ANOTHER LOOK AT THE IDENTIFICATION OF DYNAMIC DISCRETE DECISION PROCESSES, WITH AN APPLICATION TO RETIREMENT BEHAVIOR

Victor Aguirregabiria

Department of Economics. University of Toronto

Princeton University. Econometrics Seminar

December 4, 2007

MOTIVATION AND CONTEXT

• **Dynamic Discrete Structural Models** have proved to be useful tools for the **assessment of public policy initiatives**.

Examples: Social Security and retirement; Investment subsidies; Educational policies; Unemployment insurance; etc

• A common feature in these applications is the **full parametric specification** of the model structure: i.e., preferences, technology, transition probabilities of state variables, and probability distribution of unobservable variables.

MOTIVATION AND CONTEXT (2)

- In some applications, the estimates of the effects of a policy can be very sensitive to some parametric assumptions on the primitives.

- **Can we estimate DDSM and evaluate counterfactual policies when the primitives of these models are nonparametrically specified?**

- **Can we define a nonparametric test to check the validity of a parametric DDSM model in the context of the evaluation of a policy?**

- This paper addresses these two questions.

MOTIVATION AND CONTEXT (3)

• Rust (1994) and Magnac and Thesmar have shown that (in contrast to static models) preferences are not NP identified in dynamic models even when beliefs and discount factor are known.

• Instead of looking at the identification of preferences, I study the identification of behavioral and welfare effects of counterfactual policy changes.

• The contributions of the paper are:

(1) Identification results

(2) Estimation procedure and Nonparametric test.

(3) An Application to Public Pensions and Retirement Behavior

OUTLINE

1. MODEL AND ASSUMPTIONS

2. CLASS OF POLICY INTERVENTIONS

2. IDENTIFICATION RESULTS

4. ESTIMATION METHOD

5. APPLICATION

1. *MODEL AND EVALUATION PROBLEM*

• Consider a dynamic decision model. The optimal decision rule is:

$$\alpha(s_t) = \arg \max_{a \in \{0,1,...,J\}} \left\{ U(a, s_t) + \beta \int V(s_{t+1}) \, dF(s_{t+1}|a, s_t) \right\}$$

where $V(.)$ solves the Bellman equation:

$$V(s_t) = \max_{a \in \{0,1,...,J\}} \left\{ U(a, s_t) + \beta \int V(s_{t+1}) \, dF(s_{t+1}|a, s_t) \right\}$$

• The primitives of the model are $\{U, \beta, F\}$, which are nonparametrically specified.

• Suppose that the researcher has a random sample of individuals who behave according to this model.

$$Data = \{\ a_{it},\ x_{it},\ y_{it}\ : i = 1, 2, ..., N;\ t = 1, 2, ..., T_i\}$$

where x_t is a subvector of s_t; and $.y_t$ is an **outcome variable**.

• We are interested in using these data to estimate the effects on behavior (α function) and welfare (V function) of a counterfactual policy that modifies the utility function U.

• To complete the model we have to make some **assumptions on the unobservable state variables**.

- The vector of state variables s_{it} can be decomposed in 3 components:

$$s_{it} = (x_{it},\ \omega_{it},\ \xi_{it})$$

such as:

$x_{it} =$ Observable

$\omega_{it} =$ Unobservable in the outcome function: $y_{it} = Y(a_{it}, x_{it}, \omega_{it})$

$\xi_{it} =$ Other unobservables

ASSUMPTION 1: [Conditional Independence 1] Preferences are additive in the outcome variable y.

$$U(a_{it}, s_{it}) = Y(a_{it}, x_{it}, \omega_{it}) + C(a_{it}, x_{it}, \xi_{it})$$

where $Y()$ is the outcome function. Furthermore,

$$(\omega_{it} \perp \xi_{it}) \mid x_{it}$$

• Assumption 1 is needed for the NP identification of the distribution of the unobservables.

• In a semiparametric model where that distribution is parametric, Assumption 1 is not needed.

ASSUMPTION 2: [Conditional Independence 2]

- $\{\omega_{it}\}$ follows an exogenous first order Markov process such that:

$$\Pr\left(\omega_{t+1} \mid a_t, s_t\right) = F_\omega(\omega_{t+1}|\omega_t)$$

- $\{\xi_{it}\}$ follows as exogenous process such that:

$$\Pr(\xi_{t+1} \mid x_{t+1}, a_t, s_t) = F_{\xi|x}(\xi_{t+1}|x_{t+1})$$

- The transition of x_{it} is such that:

$$\Pr(x_{t+1} \mid a_t, s_t) = F_x(x_{t+1}|a_t, x_t, \omega_t)$$

ASSUMPTION 3: [Monotonicity]. The outcome function $Y(a_t, x_t, \omega_t)$ is strictly monotonic in ω_t, and ω_t is a continuous random variable.

• Again, Assumption 3 is only needed for the NP identification of the distribution of the unobservables.

- It will be convenient to define the function:

$$M(a, x_t) \equiv Median\{\ C(a, x_t, \xi_t) \mid x_t\}$$

- And the unobservable variables:

$$\varepsilon_t(a) \equiv C(a, x_t, \xi_t) - M(a, x_t)$$

and $\varepsilon_t \equiv \{\varepsilon_t(0), \varepsilon_t(1), ..., \varepsilon_t(J)\}$.

- Therefore, we can write:

$$U(a, s_t) \;=\; Y(a, x_t, \omega_t) + M(a, x_t) + \varepsilon_t(a)$$

where, by construction, $Median(\varepsilon_t(a)|x_t) = 0$.

• The set of structural functions that define the model is:

$$\left\{ Y,\ C,\ \beta,\ F_{\omega},\ F_{\xi|x},\ F_{x} \right\}$$

• We might be interested in the (semi) structural functions:

$$\{ Y,\ M,\ \beta,\ F_{\omega},\ F_{\varepsilon|x},\ F_{x} \}$$

• **Can we identify NP these primitives?** No.

• Rust (1994) and Magnac and Thesmar (Ectca, 2002).

• Instead, this paper considers a different question: **can we identify NP the effects of counterfactual policies?** Yes, for a class of counterfactual policies.

2. *POLICY INTERVENTIONS*

• Consider an hypothetical policy intervention that modifies the current utility function from the original function U to a new function U^*.

• The researcher does not know neither U nor U^*.

• Suppose that **the researcher knows the function**:

$$\tau(a,s) = U^*(a,s) - U(a,s)$$

• Furthermore, suppose that this function τ does not depend on ε:

$$\tau(a,s) = \tau(a,x,\omega)$$

• For this class of counterfactual policies, we show the NP identification of the counterfactual optimal decision rule (α^*) and the welfare change ($V^* - V$).

EXAMPLE 1: Retirement model.

- Policies that modify retirement benefits such as changes in the minimum and normal retirement age or changes in the discount for early retirement.

