Estimated Human Capital Externalities in an Endogenous Growth Framework*

Jim Malley
University of Glasgow and CESifo

Ulrich Woitek
University of Zurich and CESifo

April 3, 2019

Abstract

To better understand the quantitative implications of human capital externalities at the aggregate level, we estimate a two-sector endogenous growth model with knowledge spill-overs. To achieve this, we account for trend growth in a model consistent fashion and employ a Markov-chain Monte-Carlo (MCMC) algorithm to estimate the model's posterior parameter distributions. Using U.S. quarterly data from 1964-2017, we find significant positive externalities to aggregate human capital. Our analysis further shows that eliminating this market failure leads to sizeable increases in education-time, endogenous growth and aggregate welfare.

Keywords: Human capital externalities, endogenous growth, Bayesian estimation

JEL codes: C11, C52, E32

*We would like to thank Konstantinos Angelopoulos, Joe Byrne, Fabio Canova, David DeJong, Macro del Negro, Richard Dennis, Bernardo Xavier Fernandez, Peter Ireland, Christos Koulovantianos, Chris Papageorgiou, Apostolis Philippopoulos, Peter Rosenkranz and Marcel Thum for helpful comments and suggestions on earlier versions of this paper. We also thank seminar participants at Dresden and Zurich Universities for helpful comments and the Royal and Scottish Economic Societies for financial support. The usual disclaimer applies.

1 Introduction

Positive human capital externalities associated with education have been discussed in the economics literature for the past several centuries. Notable contributions include Smith (1776, Book V, Chaper I), Bastiat (1851, Chapter 17), Marshall (1890, Book I, Chapter IV), Bastable (1892, Book I, Chapter V), Buchanan and Tullock (1962), Becker (1964) and Lucas (1988). Despite the fact that positive externalities to education provide the economic justification for public subsidies to this activity, there is surprisingly little consensus on their quantitative importance. Empirical estimates of the net returns to education for the U.S., typically based on Mincer equations for cities, states and the aggregate economy, range from 0 to 8%.[1]

To better understand the quantitative implications of human capital externalities at the aggregate level, this paper estimates an endogenous growth model based on Lucas (1988) and Tamura (1991) using Bayesian methods which have become popular in recent years for the estimation of business cycle models (see, e.g. Canova 2007, An and Schorfheide 2007, Del Negro and Schorfheide 2007, DeJong and Dave 2011, Herbst and Schorfheide 2016 and Fernández-Villaverde *et al.* 2016). Recent papers also employing these methods for applications involving growth models include Chang *et al.* (2007), Anzoategui *et al.* (2017), and Bianchi *et al.* (2018). However, to the best of our knowledge, this is the first study in the literature attempting to econometrically estimate the extent of aggregate externalites to human capital using an endogenous growth setup.

The main contribution of this paper is that we empirically identify significant positive aggregate externalities to human capital using quarterly U.S. data from 1964-2017. For example, we find that the pre-depreciation private returns to human capital are about 90% of the social returns.[2] We further find that if the social and private returns to education were equalised, discounted lifetime aggregate welfare, in terms of the compensating consumption sup-

[1]Net returns refer to the difference between social and private returns. For recent contributions to this literature, see, e.g., Acemoglu and Angrist (2001), Davies (2003), Moretti (2004a), Moretti (2004b), Moretti (2004c), Psacharopoulos and Patrinos (2004), Ciccone and Peri (2006), Lange and Topel (2006) and Guo *et al.* (2018).

[2]To the best of our knowledge, the only other papers providing quantitative evidence relating to aggregate externalites in the U.S. are Choi (2011) and Guo *et al.* (2018). Using calibration methods Choi (2011) finds that the pre-depreciation ratio of private to social returns to human capital is nearly 75%. In contrast, Guo *et al.* (2018) econometrically estimate that one more year of average schooling leads to a 6-8% increase in individual wages. They further show that the elasticity of a firm's productivity with respect to the average human capital of the economy is 0.121.

plement, would increase by about 8.7%.[3] This implies that education-time would increase by approximately 14% and the annual growth rate of human capital by about half a percent. The latter is non-trivial considering its cumulative effects on the per capita levels in the model. For example, these quantities would double roughly 25 years earlier in the model which internalises the social returns.

Since the highly stylised growth model we estimate may be misspecified in several dimensions and the measured data employed may not necessarily be informative, we conduct several cross-validation exercises to examine whether key model predictions cohere with the data and some stylised facts more broadly. These are carried out by comparing the model implied long-run trend for human capital accumulation and the model implied cyclical behaviour of education time directly with the data. Moreover, we further assess the robustness of our findings by re-estimating the model with a modified set of observables.

The rest of the paper is organised as follows. Sections 2 and 3 lay out the economic and econometric models respectively. Sections 4, 5 and 6 present the estimation results, welfare analysis and external cross-validation in turn and Section 7 contains the conclusions.

2 Endogenous growth model

In this Section, we solve for the optimal decisions of households and firms relying on the Lucas (1988) and Tamura (1991) setups. We incorporate both goods and human capital sectors as in Lucas (1988) but follow Tamura (1991) and introduce the externality into human capital production instead of goods production. The engine of long-term growth in this model is human capital accumulation in the presence of an aggregate externality to human capital. In particular, the externality implies that the social stock of human capital increases the productivity of individuals' educational choices. Moreover, the excess of social over private benefits implied by the externality leads to an under-investment in education and in turn human capital.

The general equilibrium solution consists of a system of dynamic relations, which jointly specify the paths of output, consumption, physical capital, human capital growth, and the fractions of time allocated to work and

[3] Of course, if distortionary tax and spending policy were used to publicly provide the inefficiently low investment in education implied by the externality, the welfare gains would be reduced. Whilst analysis relating to the cost side is outwith the scope of this paper, our results provide a benchmark of the potential gains associated with equating private and social returns to education.

education. Since the Lucas (1988) and Tamura (1991) models are well known, the main purpose of this Section is simply to fix ideas, notation and variable definitions which will be used in the estimation and analysis which follow.

To facilitate econometric estimation, our deliberately minimal deviations from the Lucas (1988) and Tamura (1991) setups include: (i) non-zero depreciation rates for physical and human capital;[4] and (ii) stochastic AR(1) processes for productivity in the goods and human capital sectors.

2.1 Households

The economy is populated by a large number of identical households indexed by the superscript h and identical firms indexed by the superscript f, where $h, f = 1, 2, ..., N_t$. The population size, N_t, evolves at a constant rate $n \geq 1$, so that $N_{t+1} = nN_t$, where N_0 is given. Each household's preferences are given by the following time-separable utility function:

$$E_0 \sum_{t=0}^{\infty} \beta^t U(C_t^h) \tag{1}$$

where E_t denotes the mathematical expectations operator conditional on information available at time t; C_t^h is consumption of household h at time t; and $0 < \beta < 1$ is the discount rate. The instantaneous $CRRA$ utility function is increasing, concave and satisfies the Inada conditions:

$$U_t = \frac{\left(C_t^h\right)^{1-\sigma}}{1-\sigma}, \tag{2}$$

where $1/\sigma$ ($\sigma > 1$) is the inter-temporal elasticity of substitution of consumption.

Each household h saves in the form of investment, I_t^h, and receives interest income, $r_t K_t^h$, where r_t is the return to capital and K_t^h is the beginning-of-period private capital stock. The household has one unit of time in each period t, which is allocated between work, u_t^h, and education, e_t^h, so that:

$$u_t^h + e_t^h = 1. \tag{3}$$

A household with a stock of human capital, H_t^h, receives labor income, $w_t u_t^h H_t^h$, where w_t is the wage rate and $u_t^h H_t^h$ is h's effective work-time.

