Online Appendix to "Worker Sorting and Agglomeration Economies"

Theodore Papageorgiou

McGill University, Department of Economics[*]

March 2018

1 Model with Occupation-Specific Human Capital

In this appendix I derive optimal worker behavior when workers also accumulate occupation-specific human capital.

There are two experience levels in each occupation: workers are either experienced or inexperienced. A worker who enters an occupation for the first time is inexperienced. As in Kambourov and Manovskii (2009), a worker becomes experienced at rate θ. Output production is increased by amount u for experienced workers.

In order to solve the worker's occupational choice problem, as in the baseline model, I follow Whittle (1980, 1982) and Karatzas (1984). More specifically, I compute the retirement value at which the worker is exactly indifferent between continuing with that occupation or retiring. This retirement value serves as an index for each occupation, which corresponds to that occupation's Gittins index (see Gittins and Jones (1974) and Bergemann and Välimäki (2008)).

I start by deriving the optimal behavior of an experienced worker. The solution in this case is similar to the baseline model.

The optimal stopping rule is to retire when p reaches $\underline{p}(W)$ such that the value matching and the smooth pasting conditions hold

$$V^{\exp}\left(\underline{p}(W), W\right) = W \tag{1}$$

$$V_p^{\exp}\left(\underline{p}(W), W\right) = 0.$$

[*]855 Sherbrooke St West, Room 519, Montreal, QC H3A 2T7, Canada. E-mail: *theodore.papageorgiou@mcgill.ca* Web: *http://www.theodorepapageorgiou.com*

The solution to the above differential equation is given by

$$
\begin{aligned}
V^{\exp}(p,W) &= \frac{\alpha_G p + \alpha_B (1-p) + u + \delta J}{r+\delta} \\
&+ \frac{\alpha_G - \alpha_B}{r+\delta}\left(\underline{p}(W) + \frac{1}{2}d - \frac{1}{2}\right)^{-1} \underline{p}(W)^{\frac{1}{2}+\frac{1}{2}d}\left(1-\underline{p}(W)\right)^{\frac{1}{2}-\frac{1}{2}d} \\
&\times (p)^{\frac{1}{2}-\frac{1}{2}d}(1-p)^{\frac{1}{2}+\frac{1}{2}d},
\end{aligned} \tag{2}
$$

where

$$
\underline{p}_{ex}(W) = \frac{(d-1)\left((r+\delta)W - \alpha_B - u - \delta J\right)}{(d+1)(\alpha_G - \alpha_B) - 2\left((r+\delta)W - \alpha_B - u - \delta J\right)}, \tag{3}
$$

and $d = \sqrt{\frac{8(r+\delta)}{\left(\frac{\alpha_G - \alpha_B}{\sigma}\right)^2} + 1}$.

The index of each occupation is the highest retirement value at which the worker is indifferent between working at that occupation or retiring with $W = W(p)$, i.e.

$$
W(p) = V^{\exp}(p,W). \tag{4}
$$

For eq. (4) to hold, from eq. (1), it must be the case that

$$
p = \underline{p}(W). \tag{5}
$$

Substituting condition (5) into the threshold condition, equation (3), obtains

$$
\begin{aligned}
p &= \frac{(d-1)\left((r+\delta)W(p) - \alpha_B - u - \delta J\right)}{(d+1)(\alpha_G - \alpha_B) - 2\left((r+\delta)W(p) - \alpha_B - u - \delta J\right)} \Rightarrow \\
W(p) &= \frac{1}{r+\delta}\frac{(d+1)(\alpha_G - \alpha_B)p + (2p+d-1)(\alpha_B + u + \delta J)}{2p+d-1}.
\end{aligned} \tag{6}
$$

I now turn to the optimal behavior of an inexperienced worker.

The value of an inexperienced worker with the option of retiring and obtaining value W is given by

$$
\begin{aligned}
rV^{in}(p,W) &= \alpha_G p + \alpha_B (1-p) \\
&+ \frac{1}{2}\left(\frac{\alpha_G - \alpha_B}{\sigma}\right)^2 p^2 (1-p)^2 V^{in}_{pp}(p,W) \\
&- \delta\left(V^{in}(p,W) - J\right) - \theta\left(V^{in}(p,W) - V^{\exp}(p,W)\right).
\end{aligned} \tag{7}
$$

Substituting in the expression for $V^{\exp}(p,W)$ which was derived above (equations (2) and (3)) leads to

$$
\begin{aligned}
& (r+\theta+\delta) V^{in}(p,W) \\
= \quad & \frac{r+\delta+\theta}{r+\delta}(\alpha_G-\alpha_B)p+\frac{r+\delta+\theta}{r+\delta}\alpha_B \\
& +\frac{1}{2}\left(\frac{\alpha_G-\alpha_B}{\sigma}\right)^2 p^2(1-p)^2 V_{pp}^{in}(p,W) \\
& +\theta\frac{\alpha_G-\alpha_B}{r+\delta}\left(\underline{p}_{ex}(W)+\frac{1}{2}d-\frac{1}{2}\right)^{-1}\underline{p}_{ex}(W)^{\frac{1}{2}+\frac{1}{2}d}\left(1-\underline{p}_{ex}(W)\right)^{\frac{1}{2}-\frac{1}{2}d} \\
& \times p^{\frac{1}{2}-\frac{1}{2}d}(1-p)^{\frac{1}{2}+\frac{1}{2}d} \\
& +\frac{r+\delta+\theta}{r+\delta}\delta J+\frac{\theta}{r+\delta}u
\end{aligned}
$$

$$
\begin{aligned}
& \frac{1}{2(r+\theta+\delta)}\left(\frac{\alpha_G-\alpha_B}{\sigma}\right)^2 p^2(1-p)^2 V_{pp}^{in}(p,W) \\
= \quad & V^{in}(p,W) \\
& -\frac{1}{r+\delta}(\alpha_G-\alpha_B)p \\
& -\frac{\theta}{r+\theta+\delta}\frac{\alpha_G-\alpha_B}{r+\delta}\left(\underline{p}_{ex}(W)+\frac{1}{2}d-\frac{1}{2}\right)^{-1}\underline{p}_{ex}(W)^{\frac{1}{2}+\frac{1}{2}d}\left(1-\underline{p}_{ex}(W)\right)^{\frac{1}{2}-\frac{1}{2}d} \\
& \times p^{\frac{1}{2}-\frac{1}{2}d}(1-p)^{\frac{1}{2}+\frac{1}{2}d} \\
& -\left(\frac{1}{r+\delta}\alpha_B+\frac{\delta}{r+\delta}J+\frac{1}{(r+\theta+\delta)}\frac{\theta}{r+\delta}u\right),
\end{aligned}
$$

