Online Appendix to

"Immigration and demographics: can high immigrant fertility explain voter support for immigration?"

by Henning Bohn and Armando R. Lopez-Velasco

A Fertility Data for Europe

Sobotka (2008) compiles empirical evidence from several studies on total fertility rates for several European countries. The table below reproduces Sobotka's estimates from his tables 2.a and 2.b and the implied fertility factor (ε) is computed. In every case fertility is higher for immigrants. There are two types of estimates: countries labeled with **(a)** show the comparison between fertility of "native" versus fertility of "immigrant" women, while countries labeled with **(b)** compare "native" nationals with "foreign" nationals. For more details on the particular data sets used for the estimate for each country, see Sobotka (2008).

Table A.1. Total Fertility Rates by Native and Immigrant Status
Several European Countries and Years

Country	Native Fertility	Immigrant Fertility	Ratio (ε)	Year
Austria (b)	1.29	2.03	1.57	2001 – 2005
Belgium (b)	1.49	2.13	1.43	2001 – 2005
Flanders (Belgium)(b)	1.5	3	2	1995
Denmark (a)	1.69	2.43	1.44	1999 – 2003
England and Wales (a)	1.6	2.2	1.38	2001
France (a)	1.65	2.50	1.52	1991 – 1998
France (ii) (a)	1.70	2.16	1.27	1991 – 1998
France (b)	1.72	2.80	1.63	1999
France (b)	1.80	3.29	1.83	2004
Italy (b)	1.26	2.61	2.07	2004
Netherlands (a)	1.65	1.97	1.19	2005
Norway (a)	1.76	2.42	1.38	1997 – 1998
Spain (b)	1.19	2.12	1.78	2002
Sweden (a)	1.72	2.01	1.17	2005
Switzerland (b)	1.34	1.86	1.39	1997

Source: Sobotka (2008) tables 2.a and 2.b.

(a): Native vs. immigrant women. (b) : Native nationals vs. foreign nationals.

For (ii), data is adjusted for age of arrival and duration of stay in France.

B Analysis in Sections 2-3

B.1 Utility as Function of Factor Prices

We claim in Section 2 that (4) implies the indirect utility (6).

Proof: For log-utility ($\gamma = 1$), maximizing

$$U\left(c_t^1, c_{t+1}^2\right) = \ln c_t^1 + \beta \ln c_{t+1}^2 \tag{A.1}$$

straightforwardly yields optimal consumption and savings

$$c_t^1 = \frac{1}{1+\beta} w_t,\; c_{t+1}^2 = \frac{\beta}{1+\beta} w_t R_{t+1}, \text{ and } s_t = \frac{\beta}{1+\beta} w_t.$$

Inserting these expressions into (A.1), using rules of logarithms and collecting similar terms one obtains

$$\begin{aligned} U\left(w_t, R_{t+1}\right) &= A_1 + (1+\beta)\ln w_t + \beta \ln R_{t+1} \\ A_1 &= \left[\beta \ln \beta - (1+\beta)\ln(1+\beta)\right] \end{aligned} \tag{A.2}$$

The constant $A_1 = [\beta \ln \beta - (1+\beta)\ln(1+\beta)]$ is inessential and omitted from (6) for simplicity.

For power utility ($\gamma \neq 1$), the first order conditions of (4) imply

$$s_t = \frac{(\beta R_{t+1})^{1/\gamma}}{R_{t+1} + (\beta R_{t+1})^{1/\gamma}} w_t,\; c_{1,t} = \frac{R_{t+1}}{R_{t+1} + (\beta R_{t+1})^{1/\gamma}} w_t, \text{ and } c_{2,t+1} = (\beta R_{t+1})^{\frac{1}{\gamma}} c_{1,t}.$$

Substituting consumption into the utility function, one obtains:

$$U\left(w_t, R_{r+1}\right) = \frac{1}{1-\gamma} \frac{\left\{1 + \beta^{\frac{1}{\gamma}} R_{t+1}^{\frac{1}{\gamma}-1}\right\}}{\left(1 + \beta^{\frac{1}{\gamma}} R_{t+1}{}^{\frac{1}{\gamma}-1}\right)^{1-\gamma}} w_t^{1-\gamma} = \frac{1}{1-\gamma}\left\{1 + \beta^{\frac{1}{\gamma}} R_{t+1}^{\frac{1}{\gamma}-1}\right\}^{\gamma} w_t^{1-\gamma},$$

as claimed. *QED.*

B.2 Extensions with stochastic mortality

We claim in footnote 8 that our maintained assumption of a working-age majority is a robust result in extended versions with stochastic mortality. This appendix provides an illustration.

To model stochastic mortality within a two-period setting, suppose there is idiosyncratic uncertainty about survival at the end of working age: individuals die with probability $(1-\pi)$ and survive with probability π. These assumption largely follow Bohn (2001). The consumption/saving problem of individuals is then

$$\begin{aligned} U_t &= \underset{c_t^1, \hat{c}_{t+1}^2, s_t}{Max} \left\{u(c_t^1) + E[\hat{\beta} u(\hat{c}_{t+1}^2)]\right\} \\ &= \underset{c_t^1, c_{t+1}^2, s_t}{Max} \left\{u(c_t^1) + \pi \hat{\beta} u(c_{t+1}^2)\right\} \\ s.t. \quad & c_t^1 + s_t = w_t \text{ and } c_{t+1}^2 = \hat{R}_{t+1} s_t + Q_{t+1}, \end{aligned} \tag{A.3}$$

where $E[\cdot]$ is the expectation over survival, $\hat{\beta}$ pure time preference, $\hat{c}^2_{t+1} = c^2_{t+1}$ in case of survival, $\hat{c}^2_{t+1} = 0$ otherwise, and Q_{t+1} denotes bequests (if any). The return $\hat{R}_{t+1}$ conditional on survival depends on the availability of annuity markets. With actuarially fair annuities, $\hat{R}_{t+1} = R_{t+1}/\pi$, as all savings are allocated to survivors. Without annuities, $\hat{R}_{t+1} = R_{t+1}$ is the return on savings, and one must make assumptions about the disposal of deceased agents' assets.

In general, denote survivors return to saving by $\hat{R}_{t+1} = R_{t+1}/\pi^a$ where $a \in [0,1]$ admits intermediate degrees of annuitization. Then the first order condition for optimal saving is $u'(c^1_t) = \pi\hat{\beta}\hat{R}_{t+1}u'(c^2_{t+1}) = \pi^{1-a}\hat{\beta}R_{t+1}u'(c^2_{t+1})$. Accidental bequests are $(R_{t+1}/\pi - \hat{R}_{t+1})s_t$. Without much loss of generality, assume accidental bequests are shared by surviving members of the old generation. Then $c^2_{t+1} = \hat{R}_{t+1}s_t + Q_{t+1} = R_{t+1}s_t/\pi$ applies regardless of annuitization, and optimal saving imply $(w_t - s_t)^{-\gamma} = \pi^{1-a}\hat{\beta}R_{t+1}(R_{t+1}s_t/\pi)^{-\gamma}$, and hence $\sigma_t = \frac{s_t}{w_t} = \frac{\hat{B}_t}{1+\hat{B}_t}$, where $\hat{B}_t = (\pi^{1-a+\gamma}\hat{\beta})^{\frac{1}{\gamma}}(R_{t+1})^{\frac{1-\gamma}{\gamma}}$. Thus, individual consumption and saving are the same as in the model with 100% survival and time preference $\beta = \pi^{1-a+\gamma}\hat{\beta}$.

Since the voting population consists of $N_t = N_{t-1}\eta(1+\varepsilon\theta_{t-1})$ young and πN_{t-1} old agents, the voting share of the young is given by $\eta(1+\varepsilon\theta_{t-1})/[\pi + \eta(1+\varepsilon\theta_{t-1})]$. Since $\theta_{t-1} \geq 0$, the voting share of the young is bounded below by $\eta/(\pi+\eta)$. Empirically, U.S. life expectancy at age 65 is about 20 years (male and female averaged according to the Social Security Administration[23]), so π can be estimated as (remaining life expectancy)/(number of years in workforce), which yields $\pi \approx 1/2$. Hence $\eta > 1/2$ is a sufficient condition for the young to be the majority in absence of immigration. In section 5.1 we calibrate $\eta = 1.116$, which rules out immigration cycles by a wide margin.[24]

B.3 Utility as Function of Immigration

Section 3 claims that utility in terms of immigration quotas is given by (14) and that dynamically efficiency corresponds to $\chi \geq 1 - \alpha$.

Proof of (14): Inserting (9) into (8), one obtains

$$k_{t+1} = \frac{\beta(1-\alpha)}{\eta(1+\beta)} \frac{(1+\theta_t)}{(1+\varepsilon\theta_t)(1+\theta_{t+1})} k_t^{\alpha} \tag{A.4}$$

Substituting this into $R_{t+1} = \alpha(k_{t+1})^{-1+\alpha}$ and substitution R_{t+1} into the utility function (A.2), we obtain

[23] According to the Social Security Administration, conditional on reaching 65 years of age, a man in the US is expected to live until age 84.3, while a woman is expected to live until age 86.6. See https://www.ssa.gov/planners/lifeexpectancy.html .

[24] The particular assumptions required for immigration cycles are: (i) Fertility and mortality such that the old generation is the majority in the absence of immigration: $N_{t+1} = \eta N_t < \pi N_t$, which implies that $\eta < \pi$; and (ii) a policy space with maximum immigration quota $\theta^{\max}$ high enough that in the next period there are more young agents than (alive) old agents: $N_{t+1} = N_t\eta(1+\varepsilon\theta^{\max}) > \pi N_t$ (implies $\eta(1+\varepsilon\theta^{\max}) > \pi$). Under these assumptions, the initial old majority would choose the maximum quota available ($\theta^*_t = \theta^{\max}$). Then since the majority in next period is the young cohort because $\eta(1+\varepsilon\theta^{\max}) > \pi$, this young majority restricts immigration in order to remain the majority in the following period (when they are old), which is a period in which they liberalize immigration. That particular cohort controls policy when young and when old. The immigration quota that the young majority selects is either $\theta^*_{t+1} = 0$ or a sligthly positive number (depending on how large ε is), taking into account that the they will be in power the next period and thus choosing the maximum quota in that period ($\theta^*_{t+2} = \theta^{\max}$). A cycle of restriction and then liberalization repeats, tracking the life cycle of cohorts that remain in power during all their lifetimes and similarly, each cycle sees a generation that is never in power, and so on.

$$\widehat{V}(\theta_t, \theta_{t+1}, k_t) = A_1 + (1+\beta)\ln(1-\alpha)\,k_t^{\alpha}$$
$$- \beta(1-\alpha)\ln\left(\frac{\beta(1-\alpha)}{(1+\beta)}\frac{(1+\theta_t)}{\eta(1+\varepsilon\theta_t)(1+\theta_{t+1})}k_t^{\alpha}\right)$$

Finally, substituting $k_t = \frac{\kappa_t}{1+\theta_t}$, collecting similar terms, and dividing the equation by the constant $\beta(1-\alpha)$, we obtain the indirect utility function that depends on θ_t, θ_{t+1} and κ_t :

$$V(\theta_t, \theta_{t+1}, \kappa_t) = A + \chi \ln \kappa_t - (1+\chi)\ln(1+\theta_t) + \ln(1+\varepsilon\theta_t) + \ln(1+\theta_{t+1}). \qquad \text{(A.5)}$$

where χ is a constant given by $\chi = \frac{\alpha(1+\beta\alpha)}{\beta(1-\alpha)}$. We write $V(\theta_t, \theta_{t+1}, \kappa_t)$, where we condition on κ_t since agents take it as given. Q.E.D.