- Policies that affect labor earnings such as a wage tax; or an hypothetical change in the relative risk aversion parameter.

EXAMPLE 2: Model of educational choice.

- A change in returns to schooling

- A change in the costs of schooling.

EXAMPLE 3: Firms' investment and labor demand.

- Hypothetical changes in the production function parameters.

MEASURES OF THE POLICY EFFECTS

- Define the **choice probability function**:

$$P(a|x,\omega) \equiv \text{Pr}(\alpha(s_t) = a \mid x_t = x,\ \omega_t = \omega)$$

before the policy intervention.

- And let $P^*(a|x,\omega)$ be this choice probability function after the policy intervention. We are interested in the identification of $P^*(a|x,\omega)$.

- Define the **integrated value function:**

$$\bar{V}(x,\omega) = \int V(x,\omega,\varepsilon)\ F_{\varepsilon|x}(d\varepsilon)$$

- And let $\bar{V}^*(x,\omega)$ be this expected value function after the policy.

- We can identify $\bar{V}^*(x,\omega) - \bar{V}(x,\omega)$

3. $IDENTIFICATION: STATIC\ MODEL$ ($\beta = 0$)

• Suppose we have a sample of individuals: $\{a_i, x_i, y_i : i = 1, 2, ..., n\}$. $a_i \in \{0, 1\}$ is an action/decision by individual i.

• Since observed actions are the result of utility maximization, $a_i = 1$ iff:

$$
\begin{aligned}
& U(0, s_i) < U(1, s_i) \\
\Leftrightarrow \quad & Y(0, x_i, \omega_i) + M(0, x_i) + \varepsilon_i(0) < Y(1, x_i, \omega_i) + M(1, x_i) + \varepsilon_i(1) \\
\Leftrightarrow \quad & \varepsilon_i(0) - \varepsilon_i(1) < [Y(1, x_i, \omega_i) - Y(0, x_i, \omega_i)] + [M(1, x_i) - M(0, x_i)] \\
\Leftrightarrow \quad & \tilde{\varepsilon}_i < \tilde{Y}(x_i, \omega_i) + \tilde{M}(x_i)
\end{aligned}
$$

• Therefore, we have that:

$$P(x,\omega) \;=\; F_{\tilde{\varepsilon}|x}\Big(\, \tilde{Y}(x_i,\omega_i) + \tilde{M}(x_i) \,\Big)$$

• Similarly, under the counterfactual policy we have that $a_i = 1$ iff:

$$U^*(0,s_i) < U^*(1,s_i)$$

$$\Leftrightarrow \;\; \tilde{\varepsilon}_i < \tilde{Y}(x_i,\omega_i) + \tilde{M}(x_i) + \tilde{\tau}(x,\omega)$$

• Therefore:

$$P^*(x,\omega) \;=\; F_{\tilde{\varepsilon}|x}\Big(\, \tilde{Y}(x,\omega) + \tilde{M}(x) + \tilde{\tau}(x,\omega) \,\Big)$$

• It is clear that to identify $P^*(x,\omega)$ we have to identify $F_{\tilde{\varepsilon}|x}$, $\tilde{Y}$ and $\tilde{M}$. **Can we from the factual P?**

• **Sketch of the proof:**

(1) Based on the equation $y_i = Y(a_i, x_i, \omega_i)$ the outcome function Y can be identified under different conditions (conditional independence; control function; IV).

(2) Given the estimation of $Y()$, we can obtain consistent estimates of $\{\omega_i\}$.

(3) Then, we can estimate NP $P(x, \omega)$ using a NP regression of $E(a_i|x_i, \omega_i)$

(4) Given $x \in X$, define $\omega^*(x)$ as the value of ω that solves

$$P(x, \omega) = 0.5$$

This function $\omega^*(x)$ is identified for any x.

(5) By the zero conditional median of ε:

$$\tilde{Y}(x, \omega^*(x)) + \tilde{M}(x) = 0$$

and this identifies $\tilde{M}$.

(6) Now, for given $x \in X$ and given arbitrary real value u, define $\omega^*(x, u)$ as the value of ω that solves the equation:

$$\tilde{Y}(x, \omega) = u + \tilde{M}(x)$$

Then, by construction,

$$P(x, \omega^*(x, u)) = F_{\tilde{\varepsilon}|x}\left(\tilde{Y}(x, \omega^*(x, u)) + \tilde{M}(x) \right) = F_{\tilde{\varepsilon}|x}(u)$$

and $F_{\tilde{\varepsilon}|x}(u)$ is identified.

3. IDENTIFICATION: DYNAMIC MODEL

• Since observed actions are the result of the maximization of expected & discounted utility, $a_i = 1$ iff $V(1, s_i) > V(0, s_i)$

$$\Leftrightarrow \quad V^{(Y)}(0, x_i, \omega_i) + V^{(M)}(0, x_i, \omega_i) + \varepsilon_i(0) < V^{(Y)}(1, x_i, \omega_i) + V^{(M)}(1, x_i, \omega_i)$$

$$\Leftrightarrow \quad \tilde{\varepsilon}_i < \tilde{V}^{(Y)}(x_i, \omega_i) + \tilde{V}^{(M)}(x_i, \omega_i)$$

• We have that:

$$P(x, \omega) \;\; = \;\; F_{\tilde{\varepsilon}|x}\left(\; \tilde{V}^{(Y)}(x, \omega) + \tilde{V}^{(M)}(x, \omega) \; \right)$$

• However, it is not true that:

$$P^{*}(x, \omega) \;\; = \;\; F_{\tilde{\varepsilon}|x}\left(\; \tilde{V}^{(Y)}(x, \omega) + \tilde{V}^{(M)}(x, \omega) + \tilde{V}^{(\tau)}(x, \omega) \; \right)$$

Identification of $F_{\tilde{\varepsilon}|x}$, $\tilde{V}^{(Y)}$ and $\tilde{V}^{(M)}$ is enough.

- I obtain the following decomposition:

$$P(x,\omega) = F_{\tilde{\varepsilon}|x}\left(\tilde{V}_{10}^{(Y)}(x,\omega) + \tilde{V}_{10}^{(M)}(x) + \tilde{V}_{OPT}(x,\omega,P) \right)$$

- $\tilde{V}_{10}^{(Y)}$ is the difference between two expected-discounted values of current and future $Y's$: the value if current choice is $a = 1$ and then $a = 0$ is chosen forever in the future, minus the value of choosing $a = 0$ today and forever in the future.

- $\tilde{V}_{10}^{(M)}$ has a similar definition but for component M.

- $\tilde{V}_{OPT}$ is the difference between two expected-discounted values of future $(Y + M)'s$: the value of behaving according to P, minus the value of choosing $a = 0$ today and forever in the future.

• This decomposition has some **interesting properties**.

