of children on covariates Based on the second wave we run a regression of functions of the number of children on age dummies (weight=1) and education (weight 101). The three dependent variables are a dummy for having any children, a dummy for exactly two childre, and the number of children.

The education coefficients are in the first 6 columns of Table C6 and the age patterns in Figure C3.

Table C6
Fit of Model : Fertility

Covariate	Any Kids		Two Kids		Number Kids		Gave Birth	
	Model	Data	Model	Data	Model	Data	Model	Data
Education	-0.025	-0.026	0.003	0.002	-0.112	-0.111		
		(0.001)		(0.001)		(0.002)		
Work Last Year							-0.018	-0.008
								(0.004)
Wage Last Year							0.005	0.006
								(0.001)
Pot. Exp. Dummies	Yes	Yes	Yes	Yes	Yes	Yes	No	No
Pot. Exp. Quadradic	No	No	No	No	No	No	Yes	Yes

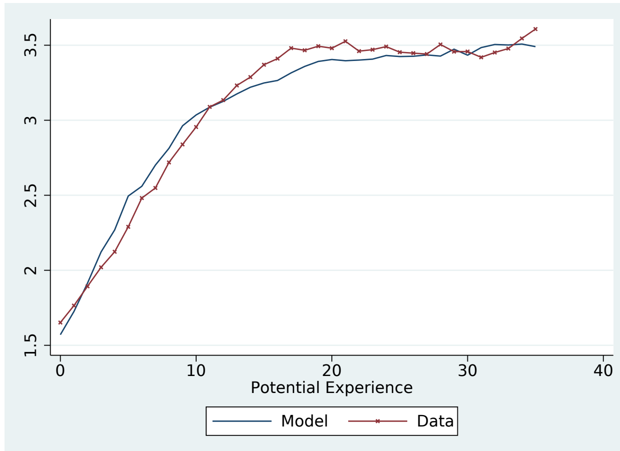
Regression of giving birth on intial wage and work The dependent variable is whether someone gave birth between any two waves in the panel and we regress that (by person) on characteristics the first wave that we see them.

- working (scale=101)
- wage for women working (scale=101)
- also in regression: intercept, education, married, number of kids <7, number of kids ≤ 18, potential experience, potential expience^2 (not matched, scale=0)

These results are presented in the last two columns of Table C5.

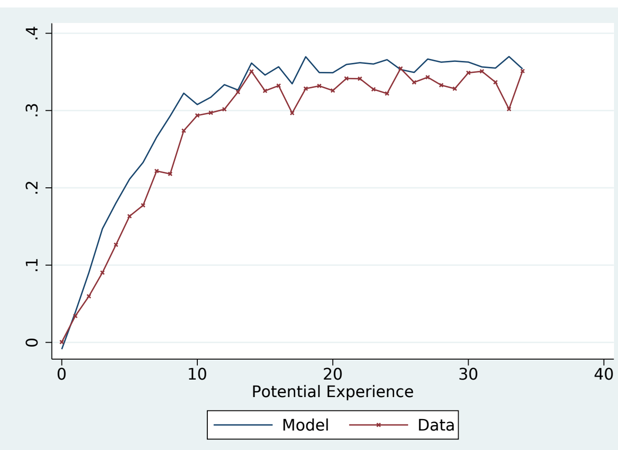
Figure C3: Fit of Model: Children

(a) Number of Children



(b) Any Children

(c) Exactly Two Children



Linear probability model of working at conditional on previous work status We regress a dummy variatble whether people were working in period t conditional on work status in period t-1 on a number of covariates. For both models we use the specification

- experience dummy variables (scale=1)
- education (scale=101)
- married (scale=101)
- number of kids <7 (scale=101)
- Work fixed effect (residualized) (scale=101)

The results are presented in Table C7 and Figure C4.