Regarding dynamic efficiency, note that the steady state ratio of return to capital is greater than (or equal to) population growth if and only if $\chi > 1-\alpha$ (or $\chi = 1-\alpha$). In detail, (A.4) implies that k_{t+1} converges to $k = \left(\frac{\beta}{1+\beta}\frac{1-\alpha}{\eta(1+\varepsilon\theta)}\right)^{1/(1-\alpha)}$ for any constant $\theta_t = \theta$, and hence R_{t+1} converges to $R = \alpha / \left(\frac{\beta}{1+\beta}\frac{1-\alpha}{\eta(1+\varepsilon\theta)}\right)$. Since $\eta(1+\varepsilon\theta)$ is population growth, the return-to-growth ratio is $\frac{R}{\eta(1+\varepsilon\theta)} = \frac{\alpha}{1-\alpha}\frac{1+\beta}{\beta}$. Moreover,

$$\chi + \alpha = \frac{\alpha(1+\alpha\beta)}{\beta(1-\alpha)} + \alpha = \frac{\alpha}{\beta(1-\alpha)}[1+\alpha\beta+\beta(1-\alpha)] = \frac{\alpha}{1-\alpha}\frac{1+\beta}{\beta}.$$

Hence $\frac{R}{\eta(1+\varepsilon\theta)} = 1$ iff $\chi = 1-\alpha$ and $\frac{R}{\eta(1+\varepsilon\theta)} > 1$ iff $\chi > 1-\alpha$.

B.4 Welfare Comparisons

In the baseline model, we claim that a transition from zero immigration to $\theta^0 = \frac{\varepsilon-(1+\chi)}{\chi\varepsilon} > 0$ will (a) increase welfare for one or more generations and (b) reduce welfare in the long run, provided the economy is dynamically efficient. The following provides a constructive proof:

First we write the evolution of capital per-native worker by substituting $k_t = \frac{\kappa_t}{(1+\theta_t)}$ into equation (A.4). We obtain

$$\kappa_{t+1} = \varpi \frac{(1+\theta_t)^{1-\alpha}}{(1+\varepsilon\theta_t)}\kappa_t^{\alpha}, \text{ where } \varpi = \frac{\beta(1-\alpha)}{\eta(1+\beta)}. \qquad \text{(A.6)}$$

Since we will compare the lifetime utility of agents with and without immigration, it is convenient to define $V_t^0 = V(\theta^0, \theta^0, \kappa_t)$ as lifetime utility in the regime with immigration and to define $\tilde{V}_t = V(0, 0, \kappa_t)$ as lifetime utility in a regime without immigration. Using (A.5), lifetimes utility in the immigration regime is given by

$$V_t^0 = V(\theta^0, \theta^0, \kappa_t) = A + \chi \ln \kappa_t + \Omega \qquad \text{(A.7)}$$

(since $\theta_t = \theta^0$ is constant), where $\Omega = \ln(1+\varepsilon\theta^0) - \chi\ln(1+\theta^0)$. Note that $\varepsilon > (1+\chi)$ implies $\Omega > 0$.

Taking logs in equation (A.6), we can write the evolution of κ_t in the regime with immigration as

$$\ln \kappa_{t+1} = \ln \varpi - \Delta + \alpha \ln \kappa_t$$
$$\text{where } \Delta = \ln\left(1+\varepsilon\theta^0\right) - (1-\alpha)\ln\left(1+\theta^0\right)$$

Note that $\Delta > \Omega$ for $\chi > (1-\alpha)$ and $\theta^0 > 0$, which applies under conditions of dynamic efficiency.

To compare utilities across regimes, we write lifetime utility in terms of the initial value κ_0 in some starting period labeled $t = 0$. For all $t > 0$, we have:

$$\ln \kappa_t = \frac{\left(1-\alpha^t\right)}{1-\alpha}\ln \varpi - \frac{\left(1-\alpha^t\right)}{1-\alpha}\Delta + \alpha^t \ln \kappa_0.$$

Using (A.7), the sequence of lifetime utilities is

$$V_t^0 = A + \chi\frac{\left(1-\alpha^t\right)}{1-\alpha}\ln \varpi + \alpha^t \ln \kappa_0 + \left[\Omega - \chi\frac{\left(1-\alpha^t\right)}{1-\alpha}\Delta\right].$$

In a regime without immigration, analogous dynamics apply with $\Omega = \Delta = 0$, so the sequence of lifetime utilities would be

$$\tilde{V}_t = A + \chi\frac{\left(1-\alpha^t\right)}{1-\alpha}\ln \varpi + \alpha^t \ln \kappa_0$$

Hence the difference between lifetime utilities with and without immigration is

$$V_t^0 - \tilde{V}_t = \left[\Omega - \chi\frac{\left(1-\alpha^t\right)}{1-\alpha}\Delta\right]$$

Notice that $V_0^0 - \tilde{V}_0 = \Omega > 0$ is positive for the generation $t = 0$ and that because $\Delta > 0$, $V_t^0 - \tilde{V}_t$ declines monotonely over time. As $t \to \infty$, $(V_t^0 - \tilde{V}_t) \to \Omega - \frac{\chi\Delta}{1-\alpha}$. Note that

$$\begin{aligned}
\Omega - \frac{\chi\Delta}{1-\alpha} &= -(1+\chi)\ln\left(1+\theta^0\right) + \ln\left(1+\varepsilon\theta^0\right) + \ln\left(1+\theta^0\right) \\
&\quad - \frac{\chi}{1-\alpha}\left\{\ln\left(1+\varepsilon\theta^*\right) - (1-\alpha)\ln\left(1+\theta^0\right)\right\} \\
&= \left(1 - \frac{\chi}{1-\alpha}\right)\ln\left(1+\varepsilon\theta^0\right)
\end{aligned}$$

Hence $\lim_{t\to\infty}(V_t^0 - V_t) < 0$ and only if $\chi > (1-\alpha)$. Monotonicity then implies there is a date $\bar{t}$ such that $V_t^0 - V_t > 0$ for all $t < \bar{t}$, whereas $V_t^0 - V_t < 0$ for all $t > \bar{t}$. Q.E.D.

C Analysis in Section 4

Section 4.1 asserts that indirect utility is (18). Proof: Let $j = I, II$ denote the states, $p_I = 1-p$, $p_{II} = p$. Then the individual problem for general CRRA utility is

$$U_t = \underset{c_t^1, c_{t+1,i}^2, s_t}{Max}\left\{u(c_t^1) + \beta\sum_j p_j \cdot u(c_{t+1,j}^2)\right\} \tag{A.8}$$

$$s.t. \quad c_t^1 + s_t = w_t,\ c_{t+1,j}^2 = R_{t+1,j}s_t$$

The first order condition for optimal savings is $u'(w_t - s_t) = \beta \sum_j p_j \cdot R_{t+1,j} u'(R_{t+1,j} s_t)$, and it implicit defines the optimal savings rate

$$\sigma_t = \frac{s_t}{w_t} = \frac{B_t}{1+B_t}, \text{ where} \tag{A.9}$$

$$B_t = \beta^{\frac{1}{\gamma}} \left[\sum_j p_j \cdot (R_{t+1,j})^{1-\gamma} \right]^{\frac{1}{\gamma}} \tag{A.10}$$

For log-utility ($\gamma = 1$), this reduces to $B_t = \beta$. Substituting consumption and savings (A.9) into (A.8), one obtains

$$U_t = (1+\beta) \ln w_t + \beta E \left[\ln R_{t+1} \right] + const$$

Substituting wages (10) and returns (10), this implies

$$\begin{aligned} \widehat{V} &= -\left(\alpha \left(1+\beta\alpha\right) + \beta \left(1-\alpha\right)\right) \ln \left(1+\theta_t\right) + \beta \left(1-\alpha\right) \ln \left(1+\varepsilon\theta_t\right) \\ &\quad + \beta \left(1-\alpha\right) E \left[\ln \left(1+\widehat{\theta}_{t+1}\right) \right] + exog \end{aligned}$$

where immigration $\widehat{\theta}_{t+1}$ is treated as random variable and *exog* summarizes exogenous terms (inessential constants and initial conditions). The possible realizations for $\widehat{\theta}_{t+1}$ are $\widehat{\theta}_{t+1} = \theta^I_{t+1}$ with probability $1-p$ (chosen by gen. $t+1$) and $\widehat{\theta}_{t+1} = \theta^{II}_{t+1} = \theta_t$ with probability p. Therefore $\widehat{V}$ can be written as

$$\begin{aligned} \widehat{V} &= -\left(\alpha \left(1+\beta\alpha\right) + \beta \left(1-\alpha\right)\right) \ln \left(1+\theta_t\right) + \beta \left(1-\alpha\right) \ln \left(1+\varepsilon\theta_t\right) \\ &\quad + \beta \left(1-\alpha\right) \left(1-p\right) \ln \left(1+\theta_{t+1}\right) + \beta \left(1-\alpha\right) p \ln \left(1+\theta_t\right) + exog \end{aligned}$$

Dividing by $\beta \left(1-\alpha\right)$ and simplifying, one obtains

$$V\left(\theta_t, \theta^I_{t+1}\right) = -\left(1+\chi-p\right) \ln \left(1+\theta_t\right) + \ln \left(1+\varepsilon\theta_t\right) + \left(1-p\right) \ln \left(1+\theta^I_{t+1}\right) + exog$$

If state I applies in period t, $\theta_t = \theta^I_t$ is set by generation t and $V\left(\theta^I_t, \theta^I_{t+1}\right)$ is the relevant indirect utility. *QED.*

We also claim that $\theta^p = \frac{\varepsilon - 1 - \chi + p}{\varepsilon(\chi - p)}$is increasing in p and in ε. As proof, note that

$$\begin{aligned} \frac{d\theta^p}{dp} &= \frac{\varepsilon\left(\chi-p\right) + \left(\varepsilon-1-\chi+p\right)\varepsilon}{\varepsilon^2 \left(\chi-p\right)^2} = \frac{\left(\varepsilon-1-\chi+p\right)\varepsilon}{\varepsilon\left[\left(\chi+p\right)-1\right]^2} = \frac{\varepsilon-1}{\varepsilon\left(\chi-p\right)^2} > 0 \\ \frac{d\theta^p}{d\varepsilon} &= \frac{\varepsilon\left(\chi-p\right) - \left(\varepsilon-1-\chi+p\right)\left(\chi-p\right)}{\varepsilon^2 \left(\chi-p\right)^2} = \frac{\chi+1-p}{\varepsilon^2 \left(\chi-p\right)} > 0. \end{aligned}$$

for all (ε, χ, p) such that $0 < \theta^p < \bar{\theta}$ is not a corner solution. *QED.*

D Details on Calibrating the CRRA model in Section 5

The parameters to be calibrated are $\{\alpha, \beta, \varepsilon, \gamma, \eta\}$. We explain each of them in this section.