• **Property 1:** Given the function Y, the value functions $\tilde{V}_{10}^{(Y)}(x,\omega)$ is known and it is strictly monotonic in ω.

• **Property 2:** The form of the function $\tilde{V}_{OPT}(x,\omega,P)$ only depends on P and on $F_{\tilde{\varepsilon}|x}$.

• **Property 3:** $\tilde{V}_{10}^{(M)}$ depends on x but not on ω.

• An **implication of Properties 1-3** is that, given $\tilde{V}_{10}^{(Y)}$ and P, the restrictions

$$P(x,\omega) = F_{\tilde{\varepsilon}|x}\left(\tilde{V}_{10}^{(Y)}(x,\omega) + \tilde{V}_{10}^{(M)}(x) + \tilde{V}_{OPT}(x,\omega,P) \right)$$

identify the functions $F_{\tilde{\varepsilon}|x}$ and $\tilde{V}_{10}^{(M)}$.

• **Property 4:** Functions $\tilde{V}_{10}^{(Y)}$ and $\tilde{V}_{10}^{(M)}$ do not depend on optimal behavior. The counterfactual version of these functions is such that:

$$\tilde{V}_{10}^{(Y)*}(x,\omega) + \tilde{V}_{10}^{(M)*}(x) = \tilde{V}_{10}^{(Y)}(x,\omega) + \tilde{V}_{10}^{(M)}(x) + \tilde{V}_{10}^{(\tau)}(x)$$

where $\tilde{V}_{10}^{(\tau)}$ is known given the policy τ.

• An **implication of Properties 1-4** is that:

$$P^*(x,\omega) = F_{\tilde{\varepsilon}|x}\left(\tilde{V}_{10}^{(Y)}(x,\omega) + \tilde{V}_{10}^{(M)}(x) + \tilde{V}_{10}^{(\tau)}(x) + \tilde{V}_{OPT}(x,\omega,P^*) \right)$$

• Given $F_{\tilde{\varepsilon}|x}$, $\tilde{V}_{10}^{(Y)}$, $\tilde{V}_{10}^{(M)}$ and $\tilde{V}_{10}^{(\tau)}$, this equation describes a contraction mapping in P^*. Therefore, P^* is identified.

• Though we can identify $F_{\tilde{\varepsilon}|x}$ and $\tilde{V}_{10}^{(M)}(x)$, we cannot identify $F_{\xi|x}$ or $\tilde{M}(x) \equiv M(1,x) - M(0,x)$.

• The function $\tilde{V}_{10}^{(M)}$ depends both on the function M and on F_x. It is a linear combination of values of M.

• This linear combination of values of M is all what we need to know to evaluate this type of policy. But that will not be enough to evaluate other counterfactual policies: for instance, a change in F_x.

• Given an estimate of P^* based on a parametric model, we can use our NP estimator of P^* to test for the **validity of the parametric assumptions in the context of this particular policy evaluation**.

5. APPLICATION: PUBLIC PENSIONS AND RETIREMENT

• We have a panel dataset $\{a_{it}, x_{it}, y_{it}\}$ from Sweden of N individuals older than 50 such that the time frequency is annual and:

• $a_{it} \in \{0, 1\}$ where $a_{it} = 1$ represents "individual i is still working at age t"

• y_{it} = Individual earnings: labor earnings if $a_{it} = 1$ and public pension earnings if $a_{it} = 0$.

• $x_{it} = (t_i, m_{it}, ra_{it}, pp_{it})$ with

t_i = Age

m_{it} = Marital status

ra_{it} = Age at retirement

pp_{it} = Pension wealth (from social security records)

• We want to use these data to evaluate the effects on retirement behavior of an hypothetical change in the schedule of public pension benefits.

• Our (simplified) description of the Swedish public pension system consists of 2 functions:

• The **transition rule of pension wealth** when working:

$$F_{pp}(pp_{t+1} \mid pp_{it},\ y_{it}, t_i, m_{it},)$$

Different authors have shown that a flexible specification of this function can represent very well the accumulation of pension wealth in many different countries (e.g., US, Sweden, Germany, etc).

• **Pension benefits function**:

$$B_{it} = B(pp_{it}, ra_{it}, t_i)$$

- **Public Pensions in Sweden:** $B_t = B(pp_t, ra_t, t)$.

- For $t < 60$, $B_t = 0$.

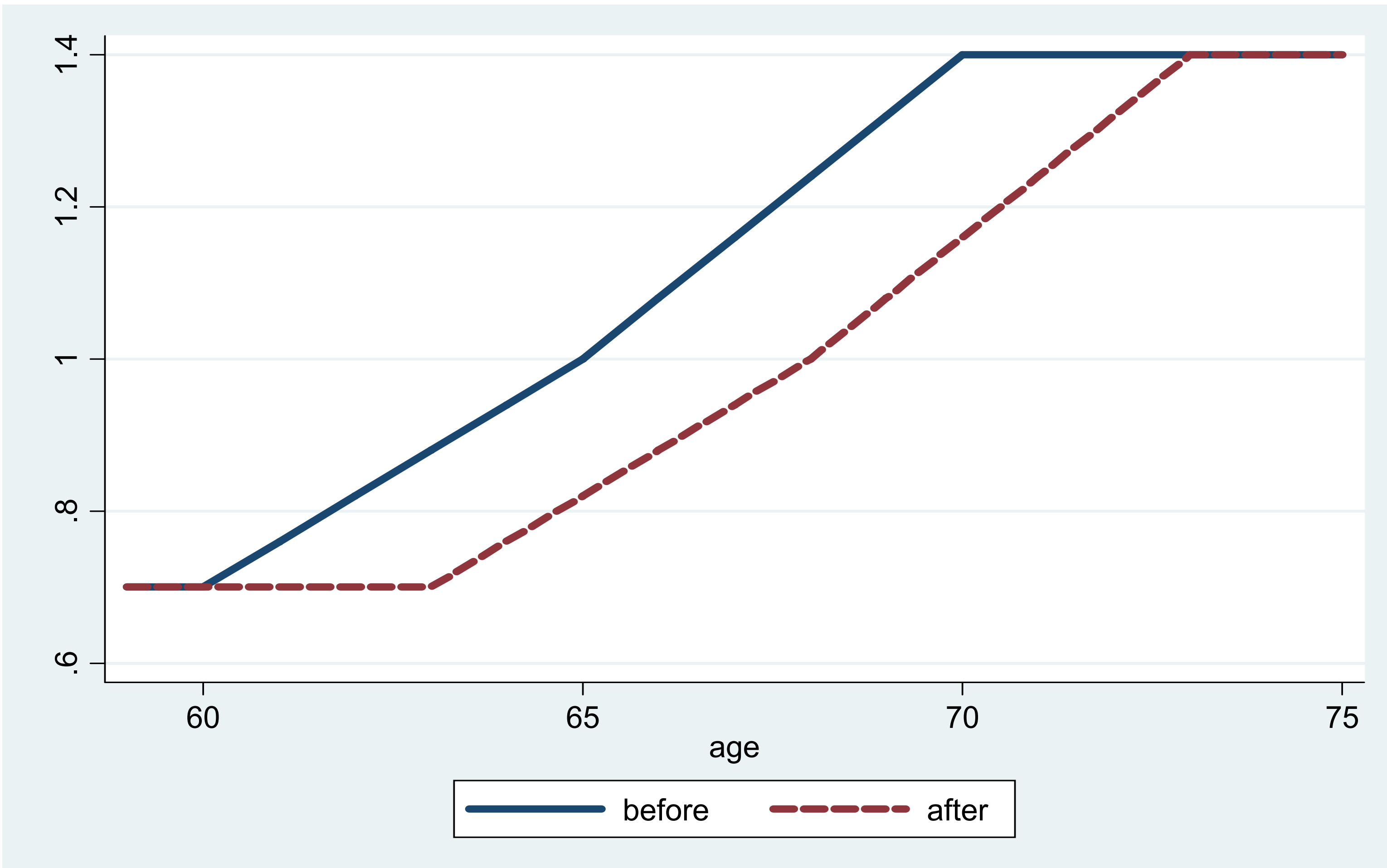
- For $t \geq 60$:

$$B_t = \begin{cases} pp_t & if \quad ra_t < 60 \\ pp_t \; (1 + \kappa_1 (ra_t - 65)) & if \quad 60 \leq ra_t < 65 \\ pp_t \; (1 + \kappa_2 (ra_t - 65)) & if \quad 65 \leq ra_t < 70 \\ pp_t \; (1 + \kappa_2 \; (70 - 65)) & if \quad ra_t \geq 70 \end{cases}$$

$\kappa_1 = 6.0\%$ (Discount for early retirement)

$\kappa_2 = 8.4\%$ (Premium for late retirement)

Figure 1. Age Profile of Pension Benefits Before and After the Reform



MODEL AND ASSUMPTIONS:

ASSUMPTION AP-1: The one-period utility is equal to expected current earnings plus the utility of leisure. Leisure takes only two values: when working and when not.

$$U_t = Y^e(a_t, x_t, \omega_t) + C(a_t, x_t, \xi_t)$$

where:

$\xi_t =$ Shock in the utility of leisure

$\omega_t =$ Shock in labor earnings

Furthermore,

$$(\omega_{it} \perp \xi_{it}) \mid x_{it}$$

ASSUMPTION AP-2: [Earnings function]

$$Y^e(a_t, x_t, \omega_t) = E\left(Y_t \mid a_t, x_t, \omega_t\right)$$

where:

$$Y_t = \begin{cases} B(pp_t, ra_t, t) & if \quad a_t = 0 \\ \\ W_t = \exp\left\{h_W(x_t) + \omega_{t+1}\right\} & if \quad a_t = 1 \end{cases}$$

and ω_{t+1} follows a Markov process, and the innovation is unknown to the individual at time t.

ESTIMATION OF WAGE EQUATION

- We observe labor earnings W_{it} only if $a_{it} = 1$ (selection problem).

$$\log(W_{it}) = h_W(x_{it}) + \omega_{i,t+1} \qquad \text{if } a_t = 1$$

- Given Assumption AP-2:

$$\omega_{i,t+1} = \rho(\omega_{it}) + v_{i,t+1} \qquad \text{where} \quad E(v_{i,t+1}|a_{it}, s_{it}) = 0$$

- Then,

$$\begin{aligned} \ln W_{it} &= h_W(x_{it}) + \rho(\omega_{it}) + v_{i,t+1} && \text{if } a_{it} = 1 \\ &= h_W(x_{it}) + \rho(\ln W_{t-1} - h_W(x_{i,t-1})) + v_{i,t+1} && \text{if } a_t = 1 \end{aligned}$$

DATA

- Blue-collar workers in Sweden who were born between 1927 and 1940.
- Observation period is 1983 to 1997.
- Working sample has 3,129 individuals and more than 31,000 observations.

Table 1
Summary Statistics
3,129 male blue-collar workers. Cohorts 1927-1940
Years 1983 to 1997

Variable	Mean	Std. Dev.	Min	Max	# Obs.
Married	0.711	0.453	0	1	34,593
Annual Labor Earnings[(1)]	195.8	52.2	1.0	1,591.0	34,593
Pension Points	4.46	0.88	0.13	6.50	34,593
Retirement Age[(2)]	63.47	2.39	51	69	834

(1) In thousands of Swedish Kronas. In 1997, US$ 1 $\simeq$ 8 Swedish Kronas.

(2) Subsample of complete (uncensored) histories.

Table 2
Estimation of Wage Equation and Stochastic Process of Wage Shock

Parameter[1]	Estimate	(Std. Error)
$h_{W,0}$ (constant)	4.7465	(0.2041)
$h_{W,1}$ (t)	0.0174	(0.0072)
$h_{W,2}$ (t^2)	-0.00014	(0.00006)
ρ_1	0.9161	(0.0037)
ρ_2	0.0019	(0.0044)
ρ_3	-0.0299	(0.0012)
$Std.Dev.\ \xi$	0.1607	
R-square	0.7160	
# Observations	30,630	

Table 3
Estimation of Stochastic Process of Pension Points

Parameter	Estimate	(Std. Error)
constant	0.0287	(0.0155)
log(pp)	0.9630	(0.0049)
log(pp)^2	0.0161	(0.0004)
age	0.0009	(0.0005)
age^2	-6.64×10^{-6}	(4.53×10^{-6})
log(pp) $\times$ *age*	-0.0003	(0.0001)
$Std.Dev.\ \eta$	0.0116	
R-square	0.9971	
# Observations	30,630	

Figure 2. Histogram for Retirement Age (subsample individuals who retire during the sample period)