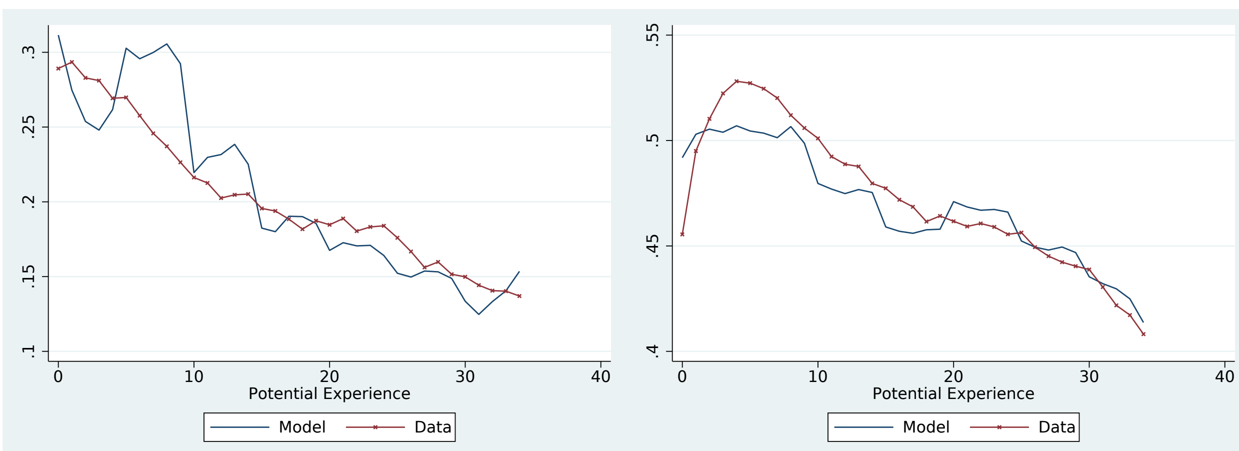
Table C7
Fit of Model : Work

Covariate	Start Work		Keep Working	
	Model	Data	Model	Data
Education	0.027	0.027	0.026	0.026
		(0.000)		(0.000)
Married	-0.041	-0.042	-0.019	-0.019
		(0.002)		(0.001)
Number of Kids < 7	-0.038	-0.038	-0.032	-0.032
		(0.001)		(0.001)
Work Fixed Effect	0.757	0.759	0.681	0.682
		(0.003)		(0.003)
Exp. Dummies	Yes	Yes	Yes	Yes

Figure C4: Fit of Model: Work Transitions

(a) Nonemployed to Employed

(b) Employed to Employed



Wage Change Regressions We regress wage growth between period t and t-1 for people working in both periods. The regressions are

- 36 experience dummy variables (scale=11)
- education (scale=801)
- married (scale=801)
- number of kids ≥ 18 (scale=801)
- education $\times$ number of kids ≥ 18 (scale=801)
- education $\times$ potential experience (scale=801)

These results are found in the first two columns of Table C8 and Figure C5.

We also run wage regressions for people with a non-employment spell in between. That is, we look at people who had a wave where they were working followed by one or more waves of non-employment, followed by a wave in which they were working. We regress the difference in wages post and pre non-employment spell on the length of the time in between with no intercept. This has a scale of 801 and is shown in Table C7.

Table C8
Fit of Model : Wage Growth

Covariate	Wage Growth×100 Continuously Employed		Wage Growth ×100 Nonemployment Spell	
	Model	Data	Model	Data
Education	0.156	0.157		
		(0.045)		
Married	-0.160	-0.165		
		(0.075)		
Total Kids > 18	-0.290	-0.286		
		(0.215)		
Educ.× Pot. Exp.	-0.009	-0.008		
		(0.002)		
Educ.× Total Kids	0.027	0.031		
		(0.016)		
Change in Potential Experience			-0.895	-0.921
				(0.429)
Pot. Exp. Dummies	Yes	Yes	Yes	Yes