Externally calibrated parameters. The externally calibrated parameters are α, ε and γ. We use baseline values $\alpha = \frac{1}{3}$ and $\varepsilon = 1.5$ as discussed in Section 3.2. Regarding γ, typical overlapping generation economies in the macroeconomics and the finance literature use risk aversion levels between 2 and 5 (see for example Auerbach and Kotlikoff (1987), Rios-Rull (1996), Constantinides et al. (2002), Conesa and Garriga (2008) and Evans et al. (2012)). We use a baseline of $\gamma = 4$, but also explore the sensitivity of the fertility required to different levels of risk aversion.

Internally calibrated parameters. The internally calibrated parameters are η and β. The total-growth factor of the population in the model is given by $\eta^+ \equiv \eta(1+\varepsilon\theta)$, which we set at 1.25 (about 1% growth per year), which is consistent with US population growth.[25] Given η^+, one can infer η from ε and θ by writing $\eta = \eta^+/(1+\varepsilon\theta)$. Using $\varepsilon = 1.5$ and $\theta = .08$, this yields $\eta = \frac{1.25}{1+1.5\cdot 0.08} = 1.116$.

For the calibration of β, we show that one can derive a calibrating expression that is invariant to the particular equilibrium concept used. Start with equation (8), which shows the evolution of capital per native worker in the CRRA case, given by

$$\kappa_{t+1} = \frac{(1+\theta_t)^{1-\alpha}}{\eta(1+\varepsilon\theta_t)}\sigma_t w_t.$$

replacing $w_t = (1-\alpha)\kappa_t^\alpha$, the evolution of capital per native worker can be written as

$$\kappa_{t+1} = \frac{(1-\alpha)(1+\theta_t)^{1-\alpha}}{\eta(1+\varepsilon\theta_t)}\sigma_t \kappa_t^\alpha$$

where σ_t is the saving rate. At steady state, this equation can be solved for $\frac{1}{\sigma}$ as

$$\frac{1}{\sigma} = \frac{(1-\alpha)}{\eta(1+\varepsilon\theta)}\left(\frac{\kappa}{1+\theta}\right)^{\alpha-1}$$

Since the gross interest rate at steady state R is given by $R = \alpha\left(\frac{\kappa}{1+\theta}\right)^{\alpha-1}$. The above expression can be written as

$$\frac{1}{\sigma} = \frac{(1-\alpha)}{\alpha}\frac{R}{\eta(1+\varepsilon\theta)}. \tag{A.11}$$

From the definition of σ_t in equation (5), at steady state $\sigma = \frac{\beta^{\frac{1}{\gamma}}R^{\frac{1}{\gamma}-1}}{1+\beta^{\frac{1}{\gamma}}R^{\frac{1}{\gamma}-1}}$, which implies that the term $\frac{1}{\sigma}$ is also given by

$$\frac{1}{\sigma} = 1 + \frac{1}{\beta^{\frac{1}{\gamma}}R^{\frac{1}{\gamma}-1}}. \tag{A.12}$$

Therefore, at steady state equating (A.11) and (A.12) yields an equality $1 + \frac{1}{\beta^{\frac{1}{\gamma}}R^{\frac{1}{\gamma}-1}} = \frac{(1-\alpha)}{\alpha}\frac{R}{\eta(1+\varepsilon\theta)}$ that can be solved for β as

$$\beta = \frac{R^{\gamma-1}}{\left[\frac{(1-\alpha)}{\alpha}\frac{R}{\eta(1+\varepsilon\theta)} - 1\right]^\gamma}.$$

[25] Using US census data from 1970 to 2010, the annual population growth rate for this 40 year period is $\left(\frac{308.45}{205.52}\right)^{1/40} - 1 = 1.02\%$.

Thus the calibration of β depends on the ratio $r \equiv \frac{R}{\eta(1+\varepsilon\theta)}$ as explained in the text. Since $\eta^{+} = \eta(1+\varepsilon\theta)$ is also empirically observed, R can be replaced by $r \cdot \eta(1+\varepsilon\theta)$ and hence β can be calibrated as

$$\widehat{\beta} = \frac{\left[(r \cdot \eta(1+\varepsilon\theta))\right]^{\gamma-1}}{\left[\frac{(1-\alpha)}{\alpha} r - 1\right]^{\gamma}}. \tag{A.13}$$

In the case of log utility ($\gamma = 1$), this simplifies to

$$\widehat{\beta}_{(\gamma=1)} = \frac{\alpha}{(1-\alpha)\, r - \alpha}.$$

as claimed in Section 3.2. For general γ and in the context of sensitivity analyses, we vary β as implied by the parameters on the r.h.s. of $\widehat{\beta}$.

E Analysis in Section 5.1

E.1 Derivation of $\frac{dR_{t+1}}{d\theta_t}$

Section 5.1 claims that $\frac{dR_{t+1}}{d\theta_t} > 0$. Proof: Combining (8), (A.9), (A.10), (9), and writing (10) as $k_{t+1} = \left(\frac{\alpha}{R_{t+1}}\right)^{\frac{1}{1-\alpha}}$, one obtains

$$\left(\frac{\alpha}{R_{t+1}}\right)^{\frac{1}{1-\alpha}} = k_{t+1} = \frac{(1-\alpha)(1+\theta_t)^{1-\alpha}(\kappa_t)^{1-\alpha}}{\eta(1+\varepsilon\theta_t)(1+\theta_{t+1})} \frac{\beta^{\frac{1}{\gamma}}(R_{t+1})^{\frac{1}{\gamma}-1}}{1+\beta^{\frac{1}{\gamma}}(R_{t+1})^{\frac{1}{\gamma}-1}}.$$

Taking logs and differentiating with respect to θ_t, one obtains

$$\begin{aligned}
-\frac{1}{1-\alpha}\frac{d\ln R_{t+1}}{d\theta_t} &= \frac{1-\alpha}{1+\theta_t} - \frac{\varepsilon}{1+\varepsilon\theta_t} + \left(\frac{1}{\gamma}-1\right)\frac{1}{1+B_t}\frac{d\ln R_{t+1}}{d\theta_t} \\
\frac{d\ln R_{t+1}}{d\theta_t} &= (1-\alpha)\left[\frac{\varepsilon}{(1+\varepsilon\theta_t)} - \frac{(1-\alpha)}{(1+\theta_t)}\right] - \phi\frac{1}{1+B_t}\frac{d\ln R_{t+1}}{d\theta_t} \\
\frac{d\ln R_{t+1}}{d\theta_t} &= \frac{(1-\alpha)}{1+\phi/(1+B_t)}\left[\frac{\varepsilon}{(1+\varepsilon\theta_t)} - \frac{(1-\alpha)}{(1+\theta_t)}\right],
\end{aligned}$$

$$\frac{dR_{t+1}}{d\theta_t} = \frac{(1-\alpha)\, R_{t+1}}{1+\phi/(1+B_t)}\left[\frac{\varepsilon}{(1+\varepsilon\theta_t)} - \frac{(1-\alpha)}{(1+\theta_t)}\right]. \tag{A.14}$$

The sign is determined by $\left[\frac{\varepsilon}{(1+\varepsilon\theta_t)} - \frac{(1-\alpha)}{(1+\theta_t)}\right]$, which is positive iff $\varepsilon > \frac{(1-\alpha)}{(1+\alpha\theta_t)}$. Since $\varepsilon > 1$ and $\frac{1-\alpha}{1+\alpha\theta_t} \leq 1-\alpha$, $\frac{dR_{t+1}}{d\theta_t} > 0$ follows. *QED.*

E.2 The First Order Condition

Section 5.1 asserts the optimality condition (22). Proof: Maximizing (6) with respect to θ_t implies (using (A.10))

$$\begin{aligned}
\frac{dU_t}{d\theta_t} &= \{1+B_t\}^{\gamma} w_t^{-\gamma}\frac{dw_t}{d\theta_t} + \{1+B_t\}^{\gamma-1} w_t^{1-\gamma}\beta^{\frac{1}{\gamma}} R_{t+1}^{\frac{1}{\gamma}-2}\frac{dR_{t+1}}{d\theta_t} \\
&= \{1+B_t\}^{\gamma} w_t^{1-\gamma}\left[\frac{1}{w_t}\frac{dw_t}{d\theta_t} + \frac{B_t}{1+B_t}\frac{1}{R_{t+1}}\frac{dR_{t+1}}{d\theta_t}\right]
\end{aligned} \tag{A.15}$$

Hence $\frac{dU_t}{d\theta_t} = 0 \Leftrightarrow \frac{B_t}{1+B_t}\frac{1}{R_{t+1}}\frac{dR_{t+1}}{d\theta_t} = -\frac{1}{w_t}\frac{dw_t}{d\theta_t}$. Note that $\frac{1}{w_t}\frac{dw_t}{dk_t} = \frac{\alpha}{k_t}$ and $\frac{dk_t}{d\theta_t} = -\frac{k_t}{(1+\theta_t)}$, so $\frac{1}{w_t}\frac{dw_t}{d\theta_t} = -\frac{\alpha}{(1+\theta_t)} < 0$. Replacing $\frac{dR_{t+1}}{d\theta_t}$ by (A.14), one finds that $\frac{dU_t}{d\theta_t} = 0$ is equivalent to

$$\frac{\alpha}{(1+\theta_t)} = (1-\alpha)\left[\frac{\varepsilon}{(1+\varepsilon\theta_t)} - \frac{(1-\alpha)}{(1+\theta_t)}\right] / \left(1 + \left(\frac{1}{\gamma} + \alpha - \frac{\alpha}{\gamma}\right)\frac{1}{B_t}\right),$$

$$1 + \left(\frac{1}{\gamma} + \alpha - \frac{\alpha}{\gamma}\right)\frac{1}{B_t} = \left(\frac{1-\alpha}{\alpha}\right)\left[\frac{\varepsilon(1+\theta_t)}{(1+\varepsilon\theta_t)} - (1-\alpha)\right] = \left(\frac{1-\alpha}{\alpha}\right)\left[\alpha - \frac{1-\varepsilon}{(1+\varepsilon\theta_t)}\right],$$

$$\left(\frac{1}{\gamma} + \alpha - \frac{\alpha}{\gamma}\right)\frac{1}{B_t} = \frac{1-\phi}{B_t} = \left[\left(\frac{1-\alpha}{\alpha}\right)\frac{\varepsilon - 1}{(1+\varepsilon\theta_t)}\right] - \alpha.$$