[4]Non-zero depreciation rates are not only necessary given that we will be taking the model to the data but also in light of the calibration findings by Jones *et al.* (2005).

Finally, each household receives dividends paid by firms, Π_t^h. Accordingly, the budget constraint of each household is:

$$C_t^h + I_t^h = r_t K_t^h + w_t u_t^h H_t^h + \Pi_t^h. \tag{4}$$

Each household's physical and human evolve according to:

$$K_{t+1}^h = (1-\delta^k)K_t^h + I_t^h; \tag{5}$$

and

$$H_{t+1}^h = (1-\delta^h)H_t^h + B_t \left(e_t^h H_t^h\right)^\theta \left(\overline{H}_t\right)^{1-\theta}, \tag{6}$$

where $0 \leq \delta^k, \delta^h \leq 1$ are constant depreciation rates on private physical and human capital respectively; and $B_t \left(e_t^h H_t^h\right)^\theta \left(\overline{H}_t\right)^{1-\theta}$ is "new" human capital created at time period t.

More specifically, B_t represents human capital productivity; $e_t^h H_t^h$ is effective education-time; $\overline{H}_t$ is the average (per household) human capital stock in the economy; and the parameters $0 < \theta, (1-\theta) < 1$ capture the efficiency of household human capital and the aggregate human capital externality respectively.[5] Households act competitively by taking market prices and aggregate outcomes as given. Thus, each household chooses $\{C_t^h, u_t^h, e_t^h, I_t^h, K_{t+1}^h, H_{t+1}^h\}_{t=0}^\infty$ to maximize (1) subject to (3)-(6), and initial conditions for the two capital stocks and the two productivity terms.

The familiar static optimality condition for consumption, C_t^h:

$$\Lambda_t^a = \left(C_t^h\right)^{-\sigma}, \tag{7}$$

states that the shadow price associated with (4), Λ_t^a, is equal to the marginal value of consumption at time t.

The Euler-relation for private physical capital, K_{t+1}^h:

$$\Lambda_t^a = \beta E_t \left[\Lambda_{t+1}^a \left(r_{t+1} + 1 - \delta^k\right)\right], \tag{8}$$

denotes that marginal cost of forgone consumption at time t is equal to the expected marginal benefit of discounted $t+1$ returns derived from investing in one unit of physical capital at time t.

The static optimality condition for time spent on education, e_t^h:

$$\Lambda_t^b = \frac{\Lambda_t^a w_t H_t^h}{B_t \theta \left(e_t^h\right)^{\theta-1} \left(H_t^h\right)^\theta \left(\overline{H}_t\right)^{1-\theta}}, \tag{9}$$

[5]The assumption that individual human capital accumulation is an increasing function of the per capita level of economy-wide human capital captures the idea that the existing know-how of the economy provides an external positive effect. Equivalently it can be thought of as a learning-by-doing effect as discussed in Romer (1986).

implies that the shadow price associated with (6), Λ_t^b, is equal to the marginal value of education at time t. In other words, this value is equal to the ratio of the marginal cost to the marginal return to time spent in education.

The Euler-equation for private human capital, H_{t+1}^h:

$$\begin{aligned}\Lambda_t^b &= E_t\beta\Lambda_{t+1}^a w_{t+1}(1-e_{t+1}^h) + \\ &\quad E_t\beta\Lambda_{t+1}^b\left[1-\delta^h+B_{t+1}\theta\left(e_{t+1}^h\right)^\theta\left(H_{t+1}^h\right)^{\theta-1}\left(\overline{H}_{t+1}\right)^{1-\theta}\right], \quad (10)\end{aligned}$$

maintains that the marginal cost of forgone labor income at time t is equal to the marginal benefit of $t+1$ returns to working plus the marginal $t+1$ returns to investing in one unit of human capital at time t.

2.2 Firm's problem

To produce its homogenous final product, Y_t^f, each firm employs private physical capital, K_t^f, and effective labor, $u_t^f H_t^f$. Thus, the production function of each firm is:

$$Y_t^f = A_t\left(K_t^f\right)^\alpha\left(u_t^f H_t^f\right)^{1-\alpha}, \quad (11)$$

where A_t represents the level of Hicks-neutral technology available to all firms, $0<\alpha<1$ and $(1-\alpha)$ are the efficiency of private capital and effective labor respectively.

Firms act competitively by taking prices and aggregate outcomes as given. Accordingly, subject to (11), each firm chooses K_t^f and $u_t^f H_t^f$ to maximize a series of static profit functions:

$$\Pi_t^f = Y_t^f - r_t K_t^f - w_t u_t^f H_t^f. \quad (12)$$

The resulting familiar first-order conditions:

$$\frac{(1-\alpha)Y_t^f}{u_t^f H_t^f} = w_t; \quad (13)$$

$$\frac{\alpha Y_t^f}{K_t^f} = r_t, \quad (14)$$

state that the firm will hire labor until the marginal product of effective labor is equal to the wage rate, w_t, and will rent capital until the marginal product of physical capital is equal to the rental rate, r_t. Finally, given the assumption of constant returns to scale in production at the firm level, factor payments exhaust the value of output, implying no economic profits are earned.

2.3 Decentralised competitive equilibrium (DCE)

The DCE is obtained when (i) households and firms optimize, as above, taking prices and aggregate outcomes as given; (ii) all constraints are satisfied; and (iii) all markets clear, i.e. $\sum K_t^h = \sum K_t^f$, $\sum(1-e_t^h)H_t^h = \sum u_t^f H^f$, $\sum \Pi_t^h = \sum \Pi_t^f = 0$. Given the N_t identical households at time period t and also N_t identical firms, economy wide magnitudes are denoted $X_t = N_t X_t^h = N_t X_t^f$. Since human capital is the engine of long-run endogenous growth, we transform variables to make them stationary, e.g. we first define *per capita* quantities for any variable X as $\overline{X}_t \equiv X_t/N_t$, where $X_t \equiv (Y_t, C_t, I_t, K_t, H_t)$ and then express these as shares of per capita human capital, e.g. $x_t \equiv \overline{X}_t/\overline{H}_t$. Finally, the gross human capital growth rate is defined as $\gamma_t \equiv \overline{H}_{t+1}/\overline{H}_t$. Using this notation and substituting out prices, $\{r_t, w_t\}_{t=0}^{\infty}$, we obtain the following stationary DCE:

$$\begin{aligned}
y_t &= c_t + n\gamma_t k_{t+1} - \left(1-\delta^k\right) k_t; \\
y_t &= A_t \left(k_t\right)^{\alpha} \left(1-e_t\right)^{(1-\alpha)}; \\
n\gamma_t &= 1-\delta^h + B_t \left(e_t\right)^{\theta}; \\
\lambda_t^a &= \left(c_t\right)^{-\sigma}; \\
\lambda_t^a &= \beta \left(\gamma_t\right)^{-\sigma} E_t \left[\lambda_{t+1}^a \left(\alpha \frac{y_{t+1}}{k_{t+1}} + 1 - \delta^k\right)\right]; \\
\lambda_t^b &= \frac{\left(c_t\right)^{-\sigma}(1-\alpha)y_t}{B_t\left(1-e_t\right)\theta\left(e_t\right)^{\theta-1}}; \\
\lambda_t^b &= \beta\left(\gamma_t\right)^{-\sigma}\{E_t\left[\left(c_{t+1}\right)^{-\sigma}(1-\alpha)y_{t+1}\right] + \\
&\quad + E_t\lambda_{t+1}^b\left[\left(1-\delta^h\right) + B_{t+1}\theta\left(e_{t+1}\right)^{\theta}\right]\},
\end{aligned} \tag{15}$$

where λ_t^a and λ_t^b are the transformed shadow prices associated with (4) and (6) respectively in the household's problem.[6] Therefore, the stationary DCE is summarised by the above system of seven equations in the paths of the following seven variables: (γ_t, y_t, c_t, e_t, k_{t+1}, λ_t^a, λ_t^b) given the paths of the exogenously set stationary AR processes whose motion is defined below.