C.2 Identification

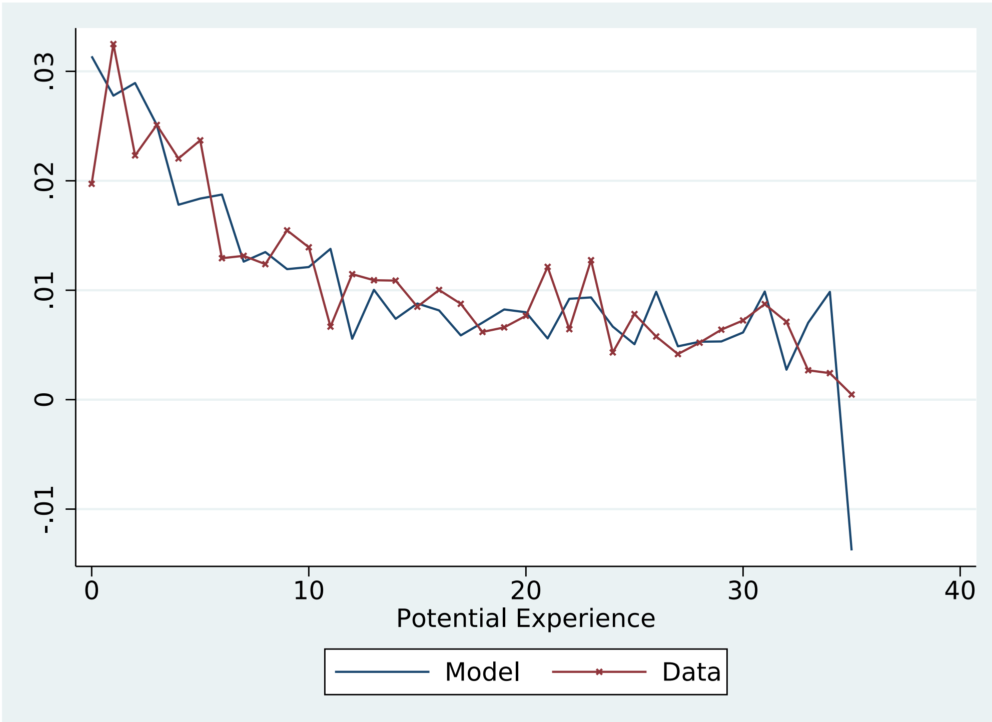
Next we discuss which moments are useful for identifying which structural parameters. This is loose as all parameters are identified by all auxiliary parameters-and we have not showed this formally. Informally this is what let us to choose the particular auxiliary parameters and in practice it feels like it is approximately correct.

We begin with a discussion of the factors. We have two factors ν_1 and ν_2. While the model is somewhat more complicated than a standard factor model we think of it in similar terms. In particular there is the standard rotation problem. We essentially normalize this by thinking of ν_1 as entering the wage equation-while ν_2 does not, so that ν_1 picks up a form of unobserved ability. Similarly we allow ν_2 to enter the hazards determining work and ν_1 is excluded from them, so ν_2 primarily picks up taste for work. We allow these two things to be correlated with each other (and in practice are very highly correlated). Note that this is not as flexible as we could be in that we have multiple equations determining labor supply, but given our relatively short panel we chose a more parsimonious model.

Hazard Model into and out of Marriage

These parameters (from the first two columns of Table 2a) are determined primarily by the regressions into and out of Marriage (Table C3) as well as the marriage and divorce regressions over the lifecycle (Figure C2).

Figure C5: Fit of Model: Wage Growth Employed



More specifically

- the education coefficient in each hazard is primarily identified through the education coefficients in the regressions
- the effect of ν_1 on marriage is primarily determined by the coefficient of wage in the previous year in the marriage/divorce regression.
- the effect of ν_2 on marriage is primarily determined by the coefficient of work in the previous year in the marriage/divorce regression.
- The parameters of the potential experience spline are primarily determined by the marriage and divorce panels across the lifecycle (i.e. the parameters presented in Figure C2)

Hazard Model into and out of Work

These parameters (from the third and fourth columns of Table 2a) are determined primarily by the fixed effect work regression (Table C2/Figure C1 panel b) and the into and out of work regressions (Table C6).