Dividing by $(1-\phi)$, using (A.10) and (10) to replace B_t, one obtains (22). *QED.*

Note that the corner solution $\theta_t = 0$ applies if $\frac{dU_t}{d\theta_t} \leq 0$ at $\theta_t = 0$, which is equivalent to

$$\frac{\alpha}{(1+\theta_t)} - (1-\alpha)\left[\frac{\varepsilon}{(1+\varepsilon\theta_t)} - \frac{(1-\alpha)}{(1+\theta_t)}\right] / \left(1 + (1-\phi)\,\beta^{-\frac{1}{\gamma}} R_{t+1}^{1-\frac{1}{\gamma}}\right) \geq 0, \text{ or}$$

$$\frac{1}{B_t} \geq \frac{1}{1-\phi}\left[\left(\frac{1-\alpha}{\alpha}\right)(\varepsilon - 1) - \alpha\right].$$

E.3 The Dynamics of Capital per Worker

Proof of (23): For CRRA utility, the evolution of capital is given by

$$k_{t+1} = \frac{(1+\theta_t)}{\eta\,(1+\varepsilon\theta_t)\,(1+\theta_{t+1})} s_t = \frac{(1+\theta_t)}{\eta\,(1+\varepsilon\theta_t)\,(1+\theta_{t+1})}\frac{B_t}{1+B_t} w_t \tag{A.16}$$

If optimal immigration has an interior solution, (22) implies

$$1 + \frac{1}{B_t} = 1 + \frac{\left(\frac{1-\alpha}{\alpha}\right)\frac{\varepsilon-1}{(1+\varepsilon\theta_t)} - \alpha}{1-\phi} = \frac{\frac{1}{\gamma}(1-\alpha) + \left(\frac{1-\alpha}{\alpha}\right)\frac{\varepsilon-1}{(1+\varepsilon\theta_t)}}{1-\phi}$$

Substituting into (A.16) and simplifying implies

$$k_{t+1} = \frac{(1+\theta_t)}{\eta\,(1+\theta_{t+1})}\frac{\alpha(1-\alpha+\alpha\gamma)}{\varepsilon\alpha(1+\theta_t) + (\gamma-\alpha)(\varepsilon-1)} k_t^{\alpha}, \tag{A.17}$$

$$\kappa_{t+1} = \frac{1}{\eta}\frac{\alpha(1-\alpha+\alpha\gamma)}{\varepsilon\alpha(1+\theta_t) + (\gamma-\alpha)(\varepsilon-1)}\,(1+\theta_t)^{1-\alpha}\,\kappa_t^{\alpha}, \tag{A.18}$$

which is (23). *QED* .

Note that for corner solutions with $\theta_t = 0$, (A.16) and $R_{t+1} = \alpha k_{t+1}^{\alpha-1}$ imply

$$\kappa_{t+1} + \beta^{-\frac{1}{\gamma}}\alpha^{1-\frac{1}{\gamma}}(\kappa_{t+1})^{\phi}(1+\theta_{t+1})^{(1-\alpha)(1-\frac{1}{\gamma})} = \frac{1}{\eta}(1-\alpha)\kappa_t^{\alpha}.$$

Since $\phi > 0$, the l.h.s. is strictly increasing in κ_{t+1}, so κ_{t+1} is determined uniquely.

E.4 Steady State

Section 5.1 claims there is a unique steady state (θ^0, κ^0). Proof: The steady state conditions are obtained by setting (θ, κ) constant in (22) and (23), which yields:

$$\kappa^0 = \frac{1}{\eta}\frac{\alpha(1-\alpha+\alpha\gamma)}{\varepsilon\alpha\theta^0 + (\gamma-\alpha)(\varepsilon-1)}\left(1+\theta^0\right)^{1-\alpha}(\kappa^0)^\alpha, \tag{A.19}$$

$$\beta^{-\frac{1}{\gamma}}(R^0)^{1-\frac{1}{\gamma}} = \beta^{-\frac{1}{\gamma}}\alpha^{1-\frac{1}{\gamma}}(\frac{\kappa^0}{1+\theta^0})^{-\phi} = \frac{\left(\frac{1-\alpha}{\alpha}\right)\frac{\varepsilon-1}{1+\varepsilon\theta^0} - \alpha}{1-\phi} \tag{A.20}$$

where (A.19) can be simplified to obtain

$$\frac{\kappa^0}{1+\theta^0} = \left\{\frac{\alpha\left[(1-\alpha)+\alpha\gamma\right]}{\eta\left(\varepsilon\alpha(1+\theta^0)+(\gamma-\alpha)(\varepsilon-1)\right)}\right\}^{\frac{1}{1-\alpha}} \tag{A.21}$$

Substituting (A.21) into (A.20), one obtains

$$\beta^{-\frac{1}{\gamma}}\alpha^{1-\frac{1}{\gamma}}\left(\frac{\eta\left(\varepsilon\alpha(1+\theta^0)+(\gamma-\alpha)(\varepsilon-1)\right)}{\alpha\left[(1-\alpha)+\alpha\gamma\right]}\right)^{(1-\frac{1}{\gamma})} = \frac{\left(\frac{1-\alpha}{\alpha}\right)\frac{\varepsilon-1}{(1+\varepsilon\theta^0)} - \alpha}{1-\phi},$$

$$\beta^{-\frac{1}{\gamma}}\left(\frac{\eta\left(\varepsilon\alpha(1+\theta^0)+(\gamma-\alpha)(\varepsilon-1)\right)}{\left[(1-\alpha)+\alpha\gamma\right]}\right)^{(1-\frac{1}{\gamma})} = \frac{\left(\frac{1-\alpha}{\alpha}\right)\frac{\varepsilon-1}{(1+\varepsilon\theta^0)} - \alpha}{1-\phi}$$

which is a univariate equation for θ^0; the solution is unique since the l.h.s. is strictly increasing in θ^0 for $\gamma > 1$ whereas the r.h.s. is strictly decreasing. Given θ^0, (A.21) provides solutions for κ^0 and $k^0 = \kappa^0/(1+\theta^1)$. *QED.*

E.5 Convergence and Stability

Section 5.1 claims that the perfect foresight path $\{\theta_t, \kappa_t\}_{t\geq t_0}$ converges to (θ^0, κ^0). To streamline the algebra, we sometimes work with $x_t = 1+\theta_t$ and $k_t = \frac{\kappa_t}{x_t}$ (since we can always recover $\{\theta_t, \kappa_t\}_{t\geq t_0}$ from $\{x_t, k_t\}_{t\geq t_0}$). To streamline, we use the constants

$$\phi_1 = \frac{\alpha\left[(1-\alpha)+\alpha\gamma\right]}{\eta} > 0$$

$$\phi_2 = (\gamma-\alpha)(\varepsilon-1)$$

$$\phi_3 = \frac{\eta^\phi\left[\alpha^{1+\alpha-\frac{\alpha}{\gamma}}\right]\left[(1-\alpha)+\alpha\gamma\right]^{\alpha+\frac{1-\alpha}{\gamma}}}{\beta^{\frac{1}{\gamma}}\gamma(1-\alpha)(\varepsilon-1)} > 0$$

$$\phi_4 = \frac{\alpha^2}{(1-\alpha)(\varepsilon-1)} > 0$$

and we omit superscripts for variables in steady state.

The dynamic system (A.16) and (22) can be written in terms of $\{x_t, k_t\}$ as

$$k_{t+1} = \phi_1\frac{x_t}{x_{t+1}\left[\varepsilon\alpha x_t + \phi_2\right]}k_t^\alpha$$

$$\frac{1}{\varepsilon x_t - (\varepsilon-1)} = \phi_3\left[\frac{x_{t+1}\left[\varepsilon\alpha x_t + \phi_2\right]}{x_t k_t^\alpha}\right]^\phi + \phi_4,$$

and the steady-state values are

$$k^1 = \left\{ \frac{\phi_1}{[\varepsilon\alpha x + \phi_2]} \right\}^{\frac{1}{1-\alpha}}$$
$$\frac{1}{\varepsilon x^1 - (\varepsilon - 1)} = \phi_3 \phi_1^{-\frac{\alpha\phi}{1-\alpha}} \left[\varepsilon\alpha x^1 + \phi_2\right]^{\frac{\phi}{1-\alpha}} + \phi_4.$$

To determine the stability of the system, we take a log-linear approximation around the steady state. Denote the percentage deviations from steady state by "^", e.g. $\widehat{z}_t = \ln z_t - \ln z^1$ for generic variable z_t. We obtain

$$\widehat{\kappa}_{t+1} = \alpha\widehat{k}_t + b_0\widehat{x}_t = \alpha\widehat{\kappa}_t + (b_0 - \alpha)\widehat{x}_t$$
$$\widehat{x}_{t+1} = \alpha\widehat{k}_t - (1/b_1)\widehat{x}_t = \alpha\widehat{\kappa}_t - (\alpha + 1/b_1)\widehat{x}_t,$$

where

$$b_0 = \frac{\phi_2}{\varepsilon\alpha x^1 + \phi_2} = \frac{1}{1 + \frac{\varepsilon\alpha x^1}{(\gamma-\alpha)(\varepsilon-1)}} \tag{A.22}$$

$$b_1 = \frac{\left[\varepsilon x^1 - (\varepsilon - 1)\right]}{\left(\frac{1}{\phi}\left\{\frac{\varepsilon x^1}{1-\varepsilon x^1\phi_4+(\varepsilon-1)\phi_4}\right\} - \left[\frac{[\varepsilon x^1-(\varepsilon-1)](\gamma-\alpha)(\varepsilon-1)}{(\varepsilon\alpha x^1+(\gamma-\alpha)(\varepsilon-1))}\right]\right)} \tag{A.23}$$

In matrix form, this is

$$\begin{bmatrix}\widehat{\kappa}_{t+1} \\ \widehat{x}_{t+1}\end{bmatrix} = \begin{bmatrix}\alpha & (b_0 - \alpha) \\ \alpha & -\left(\alpha + \frac{1}{b_1}\right)\end{bmatrix}\begin{bmatrix}\widehat{\kappa}_t \\ \widehat{x}_t\end{bmatrix}$$

Stability requires that the system has one characteristic inside the unit circle and the other outside. Here the characteristic equation is $\mu^2 + \frac{\mu}{b_1} - \frac{\alpha(1+b_0b_1)}{b_1} = 0$, which has roots

$$\mu_{1,2} = -\frac{1}{2b_1} \pm \sqrt{\frac{1}{4}\left(\frac{1}{b_1}\right)^2 + \frac{\alpha\,(1 + b_0b_1)}{b_1}}.$$

The properties of $\mu_{1,2}$ require tedious derivations, which we report in a series of Lemmas below; in combination, the Lemmas provide conditions for $\mu_1 < -1$ and $0 < \mu_2 < 1$, which are sufficient conditions for saddle-path stability and convergence.