2.4 Processes for productivity

Given the above set-up, Hicks-neutral productivity, A_t, and human capital productivity, B_t, are stochastic. Following the usual practice in the RBC

[6]Note that $\lambda_t^a = \Lambda_t^a/\overline{H}_t^{-\sigma}$ and $\lambda_t^b = \Lambda_t^b/\overline{H}_t^{-\sigma}$ where h-superscripts have been omitted since we are in a symmetric equilibrium.

literature, we assume that each follows an AR(1) process:[7]

$$\begin{aligned} A_t &= A^{(1-\rho^a)} A_{t-1}^{\rho^a} e^{\varepsilon_t^a} \\ B_t &= B^{(1-\rho^b)} B_{t-1}^{\rho^b} e^{\varepsilon_t^b} \end{aligned} \tag{16}$$

where $A, B > 0$ are constants, $0 < \rho^a, \rho^b < 1$ are the autoregressive parameters and $\varepsilon_t^a, \varepsilon_t^b$ are normally distributed random shocks with zero means and variances equal to σ_a^2 and σ_b^2 respectively.

Innovations to TFP affect the efficiency of both capital and effective labor, whereas shocks to human capital productivity are purely labor augmenting. DeJong and Ingram (2001, p. 541-42) argue that B_t can be thought of as "[...] an exogenous shock that shifts the efficiency with which hours are transformed into human capital. Examples of a negative shock are the creation of a new computer operating system that is more difficult to learn than the previous system and a decrease in funding for government-sponsored training programs. A positive shock could be a technological improvement in employee training methods".

2.5 Model solution

Following Klein (2000), we obtain the solution of the linearised stationary DCE (see, Appendix A, eq. 38) in state-space form:

$$\hat{\mathbf{y}}_t = \mathbf{Z}^{\mathrm{C}} \boldsymbol{\alpha}_t^{\mathrm{C}}; \tag{17a}$$

$$\boldsymbol{\alpha}_{t+1}^{\mathrm{C}} = \mathbf{T}^{\mathrm{C}} \boldsymbol{\alpha}_t^{\mathrm{C}} + \mathbf{R}^{\mathrm{C}} \boldsymbol{\eta}_{t+1}; \boldsymbol{\eta}_t \sim N(\mathbf{0}, \mathbf{Q}), \tag{17b}$$

where (17 a) is the measurement equation linking the vector of stationary observables $\hat{\mathbf{y}}_t = [\hat{y}_t \;\; \hat{c}_t \;\; \hat{u}_t]'$ to the unobserved state vector $\boldsymbol{\alpha}_t^{\mathrm{C}} = \left[\hat{k}_t \;\; \hat{a}_t \;\; \hat{b}_t\right]'$ and the matrix $\mathbf{Z}^{\mathrm{C}}$ contains convolutions of the model's parameters (see, Appendix A Table 5).[8] The transition equation (17 b) is the law of motion for the state vector $\boldsymbol{\alpha}_t^{\mathrm{C}}$. The matrix $\mathbf{T}^{\mathrm{C}}$ again contains convolutions of the model's parameters and $\mathbf{R}^{\mathrm{C}}$ is a matrix of zeros and ones controlling the two productivity shocks in the error vector $\boldsymbol{\eta}_t$. This vector is assumed to follow a multivariate normal distribution with zero mean and variance-covariance matrix $\mathbf{Q}$. Finally, to distinguish from other model components to be added below, superscript 'C' denotes the stationary solution.

[7]See, e.g. Kim and Lee (2007), DeJong and Ingram (2001) and Perli and Sakellaris (1998) for similar setups for the two productivity processes.

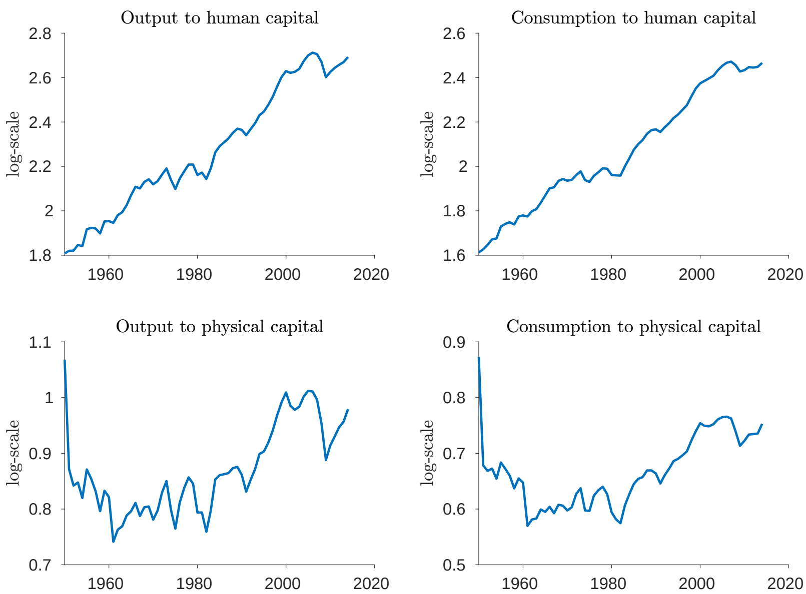
[8]Note that for any variable x_t, $\hat{x}_t = ln(x_t/x)$ and x is the model-consistent steady-state value of x_t.

2.6 Data

The measurement equation given by (17 a) requires stationary observables for output, consumption and hours. The quarterly data for these series are obtained from the St. Louis Federal Reserve FRED database, for the period 1964(1) to 2017(3). Output is the sum of real consumption plus real gross private domestic investment, both in billions of chained 2009 dollars. Hours-worked is measured by the hours of wage and salary workers on non-farm payrolls for the private sector. The civilian noninstitutional population is used to derive the per capita data. All series, except the population figures, are seasonally adjusted. It is well documented that for these three per capita series, only per capita hours is a stationary mean reverting series.

To address non-stationarity in the data when estimating business cycle models, the literature has adopted a number of different approaches, such as: (i) HP-filtering the data; (ii) incorporating a common trend in the model/estimation (see, e.g. Smets and Wouters 2007 and Fernández-Villaverde *et al.* 2016); and (iii) assuming agnostic local-linear trends (see, e.g. Ferroni 2011 and Canova 2014). In contrast to these approaches, where either the deterministic or stochastic trends are exogenous, we need to account for endogenous trend growth in a model consistent fashion.

Figure 1: Normalised Output and Consumption, 1950-2014



Whilst we could de-trend (normalise) the observable non-stationary data by another non-stationary series required by the model, Canova and Ferroni (2011) point out that "model-driven" filtering is typically problematic. Although quarterly data is not available for this purpose, we confirm this point in Figure 1 above which provides evidence of the non-stationarity of output and consumption when we de-trend using annual human capital (upper row) and annual physical capital data (lower row). The later is also used, since the DCE can be re-derived in per physical capital units.[9]

3 Econometric setup

In this Section we first discuss how we incorporate the model's endogenous trend into the estimation. We then set out our priors for the model's parameters. Details on the estimation procedure can be found in Appendix B.