subject to

$$V^{in}\left(\underline{p}(W),W\right)=W \tag{8}$$

$$V_p^{in}\left(\underline{p}(W),W\right)=0, \tag{9}$$

which is a differential equation of the form

$$k_3 t^2(1-t)^2\ddot{y}(t)=y(t)+k_1 t+g(W)t^{\beta}(1-t)^{1-\beta}+k_2,$$

where

$$y(t)=V^{in}(p,W)$$

$$k_1 = -\frac{1}{r+\delta}\left(\alpha_G - \alpha_B\right)$$

$$k_2 = -\left(\frac{\alpha_B + \delta J}{r+\delta} + \frac{1}{\left(r+\theta+\delta\right)}\frac{\theta}{r+\delta}u\right)$$

$$k_3 = \frac{1}{2r+\theta+\delta}\left(\frac{\alpha_G - \alpha_B}{\sigma}\right)^2$$

$$\begin{aligned} g\left(W\right) &= \frac{-\frac{\theta}{r+\theta+\delta}\frac{\alpha_G-\alpha_B}{r+\delta}}{\frac{(d-1)((r+\delta)W-\alpha_B-u-\delta J)}{(d+1)(\alpha_G-\alpha_B)-2((r+\delta)W-\alpha_B-u-\delta J)} + \frac{1}{2}d - \frac{1}{2}} \times \\ & \left(\frac{\left(d-1\right)\left(\left(r+\delta\right)W - \alpha_B - u - \delta J\right)}{\left(d+1\right)\left(\alpha_G-\alpha_B\right) - 2\left(\left(r+\delta\right)W - \alpha_B - u - \delta J\right)}\right)^{\frac{1}{2}+\frac{1}{2}d} \times \\ & \left(1 - \frac{\left(d-1\right)\left(\left(r+\delta\right)W - \alpha_B - u - \delta J\right)}{\left(d+1\right)\left(\alpha_G-\alpha_B\right) - 2\left(\left(r+\delta\right)W - \alpha_B - u - \delta J\right)}\right)^{\frac{1}{2}-\frac{1}{2}d} \end{aligned}$$

$$d = \sqrt{\frac{8\left(r+\delta\right)}{\left(\frac{\alpha_G-\alpha_B}{\sigma}\right)^2} + 1}$$

$$\beta = \frac{1}{2} - \frac{1}{2}\sqrt{\frac{8\left(r+\delta\right)}{\left(\frac{\alpha_G-\alpha_B}{\sigma}\right)^2} + 1}$$

subject to

$$y\left(t_0\right) = W$$

$$\dot{y}\left(t_0\right) = 0.$$

In order to solve the equation, I follow Brockett (1970), Chapter 1.

Let

$$\begin{aligned} x_1\left(t\right) &= y\left(t\right) \\ x_2\left(t\right) &= \dot{y}\left(t\right). \end{aligned}$$

Then

$$\begin{aligned} \dot{x}_1\left(t\right) &= x_2\left(t\right) \\ \dot{x}_2\left(t\right) &= \ddot{x}\left(t\right) = \frac{1}{k_3 t^2\left(1-t\right)^2}x_1\left(t\right) + Z\left(t\right), \end{aligned}$$

where

$$Z(t) = \frac{1}{k_3 t^2 (1-t)^2} \left[k_1 t + g(W) t^\beta (1-t)^{1-\beta} + k_2 \right].$$

Moreover let

$$x(t) = \begin{bmatrix} x_1(t) & x_2(t) \end{bmatrix}'.$$

Then

$$\dot{x}(t) = \begin{bmatrix} 0 & 1 \\ \frac{1}{k_3 t^2 (1-t)^2} & 0 \end{bmatrix} x(t) + \begin{bmatrix} 0 \\ 1 \end{bmatrix} Z(t).$$

I now use the "Variation of Constants formula". According to Brockett (1970), Chapter 1.6, if $\Phi(t, t_0)$ is the transition matrix for $\dot{x}(t) = A(t) x(t)$, then the unique solution of $\dot{x}(t) = A(t) x(t) + f(t)$; $x(t_0) = x_0$ is given by

$$x(t) = \Phi(t, t_0) x_0 + \int_{t_0}^{t} \Phi(t, \tau) \begin{bmatrix} 0 \\ 1 \end{bmatrix} Z(\tau) \, d\tau,$$

where $\Phi(t, t_o)$ is the transition matrix. The transition matrix is defined as the solution Φ of the matrix differential equation (Brockett (1970), Chapter 1.3)

$$\begin{aligned} \frac{d}{dt} \Phi(t, t_o) &= A(t) \Phi(t, t_o) \\ \Phi(t_0, t_o) &= I, \end{aligned}$$

where I is the identity matrix.

For the problem at hand, I have

$$\dot{x}(t) = A(t) x(t),$$

or

$$\dot{x}(t) = \begin{bmatrix} 0 & 1 \\ \frac{1}{k_3 t^2 (1-t)^2} & 0 \end{bmatrix} x(t).$$

Moreover let

$$
\begin{aligned}
x_0 &= \begin{bmatrix} x_{10} & x_{20} \end{bmatrix}' \\
x_{10} &= y(t_0) \\
x_{20} &= \dot{y}(t_0).
\end{aligned}
$$

The solution to the homogeneous differential equation

$$\ddot{y}(t) = \frac{1}{k_3 t^2 (1-t)^2} y(t),$$

is given by

$$y(t) = c_1 t^q (1-t)^{1-q} + c_2 t^{1-q} (1-t)^q,$$

where $q = \frac{1}{2}\left(1 - \sqrt{\frac{4+k_3}{k_3}}\right)$ and c_1 and c_2 are undetermined coefficients.

Given that I know the solution to the homogeneous, I can write

$$\Phi(t, t_0) x_0 = \begin{bmatrix} t^q (1-t)^{1-q} & t^{1-q} (1-t)^q \\ \frac{d}{dt}\left(t^q (1-t)^{1-q}\right) & \frac{d}{dt}\left(t^{1-q} (1-t)^q\right) \end{bmatrix} \cdot \begin{bmatrix} c_1 \\ c_2 \end{bmatrix}$$

$$\Phi(t, t_0) x_0 = \begin{bmatrix} t^q (1-t)^{1-q} & t^{1-q} (1-t)^q \\ t^{q-1} (1-t)^{-q} (q-t) & t^{-q} (1-t)^{q-1} (1-q-t) \end{bmatrix} \cdot \begin{bmatrix} c_1 \\ c_2 \end{bmatrix},$$

where c_1 and c_2 are undetermined coefficients.