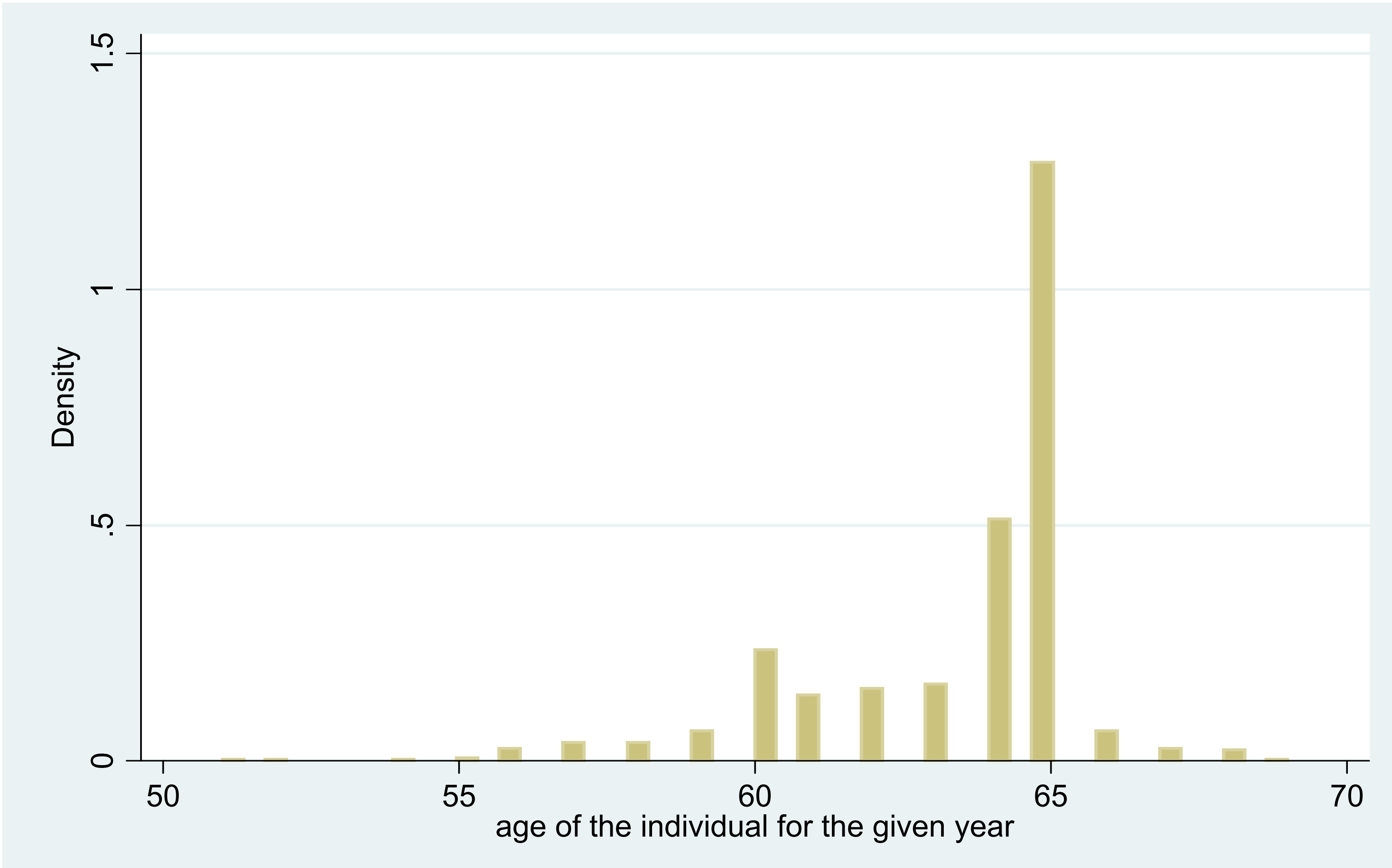


Figure 3. Estimated Function $\mathrm{h_W(t)}$

Age Profile of Labor Earnings

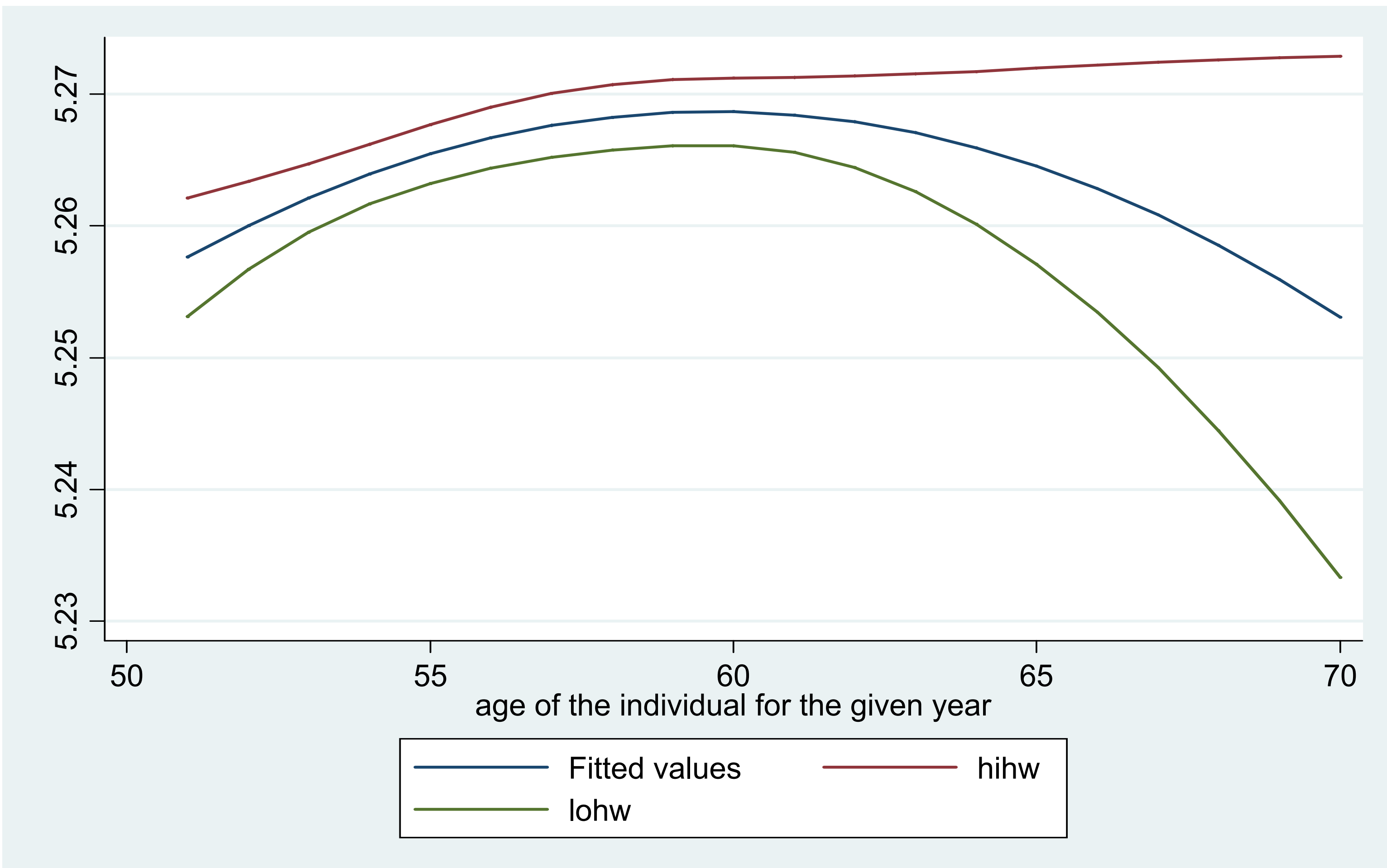


Figure 4. Estimated Function $\rho(\omega)$