3.1 Model consistent trend

Recall from Section 2.3 that stationary quantities in the model are defined as follows:

$$\hat{\mathbf{y}}_t = \ln(\mathbf{Y}_t) - \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 0 \end{pmatrix} \ln(\bar{H}_t) - \begin{pmatrix} \ln(y) \\ \ln(c) \\ \ln(u) \end{pmatrix}. \tag{18}$$

The endogenous trend $\ln(\bar{H}_t)$ is given by:

$$\ln(\bar{H}_t) = \ln(\gamma_t) + \ln(\bar{H}_{t-1}), \tag{19}$$

where $\ln(\gamma_t) = \hat{\gamma}_t + \ln(\gamma)$. To obtain stationary $\hat{\mathbf{y}}_t$ in eq. (18), requires that we remove the trend $\bar{H}_t$ from output and consumption. Using the model solution in eq. (17 a) and (17 b), this can be achieved by writing eq. (18) in first differences so that $\bar{H}_t$ drops out:

$$\Delta \ln(\mathbf{Y}_t) = \Delta\hat{\mathbf{y}}_t + \begin{pmatrix} 1 \\ 1 \\ 0 \end{pmatrix} \hat{\gamma}_t + \begin{pmatrix} 1 \\ 1 \\ 0 \end{pmatrix} \ln(\gamma). \tag{20}$$

[9]The human capital and physical capital data are obtained from FRED database. Human capital is an index on a per person basis for U.S. and the real physical capital stock is in 2011 prices (see, Feenstra *et al.* 2015). Given that human capital is in per capita terms, per capita output and consumption were used when constructing the plots in the first row of Figure 1.

In our case, $\hat{\gamma}_t$ is defined as[10]

$$\hat{\gamma}_t = \left(\frac{B\theta e^{\theta}}{n\gamma}\right)\hat{e}_t + \left(\frac{Be^{\theta}}{n\gamma}\right)\hat{b}_t, \tag{21}$$

which implies that the logged gross growth rate, $ln(\gamma_t)$, depends on the deviations of education-time, $\hat{e}_t$, and human capital productivity, $\hat{b}_t$, from their respective steady-states. Whilst $\hat{e}_t = -\left(\frac{u}{1-u}\right)\hat{u}_t$ can be implied from observable hours, $\hat{b}_t$ is not observable and is an element of the state vector $\boldsymbol{\alpha}_t^{\mathrm{C}}$ in eqs. (17 a) and (17 b). Therefore, eq. (20) becomes

$$\begin{aligned}\Delta\ln(\mathbf{Y}_t) =& \hat{\mathbf{y}}_t - \hat{\mathbf{y}}_{t-1} + \mathbf{C}_1\hat{\mathbf{y}}_t + \mathbf{C}_2\boldsymbol{\alpha}_t^{\mathrm{C}} + \begin{pmatrix}1\\1\\0\end{pmatrix}\ln(\gamma) = \\ &= \left((\mathbf{I}_3 + \mathbf{C}_1)\mathbf{Z}^C + \mathbf{C}_2\right)\boldsymbol{\alpha}_t^C - \mathbf{Z}^C\boldsymbol{\alpha}_{t-1}^C + \begin{pmatrix}1\\1\\0\end{pmatrix}\ln(\gamma),\end{aligned} \tag{22}$$

where

$$\mathbf{C}_1 = \begin{pmatrix}0 & 0 & -\frac{B\theta e^{\theta}}{n\gamma}\frac{u}{1-u}\\ 0 & 0 & -\frac{B\theta e^{\theta}}{n\gamma}\frac{u}{1-u}\\ 0 & 0 & 0\end{pmatrix}; \mathbf{C}_2 = \begin{pmatrix}0 & 0 & \frac{Be^{\theta}}{n\gamma}\\ 0 & 0 & \frac{Be^{\theta}}{n\gamma}\\ 0 & 0 & 0\end{pmatrix}.$$

Finally, to capture measurement errors as well as movements and comovements in the data which cannot be captured by the model, we follow one strand of the literature and extend the model (22) with a (3×1) measurement/specification error (MSE) vector $\boldsymbol{\nu}_t$.[11] Thus, the model to be estimated is given by:

$$\begin{aligned}\Delta\ln(\mathbf{Y}_t) - \begin{pmatrix}1\\1\\0\end{pmatrix}\ln(\gamma) &= \left((\mathbf{I}_3 + \mathbf{C}_1)\mathbf{Z}^{\mathrm{C}} + \mathbf{C}_2\right)\boldsymbol{\alpha}_t^{\mathrm{C}} - \mathbf{Z}^{\mathrm{C}}\boldsymbol{\alpha}_{t-1}^{\mathrm{C}} + \boldsymbol{\nu}_t;\\ \begin{pmatrix}\boldsymbol{\alpha}_{t+1}^{\mathrm{C}}\\ \boldsymbol{\alpha}_t^{\mathrm{C}}\end{pmatrix} &= \begin{pmatrix}\mathbf{T}^{\mathrm{C}} & \mathbf{0}\\ \mathbf{I}_3 & \mathbf{0}\end{pmatrix}\begin{pmatrix}\boldsymbol{\alpha}_t^{\mathrm{C}}\\ \boldsymbol{\alpha}_{t-1}^{\mathrm{C}}\end{pmatrix} + \mathbf{R}^{\mathrm{C}}\boldsymbol{\eta}_{t+1};\\ \boldsymbol{\nu}_{t+1} &= \mathbf{T}^{\nu}\boldsymbol{\nu}_t + \boldsymbol{\upsilon}_{t+1};\\ \boldsymbol{\eta}_t &\sim N(\mathbf{0},\mathbf{Q}); \boldsymbol{\upsilon}_t \sim N(\mathbf{0},\boldsymbol{\Sigma}).\end{aligned} \tag{23}$$

Each of the elements in $\boldsymbol{\nu}_t$ follows an AR(1) process. Therefore, $\mathbf{T}^{\nu}$ and the variance-covariance matrix $\boldsymbol{\Sigma}$ are diagonal matrices.

[10]See the linearised DCE and the parameter definitions in Appendix A eq. (38).

[11]See the discussion in Ireland (2004), p. 1209-1210 for different ways to deal with the stochastic singularity problem and references therein.

3.1.1 Priors for the parameters

The priors reported in Table 1 above regarding the supports for the prior distributions of the model's remaining parameters reflect non-sample information from: (i) the human capital model derived above; (ii) the empirical literature; (iii) technical considerations regarding the existence of a unique steady-state equilibrium and the saddle path stability of the dynamic system; and (iv) empirical considerations regarding the value of long-run human growth in the historical data.[12]

To help contextualize the quarterly rates used in Table 1, note that their annual counterparts are as follows: δ^k=(0.0456, 0.0504) and δ^h= (0.0172, 0.0188). The priors for the size of the depreciation rates and the relationship between them, i.e. $\delta^k > \delta^h$ reflect the findings of Jorgenson and Fraumeni (1989) and Jones *et al.* (2005).

In general, a uniform prior distribution was employed to be as agnostic as possible about the parameters. To help with the curvature of the posterior likelihood function, we implemented a relatively uninformative beta prior (mean: 0.5, standard deviation: 0.2) for θ and all the AR parameters.