More specifically

- the education coefficient in each hazard is primarily identified through the education coefficients in the regression into and out of work in Table C5.
- the effect of ν_2 comes from both the variance of the fixed effect (Table C3) as well as from the transition regression coefficients on the fixed effect in Table C6.

- the coefficients on married come from the coefficient on marriage in all three regressions.
- the effect of number of children on work comes from the coefficients on number of children (and at different ages) from all three regressions
- parameters of the working spline are primarily identified from the pattern of working over the lifecycle in the fixed effect regression.

Hazard Model for Fertility

These parameters are presented in the last column of Table 2a. They come from a number of different auxiliary moments. The regressions of any children, exactly 2 children, and the number of children on education and potential experience dummies (Table C5 and Figure C3), the regression of giving birth on previous labor market status (Table C5), the age difference between the youngest and oldest child (Table C4), the fraction married when giving birth (Table C4), and the fraction working when giving birth (Table C4).

In more detail

- The education coefficient and interaction between education and number of kids come primarily from the education coefficient in the first three regression mentioned above and shown in Table C5.
- The effects of ν_1 and ν_2 come primarily from the coefficient of the gave birth regression on worked last year and wage last year.
- The married coefficient comes from the fraction who were married when giving birth.
- The working coefficient comes from the fraction who were working at the time of giving birth.
- The number of kids and the age coefficient come from the first three regressions (any children/number of children/exactly two children) and the age difference between the youngest and oldest child.

Human Capital Production Function

The parameters of the human capital production are almost entirely determined by the wage growth regression (first column Table C7 and Figure C5).

The parameters determining a come from the first order magnitude of growth.

- The intercept comes from the size of the fixed effects
- The coefficient on education for a comes from the direct effect of education on wage growth
- The effect of marriage comes from the marriage coefficient

The curvature is somewhat more subtle. The overall curvature is estimated by the potential experience dummies. This can come from two sources $\bar{H}$ and λ where the former is like "actual experience" in that wages grow less when there is more accumulated human capital and the latter is like “potential experience” in that it only depends on age. The key to distinguishing between the two is the number of children over 18. Women of the same age but with more children less accumulated experience and thus should have faster wage growth. So key to the relative size of these two parameters is the coefficient on number of kids >18. To get the two education coefficients for $\bar{H}$ and λ, the interactions of eduction with age and Kids>18 are key.

Depreciation

The depreciation parameter δ is identified by the extent to which wages fall when people don't work that is presented in the right column of Table C7.

Wages

Beyond human capital we also allow several things to affect wages (Table 2c final column). All are related to the fixed effect wage regression

- The structural coefficients on children and marriage come directly from the analogous coefficients in the fixed effect regression (first column Table C1).
- The magnitude of unobserved heterogeneity, ν_1 is identified by the between variance (Table C3).
- The education coefficient comes from the regression of the fixed effect on education (second column of Table C1).

The measurement error in wages σ_ε^2 (Table 2d) is identified primarily from the within variance (Table C3).

Correlation between factors

This correlation (Table 2d) is identified primarily from the regression coefficient of the wage fixed effect on the work fixed effect (listed in Table C3).

Initial work and Work right after Giving Birth

This part of the model is a bit different than the others in that for the other parameters we started with the model and then looked for auxiliary parameters that could identify them. The motivation for these terms was the opposite-we added them to help fitting parts of the data. The intention of the initial work is to fit the initial work at the beginning of the lifecycle. The intention of the work right after birth is to fit the child age dummies in the fixed effect work specification. The education and ν_2 parameters are not directly mapped to anything and in practice are there to fit other parameters. In practice neither of these are significant and do not play a major role in what we do as few people are predicted to work initially and very few women are predicted to stop working immediately after giving birth in this specification).