I next solve for c_1 and c_2 as a function of x_0.

At $t = t_0$, since $\Phi(t_0, t_0) = I$, I have

$$
\begin{aligned}
\begin{bmatrix} x_{10} \\ x_{20} \end{bmatrix} &= \begin{bmatrix} t_0^q (1-t_0)^{1-q} & t_0^{1-q} (1-t_0)^q \\ t_0^{q-1} (1-t_0)^{-q} (q-t_0) & t_0^{-q} (1-t_0)^{q-1} (1-q-t_0) \end{bmatrix} \cdot \begin{bmatrix} c_1 \\ c_2 \end{bmatrix} \Leftrightarrow \\
\begin{bmatrix} c_1 \\ c_2 \end{bmatrix} &= \frac{1}{1-2q} \begin{bmatrix} t_0^{-q} (1-t_0)^{q-1} (1-q-t_0) & -t_0^{1-q} (1-t_0)^q \\ -t_0^{q-1} (1-t_0)^{-q} (q-t_0) & t_0^q (1-t_0)^{1-q} \end{bmatrix} \begin{bmatrix} x_{10} \\ x_{20} \end{bmatrix}.
\end{aligned}
$$

Therefore, substituting in for $\begin{bmatrix} c_1 \\ c_2 \end{bmatrix}$, in the equation I had leads to

$$\Phi(t,t_0)\,x_0 = \begin{bmatrix} t^q(1-t)^{1-q} & t^{1-q}(1-t)^q \\ t^{q-1}(1-t)^{-q}(q-t) & t^{-q}(1-t)^{q-1}(1-q-t) \end{bmatrix} \cdot \begin{bmatrix} c_1 \\ c_2 \end{bmatrix}$$

$$\begin{aligned} \Phi(t,t_0)\,x_0 &= \begin{bmatrix} t^q(1-t)^{1-q} & t^{1-q}(1-t)^q \\ t^{q-1}(1-t)^{-q}(q-t) & t^{-q}(1-t)^{q-1}(1-q-t) \end{bmatrix} \\ &\quad \frac{1}{1-2q} \begin{bmatrix} t_0^{-q}(1-t_0)^{q-1}(1-q-t_0) & -t_0^{1-q}(1-t_0)^q \\ -t_0^{q-1}(1-t_0)^{-q}(q-t_0) & t_0^q(1-t_0)^{1-q} \end{bmatrix} \begin{bmatrix} x_{10} \\ x_{20} \end{bmatrix} \end{aligned}$$

$$\begin{aligned} \Phi(t,t_0) &= \frac{1}{1-2q} \begin{bmatrix} t^q(1-t)^{1-q} & t^{1-q}(1-t)^q \\ t^{q-1}(1-t)^{-q}(q-t) & t^{-q}(1-t)^{q-1}(1-q-t) \end{bmatrix} \\ &\quad \begin{bmatrix} t_0^{-q}(1-t_0)^{q-1}(1-q-t_0) & -t_0^{1-q}(1-t_0)^q \\ -t_0^{q-1}(1-t_0)^{-q}(q-t_0) & t_0^q(1-t_0)^{1-q} \end{bmatrix}. \end{aligned}$$

In other words, the solution to the homogeneous is given by

$$x(t) = \Phi(t,t_0)\,x_0,$$

where $\Phi(t,t_0)$ is defined above.

From equations (8) and (9), I know that for $t_0 = \underline{p}+$

$$y(t_0) = W$$

$$\dot{y}(t_0) = 0.$$

Changing the notation

$$\begin{aligned} x_0 &= \begin{bmatrix} x_{10} & x_{20} \end{bmatrix}' \\ x_{10} &= y(t_0) = W \\ x_{20} &= \dot{y}(t_0) = 0. \end{aligned}$$

Then the solution to the homogeneous becomes

$$x(t) = \Phi(t, t_0) \begin{bmatrix} W \\ 0 \end{bmatrix}.$$

Using the variation of constants formula, I can now solve the original differential equation

$$x(t) = \Phi(t, t_0) x_0 + \int_{t_0}^{t} \Phi(t, \tau) \begin{bmatrix} 0 \\ 1 \end{bmatrix} Z(\tau) \, d\tau$$

$$\begin{aligned}
\begin{bmatrix} x_1(t) \\ x_2(t) \end{bmatrix} &= \frac{1}{1-2q} \begin{bmatrix} t^q (1-t)^{1-q} & t^{1-q} (1-t)^q \\ t^{q-1} (1-t)^{-q} (q-t) & t^{-q} (1-t)^{q-1} (1-q-t) \end{bmatrix} \\
& \begin{bmatrix} t_0^{-q} (1-t_0)^{q-1} (1-q-t_0) & -t_0^{1-q} (1-t_0)^q \\ -t_0^{q-1} (1-t_0)^{-q} (q-t_0) & t_0^q (1-t_0)^{1-q} \end{bmatrix} \begin{bmatrix} W \\ 0 \end{bmatrix} \\
& + \int_{t_0}^{t} \frac{1}{1-2q} \begin{bmatrix} t^q (1-t)^{1-q} & t^{1-q} (1-t)^q \\ t^{q-1} (1-t)^{-q} (q-t) & t^{-q} (1-t)^{q-1} (1-q-t) \end{bmatrix} \cdot \\
& \begin{bmatrix} \tau^{-q} (1-\tau)^{q-1} (1-q-\tau) & -\tau^{1-q} (1-\tau)^q \\ -\tau^{q-1} (1-\tau)^{-q} (q-\tau) & \tau^q (1-\tau)^{1-q} \end{bmatrix} \begin{bmatrix} 0 \\ 1 \end{bmatrix} Z(\tau) \, d\tau
\end{aligned}$$

$$\begin{bmatrix} x_1(t) \\ x_2(t) \end{bmatrix} = \frac{1}{1-2q} \cdot$$

$$\begin{bmatrix} t^q (1-t)^{1-q} t_0^{-q} (1-t_0)^{q-1} (1-q-t_0) W - t^{1-q} (1-t)^q t_0^{q-1} (1-t_0)^{-q} (q-t_0) W \\ t^{q-1} (1-t)^{-q} (q-t) t_0^{-q} (1-t_0)^{q-1} (1-q-t_0) W - t^{-q} (1-t)^{q-1} (1-q-t) t_0^{q-1} (1-t_0)^{-q} (q-t_0) W \end{bmatrix}$$