Table 1: Priors for the parameters, $\boldsymbol{\psi}$

	Parameters	Restrictions
K depreciation rate	δ^k	$U\,(0.0114, 0.0126)$
H depreciation rate	δ^h	$U\,(0.0043, 0.0047)$
utility function param.	σ	$U\,(1, 3)$
s.d. A shock	σ^a	$U\,(0, 0.224)$
s.d. B shock	σ^b	$U\,(0, 0.224)$
externality parameter	$(1-\theta)$	Beta(2.625, 2.625)
AR(1) parameter in A_t	ρ^a	Beta(2.625, 2.625)
AR(1) parameter in B_t	ρ^b	Beta(2.625, 2.625)
constant term in A_t	A	$U\,(0, 1)$
constant term in B_t	B	$U\,(0, 1)$
model solution:		equilibrium exists
measurement/specification errors:		
AR(1) parameters	ρ^y, ρ^c, ρ^u	Beta(2.625, 2.625)
variances	$\Sigma_{yy}, \Sigma_{cc}, \Sigma_{hh}$	$\Sigma > 0$

[12] Note that estimating the discount factor, β, and capital's share, α, led to unreasonably low and high values respectively. Thus, we fixed these parameters to 0.99 and to 0.33. Moreover, the steady-state γ at a quarterly rate is fixed to 1.0025 (i.e. 1.01 annually) which is its average value over the measurement period.

4 Estimation results

The results presented in this Section are based on 100,000 draws from the posterior distribution using the Chib and Ramamurthy (2010).[13] We allow for a burn-in phase of 20,000 replications, and keep every 10th draw. Table 2 summarizes the estimated parameter distributions by reporting their means, standard deviations, a measure of estimation accuracy based on numerical standard errors, NSE (see Geweke, 1992), and the quantiles of the parameter distribution. Table 3 and Figures 2 and 3 then describe some characteristics of the posterior distributions which complement the results reported in Table 2.[14] We also present trace plots of the model's parameters to assess convergence of the parameter chains in Appendix C (see Figures 7 and 8).[15]

4.1 Posterior distributions of parameters

If we start with the means of the posterior parameter distributions shown in Table 2, it's first important to observe that the non-zero estimated posterior mean for $(1-\theta)$ suggests that a positive aggregate human capital externality is supported by the data. As discussed in the introduction, except for the studies by Choi (2011) and Guo *et al.* (2018), there has been no previous robust empirical evidence corroborating the presence of an aggregate externality to human capital in the U.S..[16]

Turning to the AR processes, note that only process for goods technology implies half-life persistence in excess of a year. For example, the half-lives for the two technology shocks, A and B-shocks, are 17.56 and 0.89 years respectively. For the specification/measurement errors, they are 0.21, 0.05 and 0.55 years for output, consumption, and hours respectively.

[13]Note that an extensive search was conducted to find starting values for the parameters. Although the Chib and Ramamurthy (2010) algorithm is much more complex than the random-walk Metropolis-Hastings, it improves efficiency significantly (see, Chib and Ramamurthy 2010, Section 3.1.1).

[14]The ratio shown in column 4 of Table 2 is in percent terms, i.e. $\frac{NSE_i}{|\bar{\psi}_i|} \times 100$. The NSEs shown in this Table are based on a 15 per cent taper for the periodogram window.

[15]Note that the trace plots in Appendix C display all the draws after discarding 20,000 replications as burn in.

[16]We also estimated this model from 1948(1) to 2002(2) using the data set employed in Ireland (2004) and Malley and Woitek (2010) and found robust evidence of the aggregate human capital externality.

Table 2: Posterior distribution of parameters, $\boldsymbol{\psi}$

	Location and Spread			Quantiles		
	mean	s.d.	NSE	%25	%50	%75
A	0.58768	0.27242	0.51823	0.39018	0.60522	0.81887
B	0.02066	0.00181	0.09784	0.01940	0.02046	0.02171
δ_k	0.01184	0.00047	0.04425	0.01146	0.01152	0.01245
δ_h	0.00450	0.00014	0.03497	0.00436	0.00448	0.00463
σ	2.02778	0.56062	0.30909	1.54355	1.97965	2.53595
$1-\theta$	0.24224	0.04920	0.22704	0.20716	0.24141	0.27545
ρ^a	0.99018	0.00375	0.00423	0.98815	0.99056	0.99270
ρ^b	0.82365	0.13856	0.18807	0.71673	0.82786	0.96798
σ^a	0.00787	0.00037	0.05211	0.00762	0.00787	0.00812
σ^b	0.00235	0.00129	0.61493	0.00148	0.00197	0.00291
ρ_y	0.43596	0.22704	0.58221	0.23653	0.44494	0.64962
ρ_c	0.04138	0.01860	0.50252	0.02791	0.03885	0.05223
ρ_u	0.72843	0.04478	0.06873	0.69913	0.72996	0.75923
Σ_{yy}	0.00063	0.00052	0.92829	0.00024	0.00050	0.00090
Σ_{cc}	0.00433	0.00020	0.05163	0.00420	0.00433	0.00446
Σ_{uu}	0.00372	0.00032	0.09492	0.00350	0.00371	0.00393

Further note that there are generally no surprises regarding the sizes of the means of the remaining parameters given our priors regarding the supports for the various distributions. Examination of the trace plots and cumulative means in Appendix C suggest that the chain for each of the parameters reported in Table 2 has converged.

With respect to the spread of the parameter distributions reported in Table 2, it generally appears that the structural and all AR parameters are quite concentrated. Moreover, examination of the numerical standard errors as a share of the absolute value of the means of the posteriors reveals that our estimates are generally very precise.

4.2 Priors versus posteriors

To gain a quantitative sense of what we have learned from the data, we next present the quantiles of the prior and posterior distributions in Table 3. We then plot, for each parameter, the percentage difference between the prior and posterior median in the top panel of Figure 2 as well as the percentage difference between the prior and posterior inter-quartile range (IQR) in the bottom panel of Figure 2.

Table 3: Quantiles of Prior and Posterior Distributions

	Prior			Posterior		
	25%	50%	75%	25%	50%	75%
A	0.25000	0.50000	0.75000	0.39018	0.60522	0.81887
B	0.25000	0.50000	0.75000	0.01940	0.02046	0.02171
δ_k	0.00600	0.01200	0.01800	0.01146	0.01152	0.01245
δ_h	0.00225	0.00450	0.00675	0.00436	0.00448	0.00463
σ	1.00000	2.00000	3.00000	1.54355	1.97965	2.53595
$1-\theta$	0.34934	0.50000	0.65066	0.20716	0.24141	0.27545
ρ^a	0.34934	0.50000	0.65066	0.98815	0.99056	0.99270
ρ^b	0.34934	0.50000	0.65066	0.71673	0.82786	0.96798
σ^a	0.05590	0.11180	0.16771	0.00762	0.00787	0.00812
σ^b	0.05590	0.11180	0.16771	0.00148	0.00197	0.00291
ρ_y	0.34934	0.50000	0.65066	0.23653	0.44494	0.64962
ρ_c	0.34934	0.50000	0.65066	0.02791	0.03885	0.05223
ρ_u	0.34934	0.50000	0.65066	0.69913	0.72996	0.75923
Σ_{yy}				0.00024	0.00050	0.00090
Σ_{cc}		positive		0.00420	0.00433	0.00446
Σ_{uu}				0.00350	0.00371	0.00393

Both sets of information suggest that the data are indeed informative about the parameters since the location and spread of the posterior distributions are generally considerably different than the prior distributions. For example, Figure 2 shows that for the 13 structural parameters, only four are less than 10% different than the prior medians. Even more striking is the difference between the IQRs. In particular, the vast preponderance of these measures of spread for the posterior distributions are at least 80% smaller than their respective priors.