Stochastic Process of Labor Earnings Shock

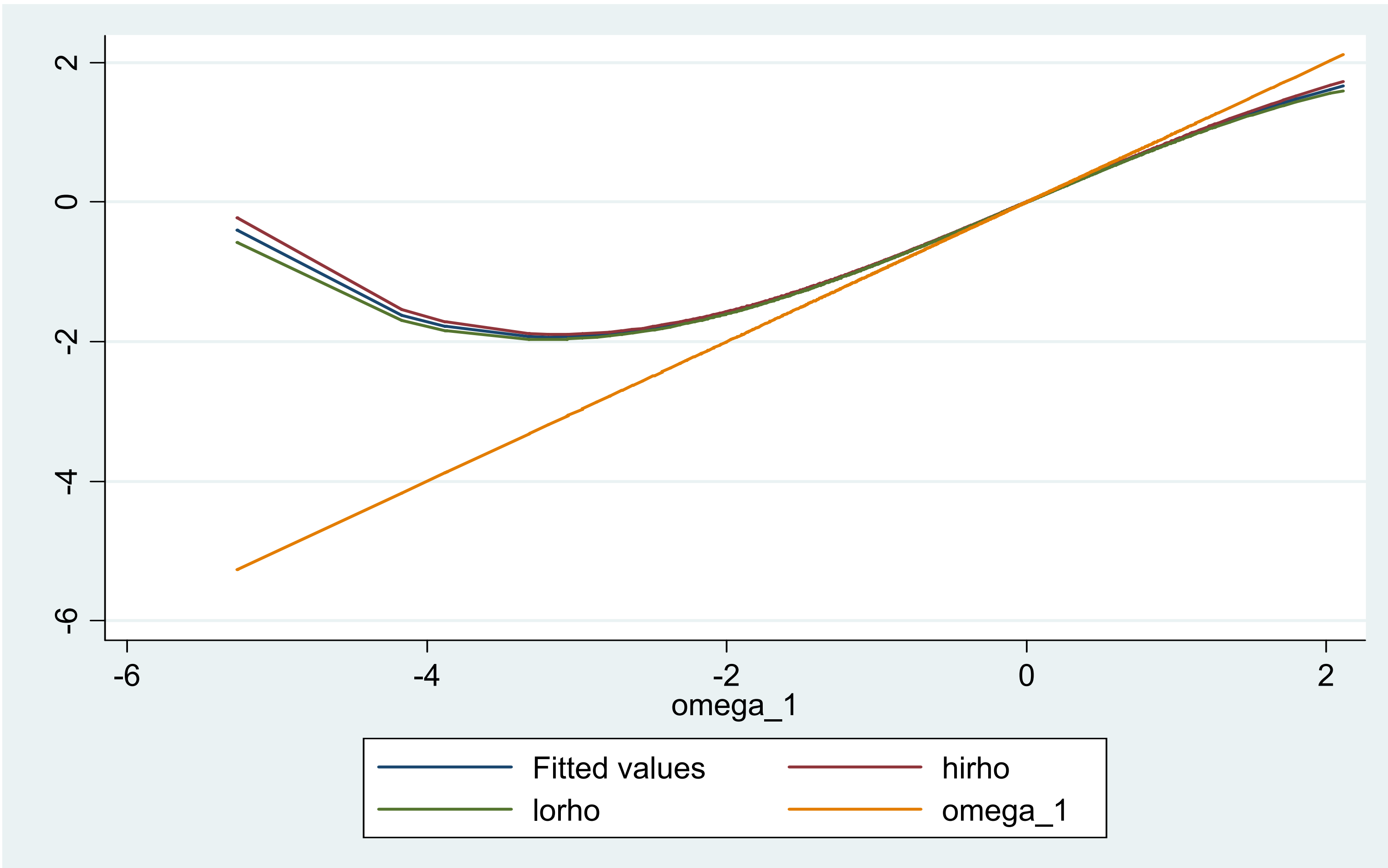


Figure 5. Estimated Density of ξ

Innovation of the labor earnings shock

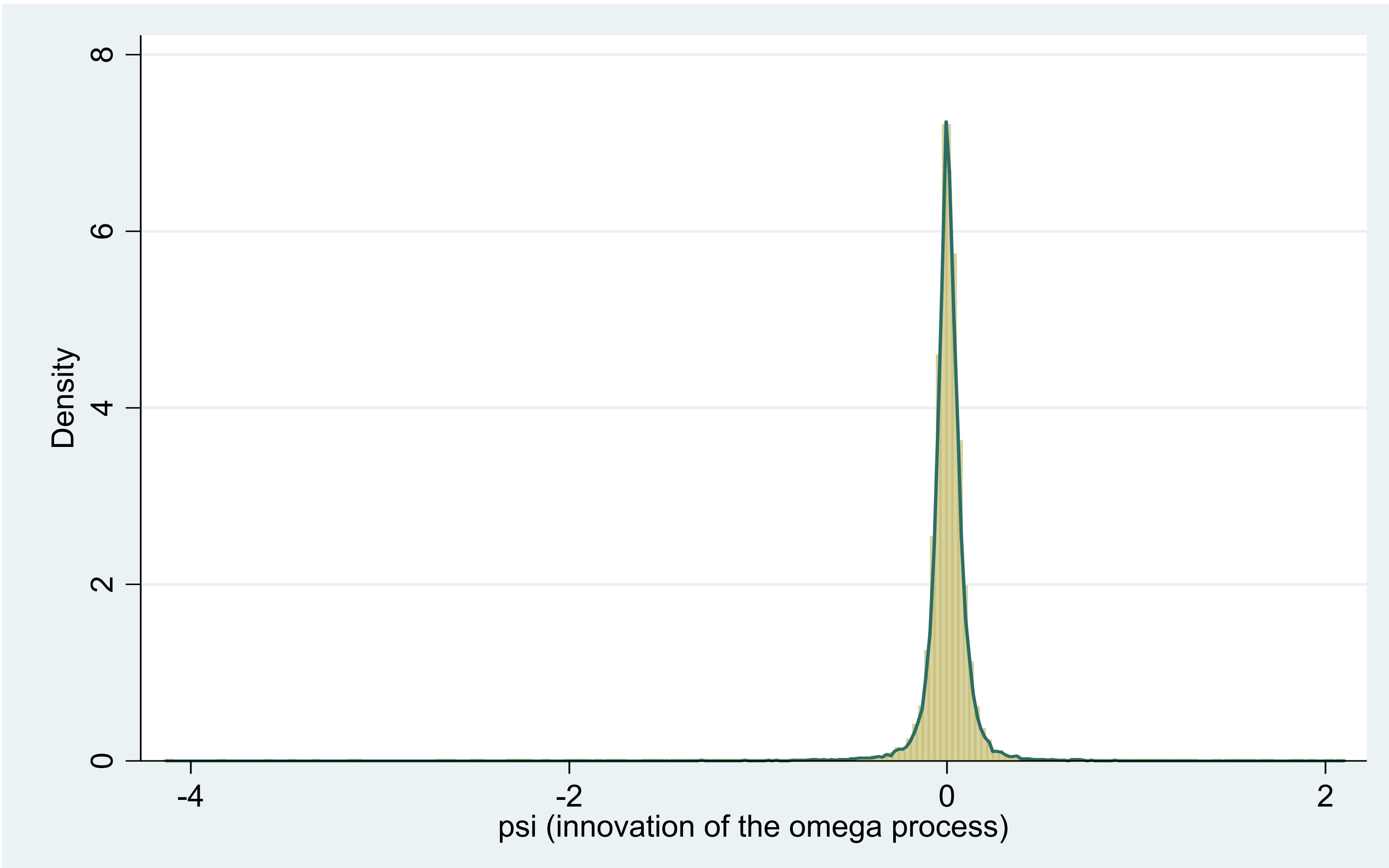


Figure 6. Estimated Density of ω