Lemma A1: $\gamma > \alpha$ implies $0 < b_0 < 1$. Proof: Follows from $\phi_2 = (\gamma - \alpha)(\varepsilon - 1) > 0$ and $x^1 \geq 1$. *QED.*

Lemma A2: $1 - \phi_4\left[\varepsilon x^1 - (\varepsilon - 1)\right] > 0$, provided $x^1 > 1$. Proof: The steady state satisfies

$$\frac{1}{\varepsilon x - (\varepsilon - 1)} = \phi_3\phi_1^{-\frac{\alpha\phi}{1-\alpha}}\left[\varepsilon\alpha x + \phi_2\right]^{\frac{\phi}{1-\alpha}} + \phi_4$$

$$\Leftrightarrow 1 - \phi_4\left[\varepsilon x - (\varepsilon - 1)\right] = \phi_3\phi_1^{-\frac{\alpha\phi}{1-\alpha}}\left[\varepsilon x - (\varepsilon - 1)\right]\left[\varepsilon\alpha x + \phi_2\right]^{\frac{\phi}{1-\alpha}}$$

The r.h.s. is positive because $\varepsilon\alpha x + \phi_2 > 0$, $\varepsilon x - (\varepsilon - 1) > 0$ and $\phi_3\phi_1^{-\frac{\alpha\phi}{1-\alpha}} > 0$. Hence $1 - \phi_4\left[\varepsilon x - (\varepsilon - 1)\right] > 0$. *QED.*

Lemma A3: $b_1 > 0$, provided $x^1 > 1$. Proof: b_1 can be written as

$$b_1 = 1/\left(\left\{\frac{\varepsilon x}{\left[1 - \phi_4\left[\varepsilon x - (\varepsilon - 1)\right]\right]\phi\left[\varepsilon x - (\varepsilon - 1)\right]}\right\} - \left[\frac{\phi_2}{(\varepsilon\alpha x + \phi_2)}\right]\right)$$

Therefore $b_1 > 0$ if $\left\{\frac{\varepsilon x}{[1-\phi_4[\varepsilon x-(\varepsilon-1)]]\phi[\varepsilon x-(\varepsilon-1)]}\right\} > \left[\frac{\phi_2}{(\varepsilon\alpha x+\phi_2)}\right] = b_0$. Using Lemma A2, this is equivalent to

$$\varepsilon x\left(\varepsilon\alpha x+\phi_2\right) > \phi_2\phi\left[1-\phi_4\left[\varepsilon x-(\varepsilon-1)\right]\right]\left[\varepsilon x-(\varepsilon-1)\right], \text{ or}$$

$$\alpha\left(\varepsilon x\right)^2+\varepsilon x\phi_2+\phi_2\phi\phi_4\left[\varepsilon x-(\varepsilon-1)\right]^2 > \phi_2\phi\left[\varepsilon x-(\varepsilon-1)\right].$$

Since $\phi_2 = (\gamma-\alpha)(\varepsilon-1) > 0$ and $0 < \phi = (1-\alpha)\left(1-\frac{1}{\gamma}\right) < 1$ if $\gamma > 1$, one can divide both sides by $\phi_2\phi$ and simplify to obtain

$$\frac{\alpha\left(\varepsilon x\right)^2}{\phi_2\phi}+\varepsilon x\left(\frac{1-\phi}{\phi}\right)+\phi_4\left[\varepsilon x-(\varepsilon-1)\right]^2+(\varepsilon-1) > 0.$$

The r.h.s. is positive, because $0 < \phi < 1$ and because ϕ_2, ϕ_4 and ε are all greater than one. By equivalence, $b_1 > 0$. *QED.*

Lemma A4: $0 < \mu_2 < 1$, provided $0 < b_0 < 1$ and $b_1 > 0$. Proof: Since $b_0 > 0$ and $b_1 > 0$, we have $\frac{\alpha(1+b_0b_1)}{b_1} > 0$, which implies $\mu_2 = -\frac{1}{2b_1}+\sqrt{\frac{1}{4}\left(\frac{1}{b_1}\right)^2+\frac{\alpha(1+b_0b_1)}{b_1}} > 0$. Given $\mu_2 > 0$ and $b_1 > 0$, $\mu_2+\frac{1}{2b_1} > 0$, so $\mu_2 = -\frac{1}{2b_1}+\sqrt{\frac{1}{4}\left(\frac{1}{b_1}\right)^2+\frac{\alpha(1+b_0b_1)}{b_1}} < 1 \Leftrightarrow \frac{1}{4}\left(\frac{1}{b_1}\right)^2+\frac{\alpha(1+b_0b_1)}{b_1} < \left(1+\frac{1}{2b_1}\right)^2$, $\Leftrightarrow 1-\alpha b_0+\frac{1-\alpha}{b_1} > 0$, which is implied by $b_0 < 1$ and $b_1 > 0$. *QED.*

Lemma A5: $\mu_1 < -1$, provided $0 < b_0 < 1$ and $0 < b_1 < 1$. Proof: Since $b_0 > 0$ and $b_1 > 0$ imply $\frac{\alpha(1+b_0b_1)}{b_1} > 0$, $\mu_2 < -\frac{1}{2b_1}-\sqrt{\frac{1}{4}\left(\frac{1}{b_1}\right)^2} < -\frac{1}{b_1}$, so $\mu_2 < -\frac{1}{b_1} < -1$ for $b_1 < 1$. *QED.*

Lemma A6: Define $\rho = \frac{\alpha}{\gamma-\alpha}$ and $z(\rho) = (2+3\rho)-2\sqrt{2\rho(1+\rho)} = \left(\sqrt{2(1+\rho)}-\sqrt{\rho}\right)^2$. Then sufficient conditions for $b_1 < 1$ are that

$$(1-\frac{1}{\gamma})z(\rho)\left(1-\alpha-\frac{\alpha^2}{\varepsilon-1}\right) < 1, \text{ or} \tag{A.24}$$

$$(1-\frac{1}{\gamma})z(\rho) < \frac{1}{1-\alpha} \tag{A.25}$$

Proof: Since (A.25) implies (A.24) holds for $\varepsilon > 1$, it suffices to prove (A.24). Since $b_1 > 0$, the restriction $b_1 < 1 \Leftrightarrow \left\{\frac{\varepsilon x}{[1-\phi_4[\varepsilon x-(\varepsilon-1)]]\phi[\varepsilon x-(\varepsilon-1)]}\right\}-\frac{\phi_2}{\varepsilon\alpha x+\phi_2} > 1 \Leftrightarrow \left\{\frac{\varepsilon x}{[1-\phi_4[\varepsilon x-(\varepsilon-1)]]\phi[\varepsilon x-(\varepsilon-1)]}\right\} > \left(\frac{\varepsilon\alpha x+2\phi_2}{\varepsilon\alpha x+\phi_2}\right) \Leftrightarrow \left\{\frac{\left(\frac{\varepsilon\alpha x}{(\gamma-\alpha)(\varepsilon-1)}+1\right)}{\left(\frac{\varepsilon\alpha x}{(\gamma-\alpha)(\varepsilon-1)}+2\right)}\frac{\varepsilon x}{[\varepsilon x-(\varepsilon-1)]}\frac{1}{\phi}\right\}+\phi_4\left[\varepsilon x-(\varepsilon-1)\right] > 1$

Define $\Psi = \frac{(\varepsilon-1)}{\varepsilon x}$, $\rho = \frac{\alpha}{(\gamma-\alpha)}$, and

$$H\left(\Psi\right) = \frac{\left(\frac{\rho}{\Psi}+1\right)}{\left(\frac{\rho}{\Psi}+2\right)}\frac{1}{1-\Psi} = \frac{(\rho+\Psi)}{(\rho+2\Psi)}\frac{1}{1-\Psi}.$$

Then

$$b_1 < 1 \Leftrightarrow \frac{1}{\phi}H\left(\Psi\right)+\phi_4\left[\varepsilon x^1-(\varepsilon-1)\right] > 1.$$

Given $\varepsilon > 1$ and given an interior solution $x^1 > 1$, we have $0 < \Psi < 1$. The terms $\frac{1}{[1-\Psi]}$ and $\frac{1}{\phi}$ are both greater than 1, while the term $\frac{\left[\frac{\rho}{\Psi}+1\right]}{\left[\frac{\rho}{\Psi}+2\right]}$ is necessarily less than 1. Note that $H(\Psi)$ has a minimum in the interval $0 < \Psi < 1$ and that

$$\frac{1}{\phi}H(\Psi) + \phi_4 [\varepsilon x - (\varepsilon - 1)] \geq \left[\min_{0<\Psi<1} H(\Psi)\right] \frac{1}{\phi} + \phi_4 \left\{\min_{x\geq 1} [\varepsilon x - (\varepsilon - 1)]\right\}$$

Since $\min_{x\geq 1} [\varepsilon x - (\varepsilon - 1)] = 1$, a sufficient condition for $b_1 < 1$ is that

$$\left[\min_{0<\Psi<1} H(\Psi)\right] \frac{1}{\phi} + \phi_4 > 1 \Leftrightarrow \min_{0<\Psi<1} H(\Psi) > \phi(1-\phi_4) \tag{A.26}$$

where $\phi(1-\phi_4) > 0$ because $(1-\phi_4) > 0$ for $x > 1$ and $\phi > 0$ for $\gamma > 0$. Note that $H'(\Psi) = \frac{(\rho+2\Psi)-2(\rho+\Psi)}{(\rho+2\Psi)^2}\frac{1}{(1-\Psi)} + \frac{(\rho+\Psi)}{(\rho+2\Psi)}\frac{1}{(1-\Psi)^2} = \frac{-\rho(1-\Psi)+(\rho+\Psi)(\rho+2\Psi)}{(\rho+2\Psi)^2(1-\Psi)^2} = \frac{2\Psi^2+4\rho\Psi-\rho(1-\rho)}{(\rho+2\Psi)^2(1-\Psi)^2} = \frac{\Psi^2+2\rho\Psi-\rho(1-\rho)/2}{2(\rho+2\Psi)^2(1-\Psi)^2} = 0$ has roots $\Psi = -\rho \pm \sqrt{\rho^2+\rho(1-\rho)/2} = -\rho \pm \sqrt{\frac{\rho}{2}(1+\rho)}$. Of these, only $\Psi^{\min} = -\rho + \sqrt{\frac{\rho}{2}(1+\rho)}$ is inside the $[0,1]$ interval, so H is minimized at $\Psi^{\min}$, where

$$\begin{aligned}
H^{\min} = H\left(\Psi^{\min}\right) &= \frac{\sqrt{\frac{\rho}{2}(1+\rho)}}{\left(-\rho+2\sqrt{\frac{\rho}{2}(1+\rho)}\right)}\frac{1}{\left(1+\rho-\sqrt{\frac{\rho}{2}(1+\rho)}\right)} \\
&= \frac{\sqrt{\frac{\rho}{2}(1+\rho)}}{-\rho(1+\rho)+2(1+\rho)\sqrt{\frac{\rho}{2}(1+\rho)}+\rho\sqrt{\frac{\rho}{2}(1+\rho)}-2\frac{\rho}{2}(1+\rho)} \\
&= \frac{\sqrt{\frac{\rho}{2}(1+\rho)}}{-2\rho(1+\rho)+(2+3\rho)\sqrt{\frac{\rho}{2}(1+\rho)}} = \frac{1}{2+3\rho-2\rho(1+\rho)/\sqrt{\frac{\rho}{2}(1+\rho)}} \\
&= \frac{1}{2+3\rho-2\sqrt{2\rho(1+\rho)}} = \frac{1}{z(\rho)}
\end{aligned}$$

Thus from (A.26), $1/z(\rho) > \phi(1-\phi_4) = \left(1-\frac{1}{\gamma}\right)\left(1-\alpha-\frac{\alpha^2}{\varepsilon-1}\right)$ is sufficient for $b_1 < 1$. *QED*.