Finally, recall that, except for the AR parameters, $(1-\theta)$ is the only parameter for which we did not assume a uniform prior. In this case we assumed a Beta distribution with mean 0.5 and standard deviation 0.2. Comparing the posterior with the prior distribution in Figure 3 shows that, the priors are indeed not overly restrictive, and that the data appear to be very informative about this parameter.

Figure 2: Comparison of Prior and Posterior Distributions

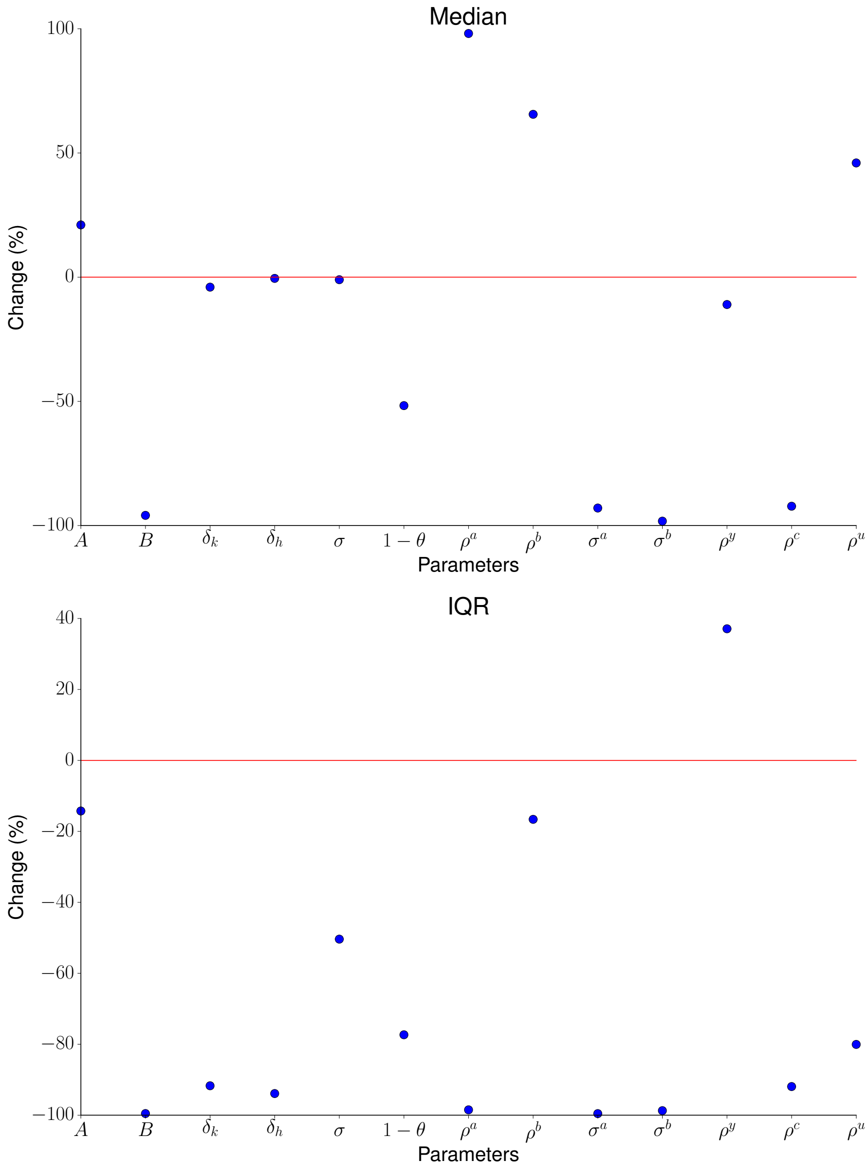
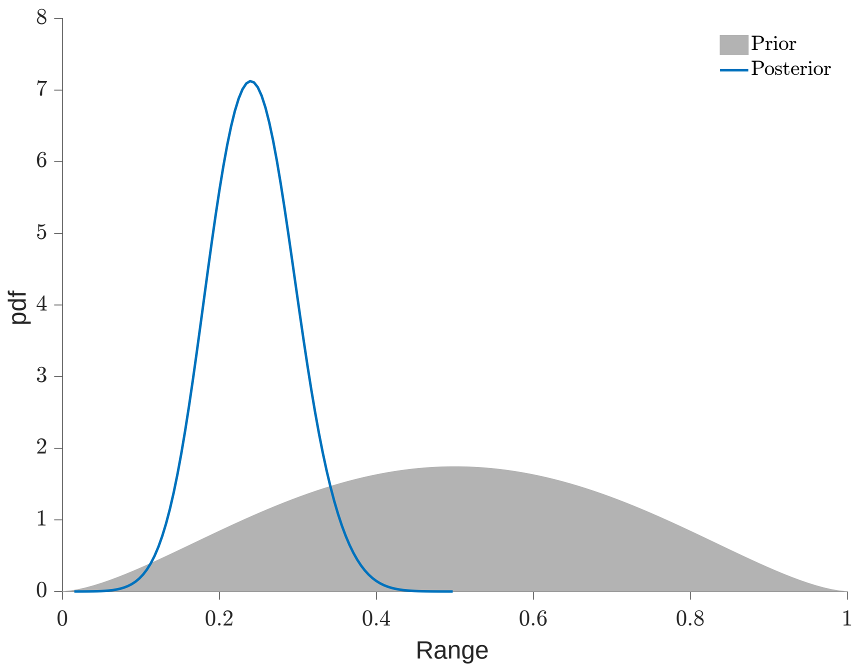


Figure 3: Posterior Plots, $1 - \theta$



5 Welfare analysis

Given that we have found econometric evidence in support of an aggregate human capital externality, it would be useful to have a quantitative sense of the difference between private and social returns to human capital implied by our estimation. In turn, it would be informative to quantify how much aggregate welfare, education-time and human capital growth would change if this ratio was equal to unity.

5.1 Inefficient versus efficient allocations

As shown in Section 2.1, the average level of human capital in the economy positively affects the productivity of education-time. This aggregate externality implies an excess of social to private returns to human capital, which in turn causes an inefficient under-investment in education and hence in human capital. Since there are no other market failures in the model, removing the externality implies the efficient allocation.

To obtain the efficient allocation requires that we: (i) set $\overline{H}_t$ equal to

H_t in the human capital eq. (6); (ii) re-derive the two non-stationary FOCs given by eqs. (9) and (10) which imply that the last equation in the stationary DCE given by eq. (15) must also be re-derived; and (iii) resolve the model as described in Section 2.

5.2 Private and social returns to human capital

To quantify the proportion of social returns to human capital which are not internalised, we next: (i) establish the equality between returns to physical and human capital in equilibrium; (ii) calculate the private return to human capital using the inefficient model; and (iii) calculate the social return to human capital using the efficient model. Recall that, in equilibrium, individual h superscripts are dropped since all markets clear and $H_t = \overline{H}_t$. Moreover, on the balanced-growth path (BGP), time subscripts are dropped for stationary variables.