$$+ \frac{1}{1-2q} \begin{bmatrix} t^q (1-t)^{1-q} & t^{1-q} (1-t)^q \\ t^{q-1} (1-t)^{-q} (q-t) & t^{-q} (1-t)^{q-1} (1-q-t) \end{bmatrix} \cdot$$

$$\begin{bmatrix} \int_{t_0}^{t} \frac{1}{k_3} \left[k_1 \tau + g(W) \tau^{\beta} (1-\tau)^{1-\beta} + k_2 \right] \left(-\tau^{-1-q} (1-\tau)^{q-2} \right) d\tau \\ \int_{t_0}^{t} \frac{1}{k_3} \left[k_1 \tau + g(W) \tau^{\beta} (1-\tau)^{1-\beta} + k_2 \right] \left(\tau^{q-2} (1-\tau)^{-1-q} \right) d\tau \end{bmatrix}.$$

In other words, the solution to the differential equation is of the following form (with no undermined coefficients)

$$
\begin{aligned}
y(t) &= \frac{1}{1-2q}\left[t^{q}(1-t)^{1-q}t_{0}^{-q}(1-t_{0})^{q-1}(1-q-t_{0})W - t^{1-q}(1-t)^{q}t_{0}^{q-1}(1-t_{0})^{-q}(q-t_{0})W\right] \\
&+\frac{1}{k_{3}(1-2q)}t^{q}(1-t)^{1-q}\left[\int_{t_{0}}^{t}k_{1}\tau+g(W)\tau^{\beta}(1-\tau)^{1-\beta}+k_{2}\right]\left(-\tau^{-1-q}(1-\tau)^{q-2}\right)d\tau \\
&+\frac{1}{k_{3}(1-2q)}t^{1-q}(1-t)^{q}\int_{t_{0}}^{t}\left[k_{1}\tau+g(W)\tau^{\beta}(1-\tau)^{1-\beta}+k_{2}\right]\left(\tau^{q-2}(1-\tau)^{-1-q}\right)d\tau
\end{aligned}
$$

$$
\begin{aligned}
y(t) &= \frac{1}{1-2q}t^{q}(1-t)^{1-q}[t_{0}^{-q}(1-t_{0})^{q-1}(1-q-t_{0})W \\
&+\frac{1}{k_{3}}(-k_{1}\int_{t_{0}}^{t}\tau^{-q}(1-\tau)^{q-2}d\tau \\
&-g(W)\int_{t_{0}}^{t}\tau^{\beta-1-q}(1-\tau)^{-\beta+q-1}d\tau-k_{2}\int_{t_{0}}^{t}\tau^{-1-q}(1-\tau)^{q-2}d\tau)] \\
&+\frac{1}{1-2q}t^{1-q}(1-t)^{q}[t_{0}^{q-1}(1-t_{0})^{-q}(t_{0}-q)W \\
&+\frac{1}{k_{3}}(k_{1}\int_{t_{0}}^{t}\tau^{q-1}(1-\tau)^{-1-q}d\tau \\
&+g(W)\int_{t_{0}}^{t}\tau^{q+\beta-2}(1-\tau)^{-\beta-q}d\tau+k_{2}\int_{t_{0}}^{t}\tau^{q-2}(1-\tau)^{-1-q}d\tau)],
\end{aligned}
$$

where $t_0 \equiv \underline{p}+$ and $q = \frac{1}{2}\left(1-\sqrt{\frac{4+k_3}{k_3}}\right)$.

But using integration by substitution (changing the variable in each of the integrals to $\frac{1}{\tau}$), the six integrals above are equal to

$$\int_{t_0}^{t}\tau^{-q}(1-\tau)^{q-2}d\tau = \frac{1}{q-1}\left(\left(\frac{1}{t_0}-1\right)^{q-1}-\left(\frac{1}{t}-1\right)^{q-1}\right)$$

$$\int_{t_0}^{t}\tau^{\beta-1-q}(1-\tau)^{-\beta+q-1}d\tau = \frac{1}{q-\beta}\left(\left(\frac{1}{t_0}-1\right)^{q-\beta}-\left(\frac{1}{t}-1\right)^{q-\beta}\right)$$

$$\int_{t_0}^{t}\tau^{-1-q}(1-\tau)^{q-2}d\tau = \frac{1}{q(q-1)}\left(\left(\frac{1}{t_0}-1\right)^{q-1}\left(q\frac{1}{t_0}-\frac{1}{t_0}+1\right)-\left(\frac{1}{t}-1\right)^{q-1}\left(q\frac{1}{t}-\frac{1}{t}+1\right)\right)$$

$$\int_{t_0}^{t}\tau^{q-1}(1-\tau)^{-1-q}d\tau = \frac{1}{q}\left(\left(\frac{1}{t}-1\right)^{-q}-\left(\frac{1}{t_0}-1\right)^{-q}\right)$$

$$\int_{t_0}^{t}\tau^{q+\beta-2}(1-\tau)^{-\beta-q}d\tau = \frac{1}{\beta+q-1}\left(\left(\frac{1}{t}-1\right)^{1-q-\beta}-\left(\frac{1}{t_0}-1\right)^{1-q-\beta}\right)$$

$$\int_{t_0}^{t} \tau^{q-2}(1-\tau)^{-1-q}\,d\tau = \frac{1}{q(q-1)}\left(\left(\frac{1}{t}-1\right)^{-q}\left(q\frac{1}{t}-1\right)-\left(\frac{1}{t_0}-1\right)^{-q}\left(q\frac{1}{t_0}-1\right)\right).$$