Lemma A7: (a) The term $\left(1-\frac{1}{\gamma}\right)z(\rho)$ in (A.24) is increasing in γ and decreasing in α provided $\alpha < \gamma/2$; (b) the $1-\alpha-\frac{\alpha^2}{\varepsilon-1}$in (A.24) increasing in ε and decreasing in α. (c) If $1-\alpha-\frac{\alpha^2}{\varepsilon-1} \leq \frac{1}{2}$ and $\alpha \leq \gamma/2$, then condition (A.24) is satisfied for all $\gamma > 1$. (d) if $1-\alpha-\frac{\alpha^2}{\varepsilon-1} > \frac{1}{2}$ and $\alpha \leq 1/2$ there exists an upper bound $\bar{\gamma} > 1$ so that condition (A.24) is satisfied for all $\gamma \in (1,\bar{\gamma})$ and not satisfied for $\gamma > \bar{\gamma}$.

Proof: Note that $z'(\rho) = 3 - 2^{1.5}\frac{1+2\rho}{2\sqrt{\rho(1+\rho)}} = 0$ has roots $\rho_{1,2} = -\frac{1}{2} \pm \sqrt{\frac{1}{4}+2} = -2, +1$ and satisfies $z'(\rho) < 0$ for $\rho \in (-1,1)$. Since $\alpha < \gamma/2$ implies $\rho < 1$ and since $\frac{\partial\rho}{\partial\gamma} < 0 < \frac{\partial\rho}{\partial\alpha}$ $\frac{\partial}{\partial\gamma}\left[\left(1-\frac{1}{\gamma}\right)z(\frac{\alpha}{\gamma-\alpha})\right] = \frac{1}{\gamma^2}z(\rho) + \left(1-\frac{1}{\gamma}\right)\frac{\partial\rho}{\partial\gamma}z'(\rho) > 0$ for all $\alpha \leq 1/2 < \gamma/2$, and $\frac{\partial\rho}{\partial\alpha} = \left(1-\frac{1}{\gamma}\right)\frac{\partial\rho}{\partial\alpha}z'(\rho) < 0$, proving (a). Part (b) holds by inspection. For (c), note that $\left(1-\frac{1}{\gamma}\right)z(\frac{\alpha}{\gamma-\alpha}) \to z(0) = 2$ as $\gamma \to \infty$, so $\left(1-\frac{1}{\gamma}\right)z(\frac{\alpha}{\gamma-\alpha}) < 2$ for any finite γ, which implies (c). For (d), note that $\left(1-\frac{1}{\gamma}\right)z(\frac{\alpha}{\gamma-\alpha}) \to 0$ as $\gamma \to 1$, so (A.24) holds in a neighborhood of $\gamma = 1$. For $1-\alpha-\frac{\alpha^2}{\varepsilon-1} > \frac{1}{2}$ and $\left(1-\frac{1}{\gamma}\right) z(\frac{\alpha}{\gamma-\alpha}) \to 2$ implies that (A.24) cannot hold as $\gamma- > \infty$. Existence of $\bar{\gamma}$ then follow from the mean value theorem and uniqueness of $\bar{\gamma}$ from monotonicity of $\left(1-\frac{1}{\gamma}\right) z(\frac{\alpha}{\gamma-\alpha})$. *QED.*

Corollary to A1-A7: If (A.25) holds for some $\alpha = \bar{\alpha}$ and $\gamma = \bar{\gamma}$, then (A.25) holds for all $\bar{\alpha} \leq \alpha \leq 1/2$ and all $1 < \gamma \leq \bar{\gamma}$. If in addition $x^1 > 1$, then $\mu_1 < -1$ and $0 < \mu_2 < 1$, so saddle-path stability holds.

Condition (A.25) can be evaluated numerically and is satisfied for plausible parameters. For example, for $\bar{\alpha} = 1/3$ and $\bar{\gamma} = 10$, one finds $\left(1-\frac{1}{\gamma}\right) z(\rho) = 1.4133 < \frac{\alpha}{1-\alpha} = 1.5$ and for $\bar{\alpha} = 0.2$ and $\bar{\gamma} = 5$, one finds $\left(1-\frac{1}{\gamma}\right) z(\rho) = 1.2285 < \frac{\alpha}{1-\alpha} = 1.25$. Thus the system is saddle-path stable for all $1/3 \leq \alpha \leq 0.5$ and $1 < \gamma \leq 10$ and for all $0.2 \leq \alpha \leq 0.5$ and $1 < \gamma \leq 5$. The sufficient condition tends to fail only if α is implausibly small and γ is large. (Note that the condition is not necessary; in some cases, one can use (A.24) to show that the system is stable even though (A.25) fails. For example, for $\alpha = 1/3$ and $\gamma = 15$, one can show that (A.24) applies for all $\varepsilon \leq 10$.)

F Analysis in Section 5.2

F.1 Derivation of $\frac{dR_{t+1}}{d\theta_t}$ under Markov Strategies

Section 5.2 asserts (25), assuming no persistence. Proof: By the Markovian assumption on strategies we can write $\theta_{t+1} = g(\kappa_{t+1})$, where g is some unknown function. Differentiating $\ln R_{t+1}$, one obtains

$$\begin{aligned}\frac{d\ln R_{t+1}}{d\theta_t} &= (1-\alpha)\left[\frac{d\ln(1+\theta_{t+1})}{d\ln\kappa_{t+1}} - 1\right]\frac{d\ln\kappa_{t+1}}{d\theta_t} \\ &= (1-\alpha)\left[\lambda_{t+1} - 1\right]\frac{d\ln\kappa_{t+1}}{d\theta_t}.\end{aligned}$$

where $\lambda_{t+1} = \frac{d\ln(1+\theta_{t+1})}{d\ln\kappa_{t+1}} = \frac{\kappa_{t+1}g'(\kappa_{t+1})}{1+g(\kappa_{t+1})}$. Differentiating

$$\begin{aligned}\ln\kappa_{t+1} &= \left[\ln(1-\alpha) - \ln\eta\right] + (1-\alpha)\ln(1+\theta_t) \\ &\quad - \ln(1+\varepsilon\theta_t) - \ln\left(1+\beta^{-\frac{1}{\gamma}}R_{t+1}^{1-\frac{1}{\gamma}}\right) + \alpha\ln\kappa_t\end{aligned} \tag{A.27}$$

with respect to θ_t, one obtains

$$\frac{d\ln\kappa_{t+1}}{d\theta_t} = \frac{1-\alpha}{(1+\theta_t)} - \frac{\varepsilon}{(1+\varepsilon\theta_t)} - \frac{\left(1-\frac{1}{\gamma}\right)}{1+B_t}\frac{d\ln R_{t+1}}{d\theta_t}$$

Substituting the above equation into $\frac{dR_{t+1}}{d\theta_t}$ above, we obtain

$$\frac{d\ln R_{t+1}}{d\theta_t} = (1-\alpha)\left[\lambda_{t+1}-1\right]\left\{\frac{1-\alpha}{(1+\theta_t)} - \frac{\varepsilon}{(1+\varepsilon\theta_t)} - \frac{\left(1-\frac{1}{\gamma}\right)}{1+B_t}\frac{d\ln R_{t+1}}{d\theta_t}\right\}.$$

Solving for $\frac{dR_{t+1}}{d\theta_t}$ and simplifying terms we obtain

$$\frac{d\ln R_{t+1}}{d\theta_t} = \frac{(1-\alpha)\left[\lambda_{t+1}-1\right]\left[\frac{1-\alpha}{(1+\theta_t)} - \frac{\varepsilon}{(1+\varepsilon\theta_t)}\right](1+1/B_t)}{(1+1/B_t) + \left[\lambda_{t+1}-1\right](1-\alpha)\left(1-\frac{1}{\gamma}\right)/B_t}$$
$$= \frac{(1-\alpha)(1-\lambda_{t+1})\left[\frac{\varepsilon}{(1+\varepsilon\theta_t)} - \frac{1-\alpha}{(1+\theta_t)}\right](1+1/B_t)}{1+(1-(1-\lambda_{t+1})\phi)/B_t} \tag{A.28}$$

As in the previous section, maximizing (6) with respect to θ_t again implies (A.15), where $\frac{dR_{t+1}}{d\theta_t} = R_{t+1}\frac{d\ln R_{t+1}}{d\theta_t}$ is now given by (A.28). An interior solution again requires $\frac{d\ln R_{t+1}}{d\theta_t}/(1+1/B_t) = \frac{\alpha}{(1+\theta_t)}$, which (after simplifying as in the previous section) reduces to (25). *QED.*

F.2 The Numerical Algorithm for the Equilibrium in Markov Strategies

We obtain a log-linearization of the model in Markov strategies and verify the quality of the solution by using numerical methods that use projection methods similar to den Haan and Marcet's (1990) parameterized expectations approach (PEA).

Specifically, we approximate period-t expectations about θ^*_{t+1} by using the function:

$$\left(1+\theta^*_{t+1}\right) = \exp\left(\sum_{i=0}^{n}\mu_i\left(\ln\left(\kappa_{t+1}\right)\right)^i\right) \tag{A.29}$$

for some unknown coefficients $\boldsymbol{\mu} = \{\mu_0, \mu_1, \mu_2, ..\mu_n\}$, and where κ_{t+1} is known at time t. The general idea is to choose some particular coefficients $\widehat{\boldsymbol{\mu}}$ such that the distance between the forecast and the optimal choices dictated by the equations of the model are minimized. For our application, the coefficients are obtained after an iterative procedure where the model is solved for different points on the state grid[26], where the solution is conditional on a previous set of coefficients. The new set of equilibrium pairs $\{\kappa_t, \theta_t\}$ are used to generate a new set of coefficients that describe the optimal policy function (and the elasticity $\frac{d\ln\left(1+\theta^*_{t+1}\right)}{d\ln\kappa_{t+1}}$), and which also update the expectation function of the model. This is repeated until the distance of the forecast and actual choices consistent with that forecast is minimized.