To undertake (i) and (ii), we first substitute the static FOC for consumption given by eq. (7) into the Euler for capital in eq. (8) and evaluate the result along the BGP to obtain the familiar condition relating discounted net consumption growth, $\left(\frac{1}{\beta}\gamma_c - 1\right)$, to the return to physical capital net of depreciation, $\left(r - \delta^k\right)$, or the net marginal product of capital, $\left(\alpha Y/K - \delta^k\right)$:

$$\left(\frac{1}{\beta}\gamma_c - 1\right) = r - \delta^k. \tag{24}$$

Next, we substitute the static FOC for consumption given by eq. (7) into the static optimality condition for education given by eq. (9) and the resulting expression into the Euler-equation for human capital in eq. (10). Evaluating this result along the BGP gives an expression which relates discounted net consumption growth to the marginal product of human capital net of depreciation:

$$\left(\frac{1}{\beta}\gamma_c - 1\right) = B\theta\left(e\right)^{\theta-1} - \delta^h. \tag{25}$$

Equations (24) and (25) thus imply that, in equilibrium, the net returns to physical and human capital are equal:

$$r - \delta^k = B\theta\left(e\right)^{\theta-1} - \delta^h. \tag{26}$$

Repeating the above for the efficient model in which social and private returns are equal, we find an analogous no-arbitrage condition between physical and human capital:

$$r^* - \delta^k = B[\theta\left(e^*\right)^{\theta-1}\left(1 - e^*\right) + \left(e^*\right)^{\theta}] - \delta^h, \tag{27}$$

where a $*$ superscript denotes steady-state quantities in the efficient model. Thus, to calculate the wedge between private, P, and social returns to human capital, S, we compute the following ratios post- and pre-depreciation on an annual basis:

$$\frac{P_{post}}{S_{post}} = \frac{r - \delta^k}{r^* - \delta^k} = 0.86 \tag{28}$$

$$\frac{P_{pre}}{S_{pre}} = \frac{r - \delta^k + \delta^h}{r^* - \delta^k + \delta^h} = 0.89 \tag{29}$$

The calibration study of Choi (2011) finds that the ratio of private to social returns to human capital post- and pre-depreciation are 0.56 and 0.73 respectively in his benchmark calibration. Whilst, the size of knowledge externalites, as measured by the non-internalised returns to human capital, $(1 - \frac{P}{S})$, are clearly smaller in our econometric estimation, we will see below that they nonetheless imply large welfare gains when the externality is eliminated.

5.3 Welfare along the balanced-growth path (BGP)

We next examine the welfare implications of equating the private and social returns to human capital. First note that resolving the model without the externality leads to increases in education time, human capital growth, and the level of per capital human capital. Recall that along the balanced-growth path, non-stationary per capita variables, $\overline{X}_t$, are defined as $\overline{X}_t = x\overline{H}_t$, where x is the stationary steady-state value. Moreover, since $\overline{H}_t = (1+\gamma)\overline{H}_0$; $\overline{X}_t$ grows according to $\overline{X}_t = x(1+\gamma)\overline{H}_0$. Thus, *ceteris paribus*, higher γ in the efficient model (without the externality) leads to higher $\overline{H}_t$ and lower stationary steady-state values for x relative to the efficient model.

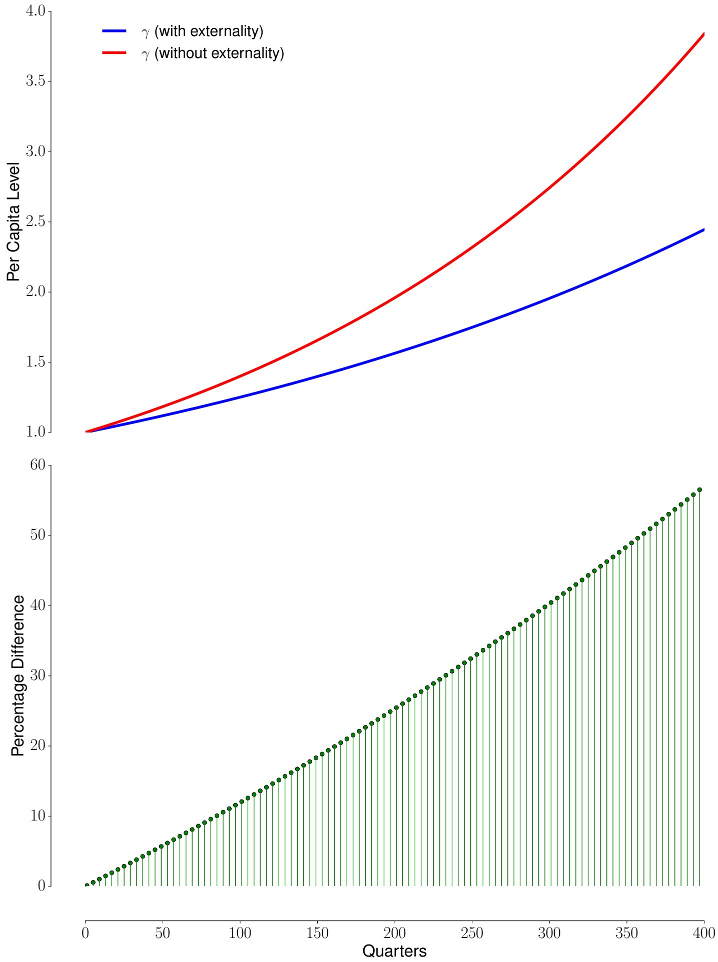
We show in Appendix D that lifetime welfare, $\overline{V}$, along the balanced-growth path depends on the starting value for the per capita human capital stock, $\overline{H}_0$, the stationary steady-state per human capital value of consumption, $c = \frac{\overline{C}_t}{\overline{H}_t}$, and the steady-state growth rate γ. Given that c falls when we move from the inefficient to the efficient model, to compare welfare effects of increased growth on consumption and hence welfare across models, $\overline{H}_0$ needs to be normalised so that both models start on the balanced-growth path from the same level of non-stationary consumption, $c\overline{H}_0$. For $\overline{H}_0 = 1$ in the inefficient model, this can be achieved by setting $\overline{H}_0 = \frac{c}{c^*}$ in the efficient model. Thus, using the steady-state lifetime welfare function, we will compare welfare across the inefficient and efficient balanced-growth paths given the same starting value for non-stationary consumption.

To quantify welfare gains/losses, we follow Lucas (1990) and Schmitt-Grohé and Uribe (2004) and compute the percentage extra private consumption that an individual would require so as to be equally well off between the inefficient and efficient regimes. The compensating consumption supplement is thus defined as follows:

$$\xi \simeq \left|\frac{1}{(1-\sigma)}\right| \ln\left(\frac{V_t^N}{V_t^B}\right) \times 100 \tag{30}$$

where V_t^N denotes lifetime utility when the externality is omitted. In turn V_t^B denotes the welfare associated with the estimated inefficient model.[17]

Figure 4: Implications for Per Capita Levels



The results of the welfare analysis are reported in Table 4.[18] These are

[17]See Appendix D for a derivation of lifetime welfare along the BGP.
[18]See eq. (51) in Appendix D

based on the means of the posterior distribution of parameters reported in Table 2. The results suggest that eliminating the externality leads to aggregate welfare gains in terms of the compensating consumption supplement of 8.68%. Table 4 also presents the percentage difference between the shocked values of education time and human capital growth from those in the base inefficient model. For example, when the externality is removed, e increases by 13.95% and annual γ by 0.45%. Whilst the latter appears to be a small change, the cumulative implications for the model's per capital levels are non-trivial. For example, Figure 4 above shows that these quantities would double roughly 25 years earlier in the model which internalises the social returns.