Labor earnings shock

Figure 8. Prob. Working vs. age

(evaluated at mean values of m_t, pp_t and ω_t)

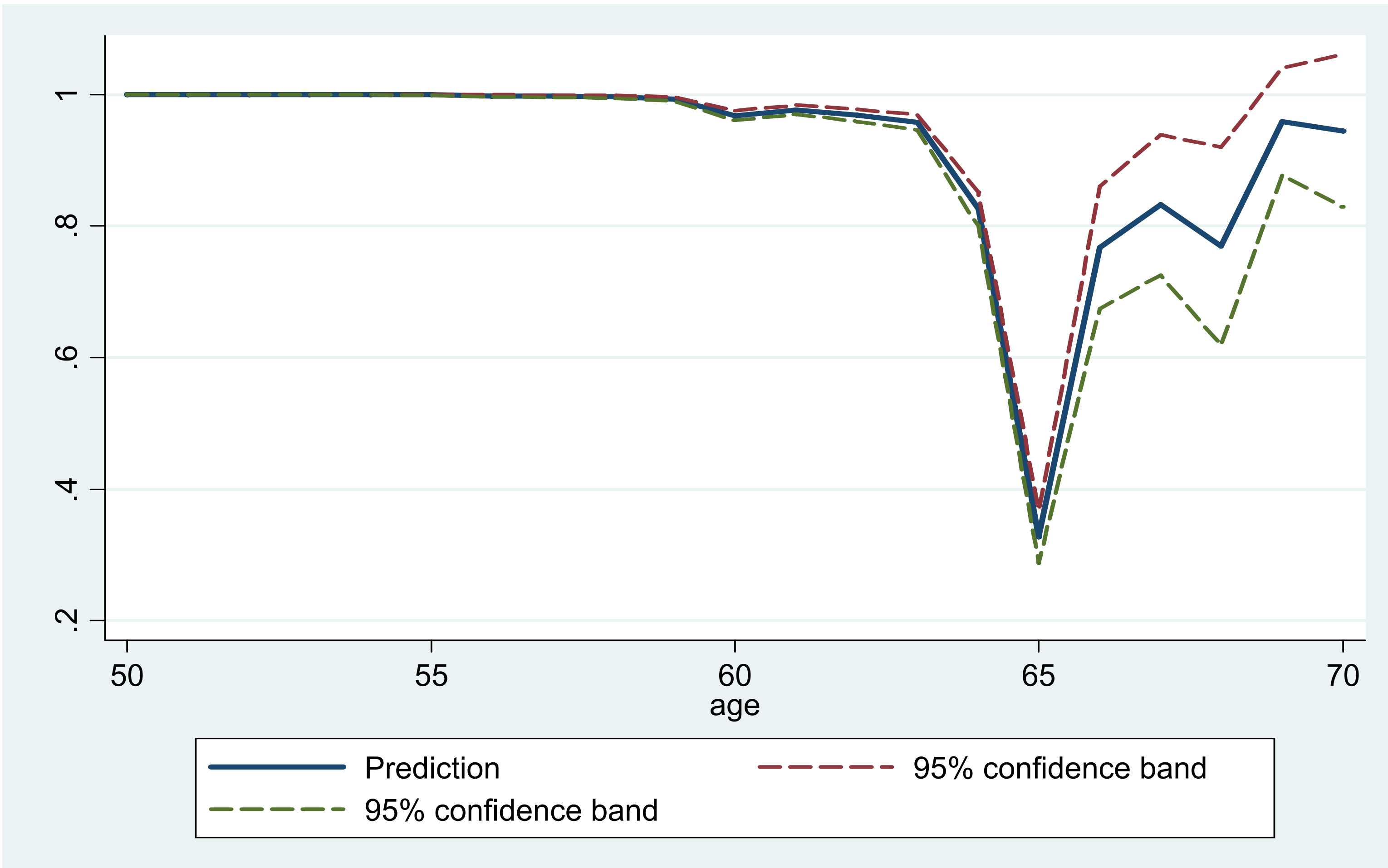


Figure 9. Prob. Working vs. ω_t

(evaluated at Age=65 and at mean values of m_t, pp_t)