Under equation (A.29) the term $\frac{d\ln\left(1+\theta^*_{t+1}\right)}{d\ln\kappa_{t+1}}$ needed in the first order condition of generation $[t]$ is given by

$$\frac{d\ln\left(1+\theta_{t+1}\right)}{d\ln\kappa_{t+1}} = \sum_{i=1}^{n} i\mu_i\left(\ln\left(\kappa_{t+1}\right)\right)^{i-1} \tag{A.30}$$

The particular steps of the algorithm are explained below.

Preliminaries. For the initial conditions that represent the grid of the state variable (κ) we use a neighborhood around the myopic steady state, given by $\left[.75\kappa^1, 1.5\kappa^1\right]$. This interval does not need to be symmetric, and in our particular case it is due to the fact that

[26] See Christiano and Fisher (2000) for the use of this step in the PEA algorithm. Since simulating the model will naturally lead to points which have a high probability and not many points of states with low probability, they suggest several variations of the algorithm that amount to collocation in the grid, rather than increasing the number of periods of the simulation. Some of their best variations of the PEA algorithm don't require many collocation points (i.e. 5 collocation points) , nor require many terms in their expectation function.

(with $\gamma > 1$) the Markov steady state level (κ^*) is always to the right of the myopic steady state (κ^1).

Step 1. Start with the solution of the rational expectations "myopic" equilibrium where immigration quotas θ_t take as given the future immigration quotas θ_{t+s}, for $s \geq 1$. The model is simulated for many periods $t = 0, 1, 2, ...T$ using perfect foresight under $j = 1, 2, ...J$ different initial conditions in the state grid. For each of the time series generated under each initial condition, a point (κ_0, θ_0^*) is obtained.[27] Hence there are J pairs $\left(\kappa_{0,j}, \theta_{0,j}^*\right)$ that show the optimal immigration quota consistent with the initial condition and with perfect foresight simulation of the model. Then those pairs are used to estimate equation $(A.29)$ in order to obtain the first set of coefficients $\left(\widehat{\boldsymbol{\mu}}_1\right)$. More details about the specific regression are explained in step (4). Then form the forecasting function $\widehat{g}\left(\kappa_{t+1}; \widehat{\boldsymbol{\mu}}_1\right)$, which is parameterized by the initial set of coefficients $\widehat{\boldsymbol{\mu}}_1$. For economy of notation in what follows define the forecasting function parameterized by the coefficients of the s^{th} iteration by $\widehat{g}\left(\kappa_{t+1}; \widehat{\boldsymbol{\mu}}_s\right) \equiv \widehat{g}_s\left(\kappa_{t+1}\right)$.

Step 2. For each initial condition $\kappa_{0,j}$ $(j = 1, 2, ...J)$ and given the coefficients $\widehat{\boldsymbol{\mu}}_s$ that parameterize the $\widehat{g}_s\left(\kappa_{t+1}\right)$, a non-linear solver is used in order to (simultaneously) solve for $\theta_{0,j}^*$ and $\kappa_{1,j}^*$ for the following two equations that describe the evolution of the system (where for simplicity we ignore the subscript j that denotes a particular initial condition):

$$\kappa_1\left(1 + \beta^{-\frac{1}{\gamma}}\alpha^{\left(1-\frac{1}{\gamma}\right)}\left(\frac{1+\widehat{g}_s\left(\kappa_1\right)}{\kappa_1}\right)^{\phi}\right) = \frac{(1-\alpha)\left(1+\theta_0\right)^{1-\alpha}\kappa_0^{\alpha}}{\eta\left(1+\varepsilon\theta_0\right)}$$

$$\beta^{-\frac{1}{\gamma}}\alpha^{\left(1-\frac{1}{\gamma}\right)}\left(\frac{1+\widehat{g}_s\left(\kappa_1\right)}{\kappa_1}\right)^{\phi} = \frac{\left[\left(\frac{1-\alpha}{\alpha}\right)\frac{(\varepsilon-1)}{(1+\varepsilon\theta_0)} - \alpha\right] - \frac{d\ln(1+\widehat{g}_s(\kappa_1))}{d\ln\kappa_1}\left[(1-\alpha) + \left(\frac{1-\alpha}{\alpha}\right)\frac{(\varepsilon-1)}{(1+\varepsilon\theta_0)}\right]}{\left\{\frac{1}{\gamma} + \alpha - \frac{\alpha}{\gamma}\right\} + \phi\frac{d\ln(1+\widehat{g}_s(\kappa_1))}{d\ln\kappa_1}}$$

where the term $\frac{d\ln(1+\widehat{g}(\kappa_1))}{d\ln\kappa_1}$ is given by equation (A.29). For each $\kappa_{1,j}$ obtained as the solution to the above equations, the non-linear solver can be used again in order to obtain $\left(\kappa_{2,j}, \theta_{1,j}^*\right)$ and repeat this step recursively up to some final period T.

Step 3. A measure of accuracy of the forecast $\widehat{g}_s\left(\kappa_t\right)$ and the optimal choice that solves the system of equations $\left(\theta_{t,j}^*\right)$ is constructed. We calculate the sum of squared residuals for simulation started by initial condition j (SSR_j) from the current optimally chosen levels (θ_t^*) and the levels that would be predicted directly by the forecasting function using the previous set of coefficients for all periods t. That is, we compute an error term (not the regression error term) obtained in (each of the J) current simulations of the equilibrium immigration quota $\left(\theta_{t,s}^*\right)$, which use perfect foresight and which assume that the future effects $\frac{d\ln(1+\theta_{t+1})}{d\ln\kappa_{t+1}}$ are given by (A.30). The error of prediction for each one of the J simulations (for each of the J initial conditions) when the expectation function uses coefficients of iteration s, for time t is given by $e_{t,j,s+1} = \theta_{t,j}^* - \widehat{g}\left(\kappa_t, \widehat{\mu}_s\right)$. Then $SSR_{j,s+1} = \sum_{t=0}^{T}\left(e_{t,j,s+1}\right)^2$. If this distance

[27] More than one point can be used (i.e $(\kappa_{0j}, \theta_{0j}^*)$ and $(\kappa_{1j}, \theta_{1j}^*)$ for the $j-th$ initial condition. We don't do this step because the next set of points are closer to each other (converging toward the steady state) and in practice didn't change the coefficients. When only the first point is used, we control exactly which points on the grid we want to approximate, which is a good thing when approximating numerically a particular function.

is at a minimum, then a solution has been found, but if not at a minimum, continue with the next step.[28]

Step 4. Given the new set of pairs $\left\{\theta^*_{0,j}, \kappa_{0,j}\right\}_{j=1}^{J}$, run the non-linear regression

$$\left(1+\theta^*_{0,j}\right) = \exp\left(\sum_{i=0}^{n} \mu_i \left(\ln\left(\kappa_{0,j}\right)\right)^i\right) + error$$

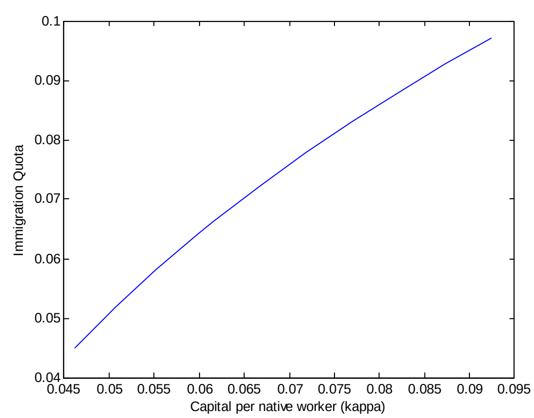
where the estimates for coefficients that parameterize this function when this is the $(s+1)$ time that this step is performed are given by $\widehat{\boldsymbol{\mu}}_{s+1}$. Given the new coefficients, go back to step 2 and repeat the steps until $\max\left[SSR_{1,s+1}, ..SSR_{J,s+1}\right]$ has been minimized.[29]

An example. Consider the case with the parameters as discussed in the text with $\alpha = \frac{1}{3}$, $\eta = \frac{1.25}{1.12}$, $\gamma = 4$, $\beta = .412$ and $\varepsilon = 1.9321$. We use $n = 3$ (a polynomial of the third degree) in order to minimize the distance. The steady state yields $\theta^* = 8.00\%$ and $\kappa^* = .0739$, with an elasticity $\frac{d\ln\left(1+\theta^*_{t+1}\right)}{d\ln \kappa_{t+1}}$ evaluated at steady state of 7.04%. In our experiments, adding more monomial terms to the expectation function doesn't result in more accuracy. The Markovian strategy of equilibrium is described in this case by

$$\left(1+\theta^*_t\right) = \exp\left(\sum_{i=0}^{3} \widetilde{\mu}_i \ln\left(\kappa_{t,j}\right)^i\right)$$

with coefficients $\{\widetilde{\mu}_0, \widetilde{\mu}_1, \widetilde{\mu}_2, \widetilde{\mu}_3\}$ given by {0.261032563347811, 0.0698516558436102, -0.000699103908625326, -0.000150142686729684}, and $max\left\{SSR_1, ...SSR_J\right\} = .5 * 10^{-15}$. The algorithm required 11 iterations in order to arrive at the solution where we used $T = 20$ and $N = 24$. The optimal policy function in Markov strategies for this particular example is shown below.

The optimal markovian policy function



[28] Alternatively, we can define this algorithm in terms of finding the coefficients $\boldsymbol{\mu}^*$ such that, using the forecast $g\left(\kappa; \boldsymbol{\mu}^*\right)$ the model equations yields a set of pairs $\left\{\kappa_j, \theta_j\right\}_j^J$ such that $\boldsymbol{\mu}^*$ also solves the regression problem. That is, returns the same coefficients used in the simulation of the model.

[29] The coefficients used for the perfect foresight simulation of the model (vector $\widetilde{\boldsymbol{\mu}}$) can be those directly dictated by the regression $(\widehat{\boldsymbol{\mu}})$, or as den Haan and Marcet do, a linear combination of the estimated coefficients for the current iteration and the estimates used in the previous iteration can be used. That is, for the i^{th} iteration if the regression coefficients are $\widehat{\boldsymbol{\mu}}_i$ and the coefficients that are fed into the model in the previous iteration are $\widetilde{\boldsymbol{\mu}}_{i-1}$, then the next iteration uses coefficients given by $\widetilde{\boldsymbol{\mu}}_i = \rho\widehat{\boldsymbol{\mu}}_i + (1-\rho)\widetilde{\boldsymbol{\mu}}_{i-1}$, for a specific $0 < \rho < 1$. For our model we can update the new coefficients directly $(\rho = 1)$.