So I end up with

$$\begin{aligned}
y(t) &= \frac{1}{1-2q}t^q(1-t)^{1-q}t_0^{-q}(1-t_0)^{q-1}(1-q-t_0)W \\
&+\frac{1}{1-2q}\frac{1}{k_3}t^q(1-t)^{1-q}[-k_1\frac{1}{q-1}\left(\left(\frac{1}{t_0}-1\right)^{q-1}-\left(\frac{1}{t}-1\right)^{q-1}\right) \\
&-g(W)\frac{1}{q-\beta}\left(\left(\frac{1}{t_0}-1\right)^{q-\beta}-\left(\frac{1}{t}-1\right)^{q-\beta}\right) \\
&-k_2\frac{1}{q(q-1)}\left(\left(\frac{1}{t_0}-1\right)^{q-1}\left(q\frac{1}{t_0}-\frac{1}{t_0}+1\right)-\left(\frac{1}{t}-1\right)^{q-1}\left(q\frac{1}{t}-\frac{1}{t}+1\right)\right)] \\
&+\frac{1}{1-2q}t^{1-q}(1-t)^q t_0^{q-1}(1-t_0)^{-q}(t_0-q)W \\
&+\frac{1}{1-2q}\frac{1}{k_3}t^{1-q}(1-t)^q[k_1\frac{1}{q}\left(\left(\frac{1}{t}-1\right)^{-q}-\left(\frac{1}{t_0}-1\right)^{-q}\right) \\
&+g(W)\frac{1}{\beta+q-1}\left(\left(\frac{1}{t}-1\right)^{1-q-\beta}-\left(\frac{1}{t_0}-1\right)^{1-q-\beta}\right) \\
&+k_2\frac{1}{q(q-1)}\left(\left(\frac{1}{t}-1\right)^{-q}\left(q\frac{1}{t}-1\right)-\left(\frac{1}{t_0}-1\right)^{-q}\left(q\frac{1}{t_0}-1\right)\right)].
\end{aligned}$$

Finally I still need to pin down the value of t_0. I consider the case where $p \to 1$.

For an experienced worker I know from equation (2) that

$$V^{\exp}(1,W) = \frac{\alpha_G + u + \delta J}{r+\delta}.$$

Similarly for an inexperienced worker, from equation (7) I have

$$(r+\delta+\theta)V^{in}(1,W) = \alpha_G + \delta J + \theta V^{\exp}(1,W).$$

Substituting in for the value of an inexperienced worker, obtains

$$(r+\delta+\theta)V^{in}(1,W) = \alpha_G + \delta J + \theta\frac{\alpha_G+u+\delta J}{r+\delta}$$

$$V^{in}(1,W) = \frac{\alpha_G+\delta J}{r+\delta+\theta} + \frac{\theta}{r+\delta+\theta}\frac{\alpha_G+u+\delta J}{r+\delta}.$$

Put differently, using the notation above I have

$$\lim_{t\to 1} y\left(t\right)=\frac{\alpha_G+\delta J}{r+\delta+\theta}+\frac{\theta}{r+\delta+\theta}\frac{\alpha_G+u+\delta J}{r+\delta}. \tag{10}$$

The above equation pins down $t_0\left(\underline{p}\right)$ for a a given W.

As before the Gittins index of an occupation is the highest retirement value at which the worker is indifferent between working at that occupation or retiring with $W=W\left(p\right)$, i.e.

$$W\left(p\right)=V^{in}\left(p,W\right). \tag{11}$$

For eq. (11) to hold, it must be the case that

$$p=\underline{p}\left(W\right). \tag{12}$$

Substituting the above equation, (12), into the threshold condition, equation (10), obtains

$$\begin{aligned}
W\left(p\right)=&\lim_{x\to 1} \qquad\qquad (13)\\
&\frac{1}{x^{\beta_1}\left(1-x\right)^{1-\beta_1}p^{-\beta_1}\left(1-p\right)^{\beta_1-1}\left(1-\beta_1-p\right)+x^{1-\beta_1}\left(1-x\right)^{\beta_1}p^{\beta_1-1}\left(1-p\right)^{-\beta_1}\left(p-\beta_1\right)}\\
&[\left(1-2\beta_1\right)\left(\frac{\alpha_G+\delta J}{r+\delta+\theta}+\frac{\theta}{r+\delta+\theta}\frac{\alpha_G+u+\delta J}{r+\delta}\right)\\
&-\frac{x^{\beta_1}\left(1-x\right)^{1-\beta_1}}{k_3}[-\frac{k_1}{\beta_1-1}\left(\left(\frac{1}{p}-1\right)^{\beta_1-1}-\left(\frac{1}{x}-1\right)^{\beta_1-1}\right)\\
&-\frac{g\left(W\right)}{\beta_1-\beta_2}\left(\left(\frac{1}{p}-1\right)^{\beta_1-\beta_2}-\left(\frac{1}{x}-1\right)^{\beta_1-\beta_2}\right)\\
&-\frac{k_2}{\beta_1\left(\beta_1-1\right)}\left(\left(\frac{1}{p}-1\right)^{\beta_1-1}\left(\frac{\beta_1-1}{p}+1\right)-\left(\frac{1}{x}-1\right)^{\beta_1-1}\left(\frac{\beta_1-1}{x}+1\right)\right)]\\
&-\frac{x^{1-\beta_1}\left(1-x\right)^{\beta_1}}{k_3}[\frac{k_1}{\beta_1}\left(\left(\frac{1}{x}-1\right)^{-\beta_1}-\left(\frac{1}{p}-1\right)^{-\beta_1}\right)\\
&+\frac{g\left(W\right)}{\beta_1+\beta_2-1}\left(\left(\frac{1}{x}-1\right)^{1-\beta_1-\beta_2}-\left(\frac{1}{p}-1\right)^{1-\beta_1-\beta_2}\right)\\
&+\frac{k_2}{\beta_1\left(\beta_1-1\right)}\left(\left(\frac{1}{x}-1\right)^{-\beta_1}\left(\frac{\beta_1}{x}-1\right)-\left(\frac{1}{p}-1\right)^{-\beta_1}\left(\frac{\beta_1}{p}-1\right)\right)]],
\end{aligned}$$

where

$$g(W) = \frac{-\frac{\theta}{r+\theta+\delta}\frac{\alpha_G-\alpha_B}{r+\delta}}{\frac{(d-1)((r+\delta)W-\alpha_B-u-\delta J)}{(d+1)(\alpha_G-\alpha_B)-2((r+\delta)W-\alpha_B-u-\delta J)}+\frac{1}{2}d-\frac{1}{2}} \times$$
$$\left(\frac{(d-1)((r+\delta)W-\alpha_B-u-\delta J)}{(d+1)(\alpha_G-\alpha_B)-2((r+\delta)W-\alpha_B-u-\delta J)}\right)^{\frac{1}{2}+\frac{1}{2}d} \times$$
$$\left(1-\frac{(d-1)((r+\delta)W-\alpha_B-u-\delta J)}{(d+1)(\alpha_G-\alpha_B)-2((r+\delta)W-\alpha_B-u-\delta J)}\right)^{\frac{1}{2}-\frac{1}{2}d},$$

and

$$k_1 = -\frac{1}{r+\delta}(\alpha_G-\alpha_B)$$

$$k_2 = -\left(\frac{\alpha_B+\delta J}{r+\delta}+\frac{1}{(r+\theta+\delta)}\frac{\theta}{r+\delta}u\right)$$

$$k_3 = \frac{1}{2}\frac{1}{r+\theta+\delta}\left(\frac{\alpha_G-\alpha_B}{\sigma}\right)^2$$

$$q = \frac{1}{2}\left(1-\sqrt{\frac{4+k_3}{k_3}}\right)$$

$$\beta = \frac{1}{2}-\frac{1}{2}\sqrt{\frac{8(r+\delta)}{\left(\frac{\alpha_G-\alpha_B}{\sigma}\right)^2}+1}.$$