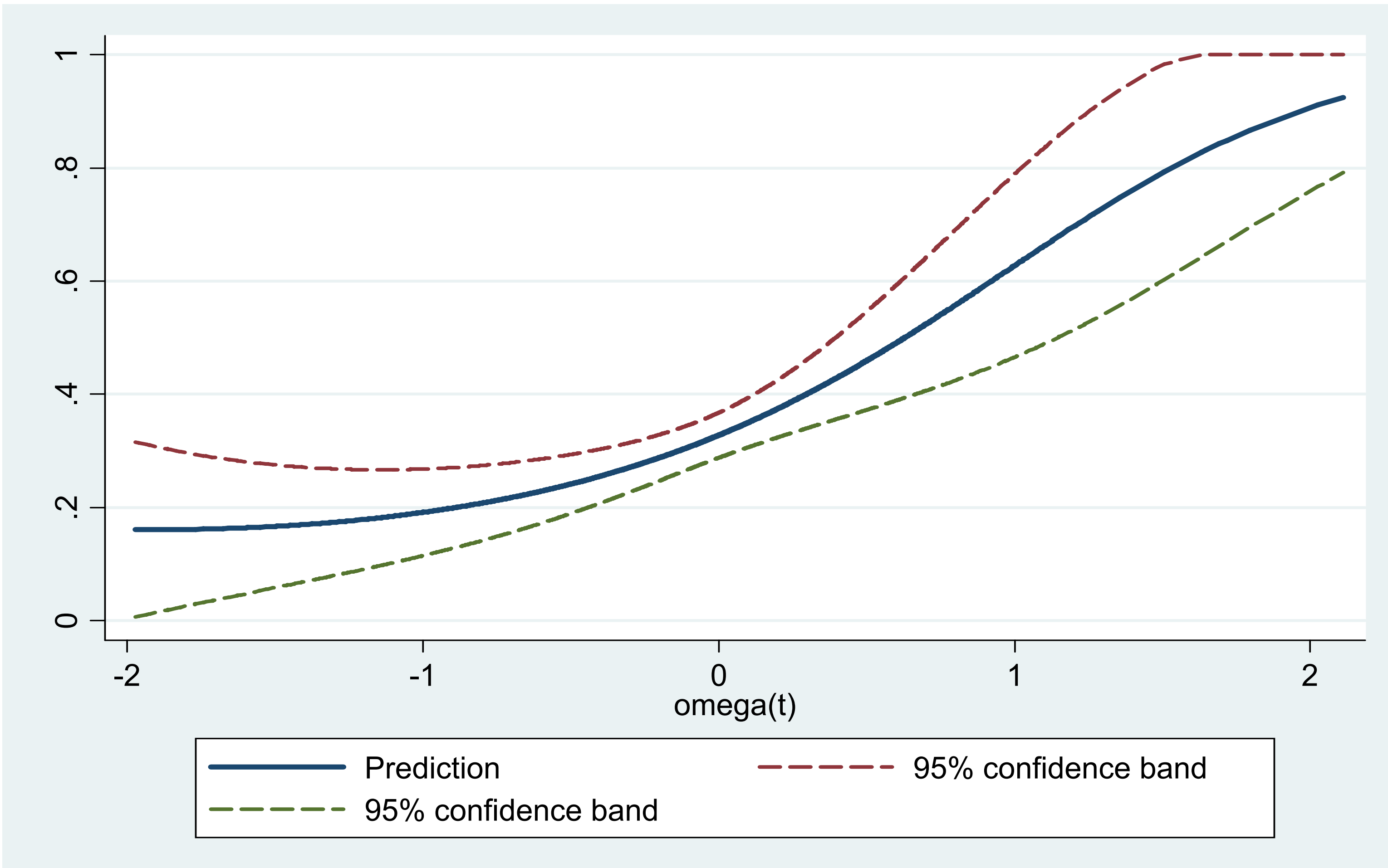
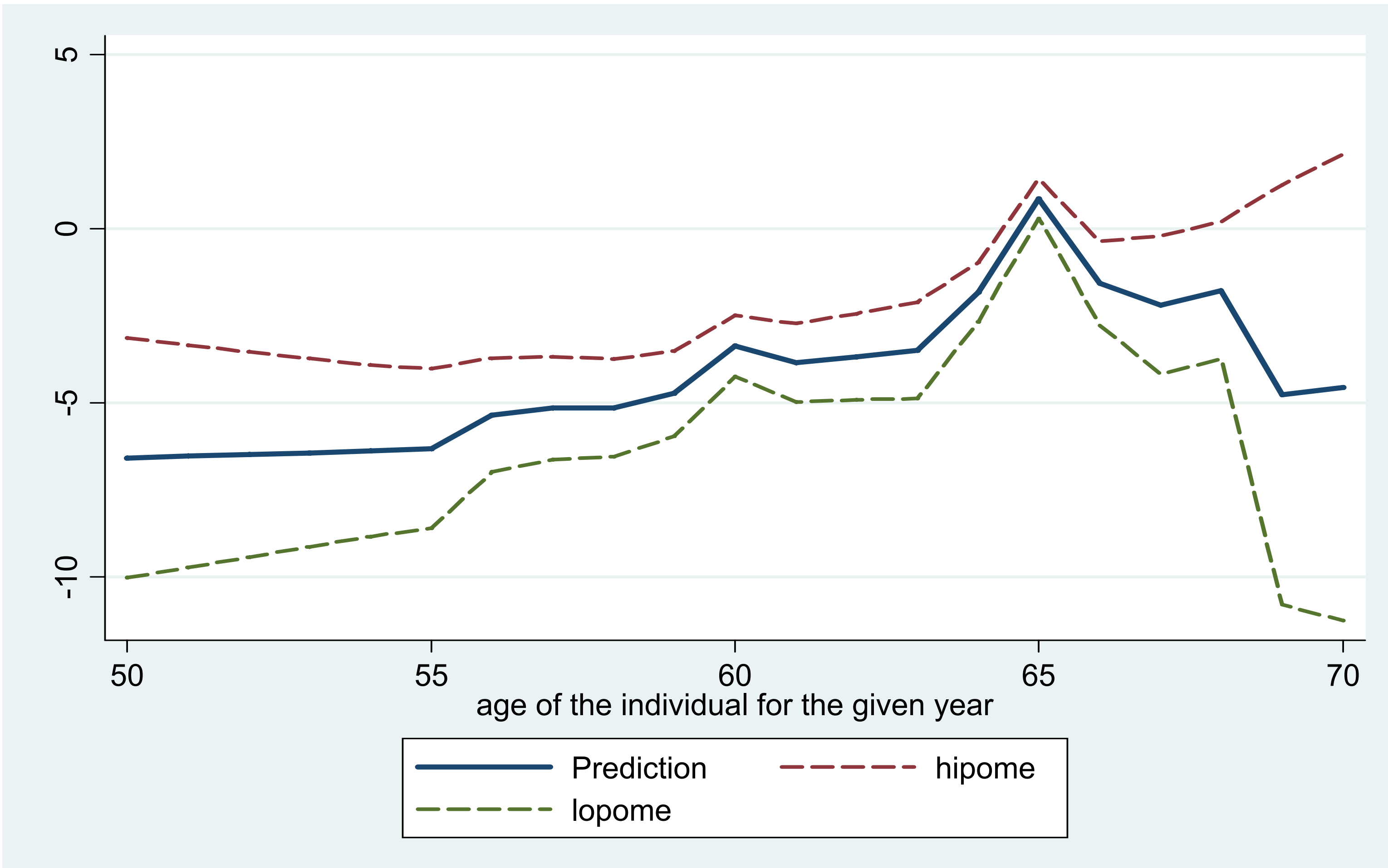


Figure 11. Function $\omega^*(m, t, pp, 0)$ vs. age

(evaluated at mean values of m_t and pp_t)



Policy Evaluation: Nonparametric and Parametric Models Bootstrap Standard Errors in Parentheses		
	Nonparametric	**Parametric**
Δ Retirement Age	1.232 (0.363)	2.108 (0.202)
Δ Present Value Life-Time Earnings	-4.76% (1.07)	-1.88% (0..62)