Robustness. This algorithm is robust with respect to many variations in the procedure: to the number of periods T (Using $T = 10$ or $T = 200$ yields identical solutions in θ_t^* up to several decimal points), the number of initial conditions J (can use J of at least 4 and get very accurate results provided we use the same grid), as well as the number of pairs used for the regression procedure (can use pairs $\left\{\theta_{t,j}^*, \kappa_{t,j}\right\}_{j=1}^{J}$ for $t = \{0\}$ or $t = \{0, 1\}$, or $t = \{0, 1, 2\}$ in the regression and still obtain the same results). Variations in the size of the state grid don't seem to affect the steady state results even when we used a much smaller grid given by $\left[.99\kappa^1, 1.10\kappa^1\right]$.

Simplified Version. For $n = 1$, the projection (A.29) reduces to a log-linear approximation around the steady state. For purposes of computing steady states–our objects of main interest–the only relevant feature of g is the elasticity $\lambda^* = \frac{g'(\kappa^*)\kappa^*}{1+g(\kappa^*)} = \frac{d\ln(1+\theta^*)}{d\ln\kappa}$ at the steady state capital stock κ^*, because only λ_{t+1} appears in (25). Log-linearization yields an analytical solution for $g'(\kappa^*)$, which means $\{\kappa^*, \theta^*, \lambda^*\}$ can be written as a system of three non-linear equations that can be solved numerically (i.e., without having to approximate g away from κ^*). The log-linearizations turn out to provide values for θ^* that are very close to values obtained from solutions to the PEA algorithm. For example, using the parameters discussed above, the log-linearization yields $\theta^* = 7.98\%$, as compared to $\theta^* = 8.00\%$ with PEA.

G Analysis in Section 5.3

We claim in Section 5.3. that the CRRA model with persistence has a first order condition that can be used to compute optimal solutions.

Proof: In the CRRA model with persistence, $R_{t+1,j}$ depends on the state $j = I, II$, as in Section 4.1. Now $\theta_{t+1}^I = g_p(\kappa_{t+1})$ and $\theta_{t+1}^{II} = \theta_t$. The individual problem (A.8) again implies (A.9), where

$$B_t = \left[(\kappa_{t+1})^{(\gamma-1)(1-\alpha)}\beta\alpha^{(1-\gamma)}\left(\sum_j p_j \cdot (1+\theta_{t+1,j})^{(1-\gamma)(1-\alpha)}\right)\right]^{1/\gamma} = (\kappa_{t+1})^{\phi}E_t$$

is decomposed multiplicatively into a function of κ_{t+1} and the expectational term

$$E_t = \beta^{\frac{1}{\gamma}}\alpha^{-(1-\frac{1}{\gamma})}\left(\sum_j p_j \cdot (1+\theta_{t+1,j})^{(1-\gamma)(1-\alpha)}\right)^{\frac{1}{\gamma}}.$$

Note that $\kappa_{t+1} = \psi_t w_t \frac{B_t}{1+B_t}$, where $\psi_t = \frac{(1+\theta_t)}{\eta(1+\varepsilon\theta_t)}$. Hence

$$B_t = (\kappa_{t+1})^{\phi}E_t = (\psi_t w_t \frac{B_t}{1+B_t})^{\phi}E_t = z_t(\frac{B_t}{1+B_t})^{\phi}$$

is (implicitly) a function $B_t = B(z_t)$, where $z_t = (\psi_t w_t)^{\phi}E_t$.

Voters maximize $U_t = \frac{1}{1-\gamma}w_t^{1-\gamma}\left[1+B_t\right]^{\gamma}$ by choice of θ_t. This implies

$$\frac{dU_t}{d\theta_t} = \frac{(1-\gamma)U_t}{w_t}\frac{dw_t}{d\theta_t} + \frac{\gamma U_t}{1+B(z_t)}B'(z_t)\frac{dz_t}{d\theta_t}, \text{ or}$$
$$\frac{d\ln U_t}{(1-\gamma)d\theta_t} = \frac{d\ln w_t}{d\theta_t} - \frac{\epsilon_B(z_t)}{1-1/\gamma}\frac{d\ln z_t}{d\theta_t}$$

where $\epsilon_B(z_t) = \frac{B'(z_t)z_t}{1+B(z_t)} = \frac{B_t}{1-\phi+B_t}$. Hence interior solutions require

$$\frac{1}{\epsilon_B(z_t)} \frac{d\ln w_t}{d\theta_t} = \frac{1}{1-1/\gamma} \frac{d\ln z_t}{d\theta_t}. \tag{A.31}$$

Taking derivatives:

$$\frac{d\ln z_t}{d\theta_t} = \frac{d\ln E_t}{d\theta_{t+1,II}} + \frac{d\ln E_t}{d\theta_{t+1,I}} \frac{\partial\theta_{t+1,I}}{\partial\ln\kappa_{t+1}} \frac{\partial\ln\kappa_{t+1}}{\partial\theta_t} + \phi\left(\frac{d\ln w_t}{d\theta_t} + \frac{d\ln\psi_t}{d\theta_t}\right);$$

$$\frac{d\ln E_t}{d\theta_{t+1,II}} = -(1-\frac{1}{\gamma})\frac{1-\alpha}{1+\theta_t}P_{II}, \text{ where}$$
$$P_{II} = P_{II}(\theta^{II}_{t+1}, E_t) = p\frac{\beta\alpha^{(1-\gamma)}(1+\theta^{II}_{t+1})^{(1-\gamma)(1-\alpha)}}{E_t^\gamma};$$

$$\frac{d\ln E_t}{d\theta_{t+1,I}} \frac{\partial\theta_{t+1,I}}{\partial\ln\kappa_{t+1}} = \frac{1}{\gamma}(1-P_{II})(1-\gamma)(1-\alpha)\cdot\lambda_{t+1};$$

$$\ln\kappa_{t+1} = \ln\psi_t + \ln w_t + \ln B_t - \ln(1+B_t) \text{ implies}$$
$$\begin{aligned}\frac{\partial\ln\kappa_{t+1}}{\partial\theta_t} &= \left(\frac{d\ln w_t}{d\theta_t} + \frac{d\ln\psi_t}{d\theta_t}\right) + \frac{1}{B_t}\frac{\partial B_t}{\partial\theta_t} - \frac{1}{1+B_t}\frac{\partial B_t}{\partial\theta_t} \\ &= \left(\frac{d\ln w_t}{d\theta_t} + \frac{d\ln\psi_t}{d\theta_t}\right) + \frac{\epsilon_B(z_t)}{B_t}\frac{d\ln z_t}{d\theta_t}; \text{ and}\end{aligned}$$

$$\frac{d\ln w_t}{d\theta_t} + \frac{d\ln\psi_t}{d\theta_t} = \frac{\alpha}{(1+\theta_t)} + \frac{1}{(1+\theta_t)}\frac{\varepsilon-1}{(1+\varepsilon\theta_t)},$$

one obtains,

$$\begin{aligned}\frac{d\ln z_t}{d\theta_t} = &-(1-\frac{1}{\gamma})\frac{1-\alpha}{1+\theta_t}P_{II} + \phi\left(\frac{d\ln w_t}{d\theta_t} + \frac{d\ln\psi_t}{d\theta_t}\right) \\ &- (1-P_{II})\phi\lambda_{t+1}\left[\left(\frac{d\ln w_t}{d\theta_t} + \frac{d\ln\psi_t}{d\theta_t}\right) + \frac{\epsilon_B(z_t)}{B_t}\frac{d\ln z_t}{d\theta_t}\right],\end{aligned}$$

$$\frac{d\ln z_t}{d\theta_t} = \frac{-(1-\frac{1}{\gamma})\frac{1-\alpha}{1+\theta_t}P_{II} + \phi\left(\frac{\alpha}{(1+\theta_t)} + \frac{1}{(1+\theta_t)}\frac{\varepsilon-1}{(1+\varepsilon\theta_t)}\right)[1-(1-P_{II})\lambda_{t+1}]}{1+\frac{\phi}{1-\phi+B_t}(1-P_{II})\lambda_{t+1}\phi}.$$

provided $1+\frac{\phi}{1-\phi+B_t}(1-P_{II})\lambda_{t+1}\phi > 0$. Inserting into (A.31) and combining terms, the first order condition is

$$\frac{1}{\epsilon_B(z_t)}\frac{d\ln w_t}{d\theta_t} = \frac{\frac{1}{1-1/\gamma}\frac{\partial\ln E_t}{\partial\theta_t} + (1-a)\left(\frac{d\ln w_t}{d\theta_t} + \frac{d\ln\psi_t}{d\theta_t}\right)[1-(1-P_{II})\lambda_{t+1}]}{1+(1-P_{II})\lambda_{t+1}\phi\frac{\epsilon_B(z_t)}{B_t}}$$

Using $\frac{1}{\epsilon_B(z_t)} = \frac{1-\phi+B_t}{B_t} = 1+\frac{1-\phi}{B_t}$, this can be written as

$$\frac{1}{B_t} = \frac{\left(1-a+\left(\frac{1-\alpha}{\alpha}\right)\frac{\varepsilon-1}{(1+\varepsilon\theta_t)}\right)[1-(1-P_{II})\lambda_{t+1}] + \left(\frac{1-\alpha}{\alpha}\right)P_{II} - 1}{1-\phi+\phi(1-P_{II})\lambda_{t+1}}. \tag{A.32}$$

This first order condition generalizes (25). *QED.*

Note that in any steady state, $\theta^I_{t+1} = \theta^{II}_{t+1} = \theta^*$ implies $P_{II} = p$. Constant $B_t = B^*$ and $\theta_t = \theta^*$ imply

$$\frac{1}{B^*} = \frac{p\left(\frac{1-\alpha}{\alpha}\right) + \left(1 - a + \left(\frac{1-\alpha}{\alpha}\right)\frac{\varepsilon-1}{(1+\varepsilon\theta^*)}\right)[1-(1-p)\lambda^*] - 1}{1-\phi+(1-p)\lambda^*\phi}, \text{ and}$$

$$B^* = \beta^{\frac{1}{\gamma}}\left(\frac{1-\alpha}{\alpha\eta(1+\varepsilon\theta^*)}\frac{B^*}{1+B^*}\right)^{1-\frac{1}{\gamma}}$$

where $\lambda^* = \frac{g'(\kappa^*)\kappa^*}{1+g(\kappa^*)}$ depends on the policy function at κ^*. Hence given a numerical approximation for g, a steady state $\{B^*, \theta^*, \lambda^*\}$ can be characterized as solution to this system of non-linear equations. We obtain steady states numerically in two ways: by approximating g using PEA (as discussed above); and by deriving an analytical solution for $g'(\kappa^*)$ from log-linearizing (A.32) and (A.18), and then solving the steady state equations numerically. In our applications, both approaches yield virtually identical answers.