The above equation implicitly defines a retirement value $W(p)$ for any value, p, of the posterior, at which the inexperienced worker is exactly indifferent between working at that occupation or retiring. As discussed above, $W(p)$ corresponds to the Gittins index for that occupation.

I also allow workers the option of moving to another city, which provides known value to the worker, J. As in the model of the main text, J is the retirement value associated with the option of moving and therefore corresponds to its Gittins index. A worker therefore moves, if and only if the retirement value (Gittins index) of all other arms is lower than J.

I begin with the case of an experienced worker. In order to find the value of the posterior, $\underline{p}$, where the worker chooses to move, I use equation (6) and substitute J for $W(p)$ to obtain

$$\underline{p} = \frac{(d-1)(rJ-\alpha_B-u)}{(d+1)(\alpha_G-\alpha_B)-2(rJ-\alpha_B-u)}.$$

Similarly, for an inexperienced worker, using (13), the moving threshold, $\underline{p}$ is defined by the following equation

$$J = \lim_{x\to 1} \frac{1}{x^{\beta_1}(1-x)^{1-\beta_1}\underline{p}^{-\beta_1}(1-\underline{p})^{\beta_1-1}(1-\beta_1-\underline{p}) + x^{1-\beta_1}(1-x)^{\beta_1}\underline{p}^{\beta_1-1}(1-\underline{p})^{-\beta_1}(\underline{p}-\beta_1)}$$

$$[(1-2\beta_1)\left(\frac{\alpha_G+\delta J}{r+\delta+\theta}+\frac{\theta}{r+\delta+\theta}\frac{\alpha_G+u+\delta J}{r+\delta}\right)$$

$$-\frac{x^{\beta_1}(1-x)^{1-\beta_1}}{k_3}[-\frac{k_1}{\beta_1-1}\left(\left(\frac{1}{\underline{p}}-1\right)^{\beta_1-1}-\left(\frac{1}{x}-1\right)^{\beta_1-1}\right)$$

$$-\frac{g(J)}{\beta_1-\beta_2}\left(\left(\frac{1}{\underline{p}}-1\right)^{\beta_1-\beta_2}-\left(\frac{1}{x}-1\right)^{\beta_1-\beta_2}\right)$$

$$-\frac{k_2}{\beta_1(\beta_1-1)}\left(\left(\frac{1}{\underline{p}}-1\right)^{\beta_1-1}\left(\frac{\beta_1-1}{\underline{p}}+1\right)-\left(\frac{1}{x}-1\right)^{\beta_1-1}\left(\frac{\beta_1-1}{x}+1\right)\right)]$$

$$-\frac{x^{1-\beta_1}(1-x)^{\beta_1}}{k_3}[\frac{k_1}{\beta_1}\left(\left(\frac{1}{x}-1\right)^{-\beta_1}-\left(\frac{1}{\underline{p}}-1\right)^{-\beta_1}\right)$$

$$+\frac{g(J)}{\beta_1+\beta_2-1}\left(\left(\frac{1}{x}-1\right)^{1-\beta_1-\beta_2}-\left(\frac{1}{\underline{p}}-1\right)^{1-\beta_1-\beta_2}\right)$$

$$+\frac{k_2}{\beta_1(\beta_1-1)}\left(\left(\frac{1}{x}-1\right)^{-\beta_1}\left(\frac{\beta_1}{x}-1\right)-\left(\frac{1}{\underline{p}}-1\right)^{-\beta_1}\left(\frac{\beta_1}{\underline{p}}-1\right)\right)]],$$

where

$$g(J) = \frac{-\frac{\theta}{r+\theta+\delta}\frac{\alpha_G-\alpha_B}{r+\delta}}{\frac{(d-1)(rJ-\alpha_B-u)}{(d+1)(\alpha_G-\alpha_B)-2(rJ-\alpha_B-u)}+\frac{1}{2}d-\frac{1}{2}} \times \left(\frac{(d-1)(rJ-\alpha_B-u)}{(d+1)(\alpha_G-\alpha_B)-2(rJ-\alpha_B-u)}\right)^{\frac{1}{2}+\frac{1}{2}d} \times \left(1-\frac{(d-1)(rJ-\alpha_B-u)}{(d+1)(\alpha_G-\alpha_B)-2(rJ-\alpha_B-u)}\right)^{\frac{1}{2}-\frac{1}{2}d}.$$

2 Additional Results on City Size and Number of Occupations

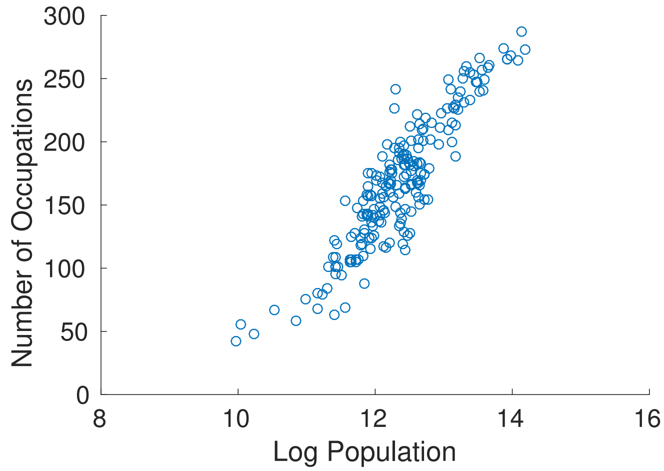

Figure 1: Number of Occupations vs. County Population - JobCentrePlus Vacancy Data, UK

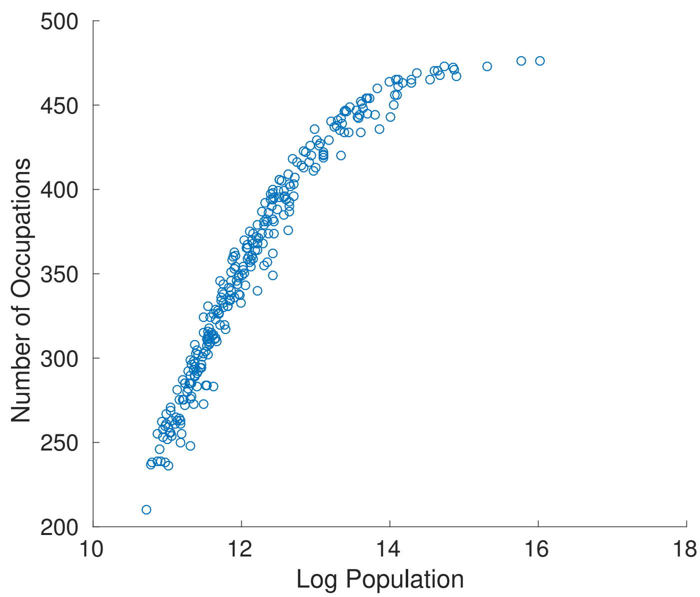

Figure 2: Number of Occupations vs. City Population - 2000 Census Data

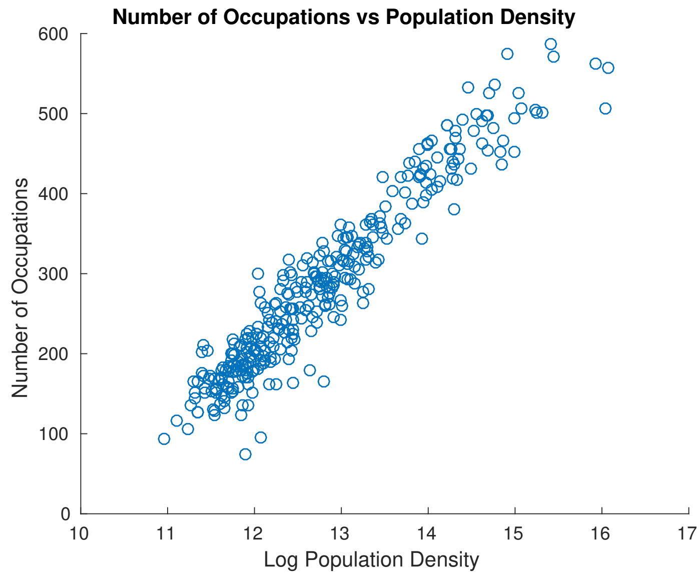

Figure 3: Number of Occupations vs. Population. Source: 2000 Occupational Employment Statistics.

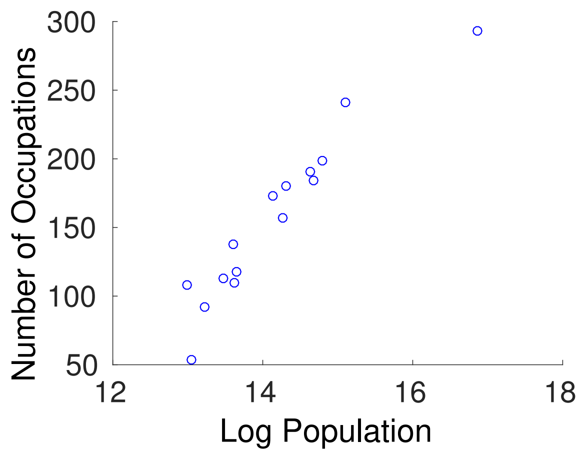

Figure 4: Number of Occupations vs. City Population - Brazilian RAIS data. Males with a College Degree in the State of São Paulo.

3 Endogenous Occupation Creation with Trade across Locations

Here I consider the economy described in Section 6 of the paper, but now allow goods to be tradable across locations. In each location, workers derive utility from the consumption of the final good given by

$$C_t = \left(\sum_{l=1}^{L} \left(\sum_{k=1}^{m^l} \left(c_{kt}^{l} \right)^{\frac{\gamma-1}{\gamma}} \right) \right)^{\frac{\gamma}{\gamma-1}}$$

where $\gamma > 1$ and c^l_{kt} is the consumption at time t of good k produced in location l and m^l is the number of goods produced in location l. In other words, workers can now consume goods produced from *all* locations, rather than just the goods that are locally produced.

The remaining model remains the same as in Section 6, namely increased population in location ν causes a negative externality to workers, captured by $z\left(N^\nu_t\right)$, flow utility is given by

$$C_t - z\left(N_t\right)$$

wage income is given by

$$w_k\left(\alpha_G p^k + \alpha_B\left(1 - p^k\right)\right)$$

and the production function is linear in labor, i.e.

$$q_k = l_k \tag{14}$$

Finally, there is free entry of intermediate good producers in every location. To reduce notational congestion I drop the time subscript in what follows.

Following similar steps to those in the model of Section 6, it is straightforward to show that demand for good j produced in location ν is

$$q^\nu_j = \sum_{l=1}^{L}\left(\frac{b^\nu_j}{P}\right)^{-\gamma}\frac{W^l}{P} + \sum_{l=1}^{L}\left(\frac{b^\nu_j}{P}\right)^{-\gamma} f m^l \tag{15}$$

where

$$P = \left(\sum_{l=1}^{L}\left(\sum_{k=1}^{m^l}\left(b^l_k\right)^{1-\gamma}\right)\right)^{\frac{1}{1-\gamma}}$$

is the *national* price index and

$$W^l = \sum_{k=1}^{m^l} w^l_k l^l_k + \sum_{k=1}^{m^l}\pi^l_k,$$

is total expenditure by city l residents, m^l is the number of occupations in city l, while w^l_k and π^k_k are wages and profits respectively in occupation k of city l.

Each intermediate good producer chooses a price, b^ν_j, given the (national) demand he faces, (15). By choosing the price, the producer also pins down the quantity produced, q^ν_j and from condition (14), this also pins down the equilibrium quantity of labor demanded. I again consider a symmetric equilibrium, where all producers choose the same price, b ($b_k = b$ for all k) and commit to it.

Labor supply, as before, is given by

$$l_k = \theta_k (w_k|w_{-k}) N \int \left(\alpha_G p^k + \alpha_B \left(1 - p^k\right)\right) h_k \left(p^k|w_k, w_{-k}\right) dp^k. \tag{16}$$

It is worth noting that all else equal, labor supply is higher in a city with larger populations, N^ν. Given that labor demand now does not depend on N^ν, this implies that profits, π_j^ν, are increasing in N^ν. To see this note that when goods are tradable across locations, profits of producer j in location ν are given by

$$\pi_j^\nu = b_j^\nu q_j^\nu - w_j^\nu l_j^\nu - Pf. \tag{17}$$

Consider two cities, one small (low N^ν) and one large (high N^ν). Assume both cities have the same number of occupations, m^ν. As before we're focusing on a symmetric equilibrium and let $\bar{b}$ denote the (optimal) price that small city producers choose. To understand why in equilibrium large city producers make higher profits if the number of occupations is the same, note that they can always pick the same price, $\bar{b}$ as the small city producers. Let $\bar{l}$, be the associated amount of effective labor demanded at that price level, as determined by equations (14) and (15). Since all producers sell in the same national goods market, they face the same demand (15) and the first term on the right hand side of (17) is the same. In addition, all producers pay the same amount to cover the fixed cost, Pf as they face the same price index. The large city producers however, all else equal, face a higher labor supply (16), which implies that in order to attract $\bar{l}$ units of effective labor they pay a lower per-unit wage. As a result the second term on the right hand side of (17) is lower and profits are higher. The higher profits combined with the free entry condition imply that more producers enter the high N^ν city, pushing up m^ν and restoring the zero profit condition.

As in the model with no trade across locations in Section 6 of the paper, this immediately implies that cities with larger populations, N^ν, have more occupations, m^ν. Another way of seeing this is to consider are two cities with the same number of occupations, but one has a larger workforce than the other. Then entry of a new occupation is more profitable in the large city, as one can hire the same number of workers at a lower rate.

It is worth noting that the above result is not driven by larger cities having bigger goods markets. Instead, when there is trade across locations, wages can no longer be normalized to 1, but wage differences across locations matter. Now the higher labor supply of larger cities makes them a more attractive destination for producers to locate and from there serve all locations.

In the symmetric equilibrium, within a city ν all producers face the same labor supply and labor demand conditions and therefore pick the same price, $b_k^\nu = b^\nu$ for all k, which implies to the same wage $w_k^\nu = w^\nu$ for all k.[1] In other words, within a city, all occupations offer the same wage per unit of effective labor, so workers' occupational choices depend only on their beliefs as in the model developed in Section 4 of the paper. That model's predictions regarding occupational mobility hold here as well. Workers in cities greater populations, N, have more occupations, m and are better matched in equilibrium.

Finally, following the same steps as in Section 4.2, a worker moves to another city when the posterior of all his occupations reaches:

$$\underline{p}\left(N^\nu, w^\nu\right) = \frac{(d-1)\left(rJ - \frac{w^\nu \alpha_B}{P} + z\left(N^\nu\right)\right)}{(d+1)\frac{w^\nu(\alpha_G - \alpha_B)}{P} - 2\left(rJ - \frac{w^\nu \alpha_B}{P} + z\left(N^\nu\right)\right)}$$

where

$$d = \sqrt{\frac{8\left(r+\delta\right)}{\left(\frac{\alpha_G - \alpha_B}{\sigma}\right)^2} + 1},$$

and J is the value of a worker about to move to another city

$$J = -c + \overline{V},$$

where

$$\overline{V} = \max_l E_{\mathbf{p}} V\left(\mathbf{p}_{m_l}, N^l, w^l\right).$$

and again the worker moves to the city, l, that maximizes his ex ante utility.

The moving probability depends on the level of the negative externality, $z\left(N^\nu\right)$ and also on the city wage rate per effective unit of labor, w^ν. In equilibrium, all workers that move are indifferent across locations. As before, the endogenous moving decision, as well as the inflow decisions of movers pin down city population, N^ν.

[1] Straightforward calculations show that the optimal price is given $b_k = \frac{\gamma}{\gamma - 1 + \frac{dw\left(q_k | w_{-k}\right)}{db_k}} w\left(q_k | w_{-k}\right)$.

4 Full Set of Controls of Mincer Regression

In Table 1, I present the full set of the Mincerian controls for the wage premium regression. Their signs and magnitudes are consistent with the prior literature.

	Full Sample
	ln(wage)
ln(current city pop)	0.041
	$(0.001)^{***}$
Age	-0.033
	$(0.008)^{***}$
Age^2	0.002
	$(0.0003)^{***}$
Age^3	-0.00004
	$(5.40e-06)^{***}$
Age^4	2.18e-07
	$(3.16e-08)^{***}$
Female dummy	-0.15
	$(0.005)^{***}$
High School dummy	0.132
	$(0.005)^{***}$
College dummy	0.248
	$(0.01)^{***}$
Married dummy	0.073
	$(0.005)^{***}$
Non-White dummy	-0.058
	$(0.005)^{***}$
Firm>100 employees	0.131
	$(0.005)^{***}$
Firm 25 - 99 employees	0.03
	$(0.006)^{***}$
constant	1.916
	$(0.127)^{***}$
1-Digit Industry & Occup fixed effects	Yes
Number of Obs	169536

Table 1: Wage Premium. Source: 1996 Panel of Survey of Income and Program Participation. Population data from 2000 Census. Controls include 11 industry dummies and 13 occupation dummies. Standard errors clustered by individual. ***, ** and * indicate statistical significance at 1, 5 and 10 percent respectively.

References

[1] Bergemann, D. and J. Välimäki (2008): "Bandit Problems," in *The New Palgrave Dictionary of Economics,* ed. by S. Durlauf and L. Blume, New York: Macmillan.

[2] Brockett, R. W. (1970): *Finite Dimensional Linear Systems*, New York: Wiley.

[3] Gittins, J. C. and D. M. Jones (1974): "A Dynamic Allocation Index for the Sequential Design of Experiments," in *Progress in Statistics*, vol. 1, ed. by J. M. Gani, K. Sarkadi, and I. Vincze, 241-266, Amsterdam: North-Holland.

[4] Kambourov, G. and I. Manovskii (2009): "Occupational Mobility and Wage Inequality," *Review of Economic Studies*, 76(2), 731-759.

[5] Karatzas, I. (1984): "Gittins Indices in the Dynamic Allocation Problem for Diffusion Processes," *The Annals of Probability*, 12(1), 173-192.

[6] Whittle, P. (1980): *Optimization over Time: Programming and Stochastic Control*, New York: Wiley.

[7] Whittle, P. (1982): "Multi-Armed Bandits and the Gittins Index," *Journal of the Royal Statistical Society*, Series B, 42(2), 143-149.