Table 4: Eliminating the Externality

	percent change
Welfare (ξ)	8.68
Education time (e)	13.95
Growth (γ)	0.45

6 External cross-validation

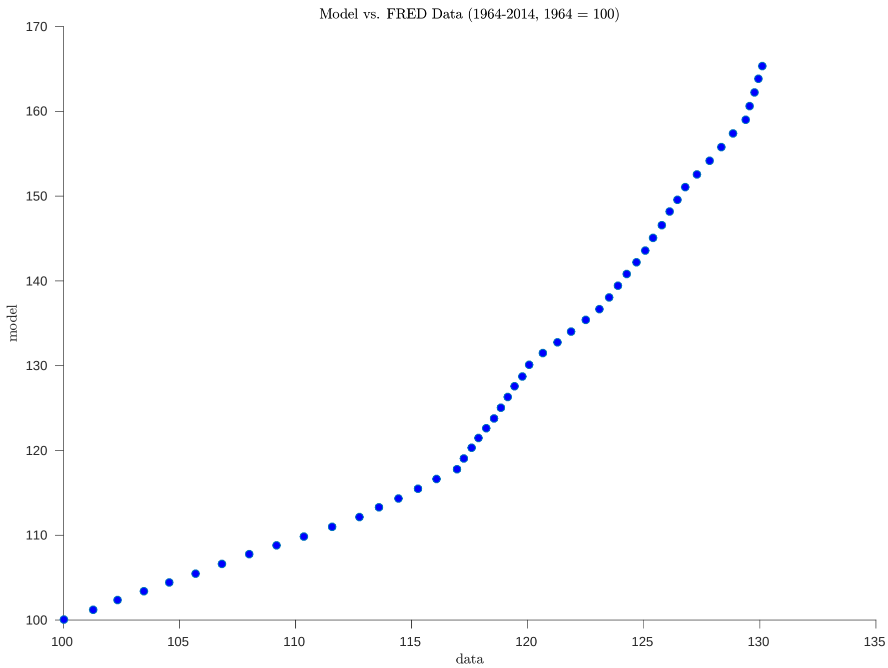
As pointed out in the introduction, since the highly stylised growth model we estimate may be misspecified in several dimensions and the measured data employed might not necessarily be informative, we conduct several cross-validation exercises to examine whether key model predictions cohere with the data and some stylised facts more broadly.

6.1 Human capital: model versus data

Given that human capital is not available at a quarterly frequency, it is treated as an unobservable in the estimation. Nonetheless, we can imply a model prediction for H_t by using eqs. (19) and (21) to generate $\hat{\gamma}_t$ and in turn imply H_t for each parameter draw. In eq. (21), unobservable $\hat{b}_t$ is obtained from the updating step of the Kalman filter (see eq. 40 in Appendix A). Finally, to obtain an annual index H_t from eq. (19), we annualize the growth rate, γ_t.

As discussed in Section 2.6, annual data relating to human capital is available from the FRED database (see section 2.6 for details). To check whether the annualised H_t implied by the model coheres with this data, we present a scatter plot in Figure 5 above which suggests a strong positive correlation. For example, the correlation coefficients for Q1, the median and Q3 are 0.9591, 0.9594 and 0.963. Thus, it appears that the model implied trend for human capital fits well with actual data not used to estimate the model.

Figure 5: Scatter Plot of Human Capital Per Capita (Model vs Data)



6.2 Cyclical behaviour of education-time

There is an abundance of evidence in the literature suggesting that college enrollment or education/training time more generally increases during recessions.[19] In recent research, Brown and Hoxby (2015) point out the college enrollment in the U.S. has increased in every recession since the 1960s. This counter-cyclical movement of education/training time is consistent with the predictions of economic theory since during downturns the opportunity cost of education, in form of forgone wages, as well as the opportunity cost of on-the-job training, in the form of foregone production, are lower.

[19]See, e.g. Brown and Hoxby (2015) and the papers within this volume as well as the papers by Dellas and Sakellaris (2003), Dellas and Koubi (2003), DeJong and Ingram (2001) and Sakellaris and Spilimbergo (2000).

In light of these findings, we next examine the quantitative predictions of our model with respect to the cyclical behaviour of education time. For each of the parameter draws, we construct the parameter matrices $\mathbf{Z}^{\mathrm{C}}, \mathbf{T}^{\mathrm{C}}, \mathbf{R}^{\mathrm{C}}$ and the variance covariance matrix of the two structural shocks $\mathbf{Q}^{\mathrm{C}}$ in eq. (23) we can calculate the cross correlations implied by the model. We start by multiplying the transition equation $\boldsymbol{\alpha}^{\mathrm{C}}_{t+1} = \mathbf{T}^{\mathrm{C}}\boldsymbol{\alpha}_t + \mathbf{R}^{\mathrm{C}}\boldsymbol{\eta}_{t+1}$ from the right with ${\boldsymbol{\alpha}^{\mathrm{C}}_{t+1}}'$:

$$\boldsymbol{\alpha}^{\mathrm{C}}_{t+1}{\boldsymbol{\alpha}^{\mathrm{C}}_{t+1}}' = \mathbf{T}^{\mathrm{C}}\boldsymbol{\alpha}_t{\boldsymbol{\alpha}^{\mathrm{C}}_{t+1}}' + \mathbf{R}^{\mathrm{C}}\boldsymbol{\eta}_{t+1}{\boldsymbol{\alpha}^{\mathrm{C}}_{t+1}}' \tag{31}$$

and take expectations to obtain the variance-covariance matrix $\boldsymbol{\Gamma}_{\boldsymbol{\alpha}}(0)$:

$$\boldsymbol{\Gamma}_{\boldsymbol{\alpha}}(0) = \mathbf{T}^{\mathrm{C}}\boldsymbol{\Gamma}_{\boldsymbol{\alpha}}(-1) + \mathbf{R}^{\mathrm{C}}\mathbf{Q}\mathbf{R}^{\mathrm{C}\prime}. \tag{32}$$

For lag 1,

$$\boldsymbol{\Gamma}_{\boldsymbol{\alpha}}(1) = \mathbf{T}^{\mathrm{C}}\boldsymbol{\Gamma}_{\boldsymbol{\alpha}}(0). \tag{33}$$

With $\boldsymbol{\Gamma}_{\boldsymbol{\alpha}}(1) = \boldsymbol{\Gamma}_{\boldsymbol{\alpha}}(-1)'$ we obtain:

$$\begin{aligned} \boldsymbol{\Gamma}_{\boldsymbol{\alpha}}(0) =& \mathbf{T}^{\mathrm{C}}\boldsymbol{\Gamma}_{\boldsymbol{\alpha}}(0)\mathbf{T}^{\mathrm{C}\prime} + \mathbf{R}^{\mathrm{C}}\mathbf{Q}\mathbf{R}^{\mathrm{C}\prime}; \\ \operatorname{vec}\left(\boldsymbol{\Gamma}_{\boldsymbol{\alpha}}(0)\right) =& \left(\mathbf{I}_{n^2} - \mathbf{T}^{\mathrm{C}} \otimes \mathbf{T}^{\mathrm{C}}\right)^{-1} \operatorname{vec}\left(\mathbf{R}^{\mathrm{C}}\mathbf{Q}\mathbf{R}^{\mathrm{C}\prime}\right). \end{aligned} \tag{34}$$

Once we have $\boldsymbol{\Gamma}_{\boldsymbol{\alpha}}(0)$, the covariance matrices for $\tau > 0$ can be calculated as:

$$\boldsymbol{\Gamma}_{\boldsymbol{\alpha}}(\tau) = \mathbf{T}^{\mathrm{C}}\boldsymbol{\Gamma}_{\boldsymbol{\alpha}}(\tau - 1), \tau > 0. \tag{35}$$

Using $\hat{e}_t = -\left(\frac{u}{1-u}\right)\hat{u}_t$, we can modify $\mathbf{Z}^{\mathrm{C}}$ so that the lhs of the measurement equations is the vector $\begin{pmatrix}\hat{y}_t & \hat{c}_t & \hat{u}_t & \hat{e}_t\end{pmatrix}$:

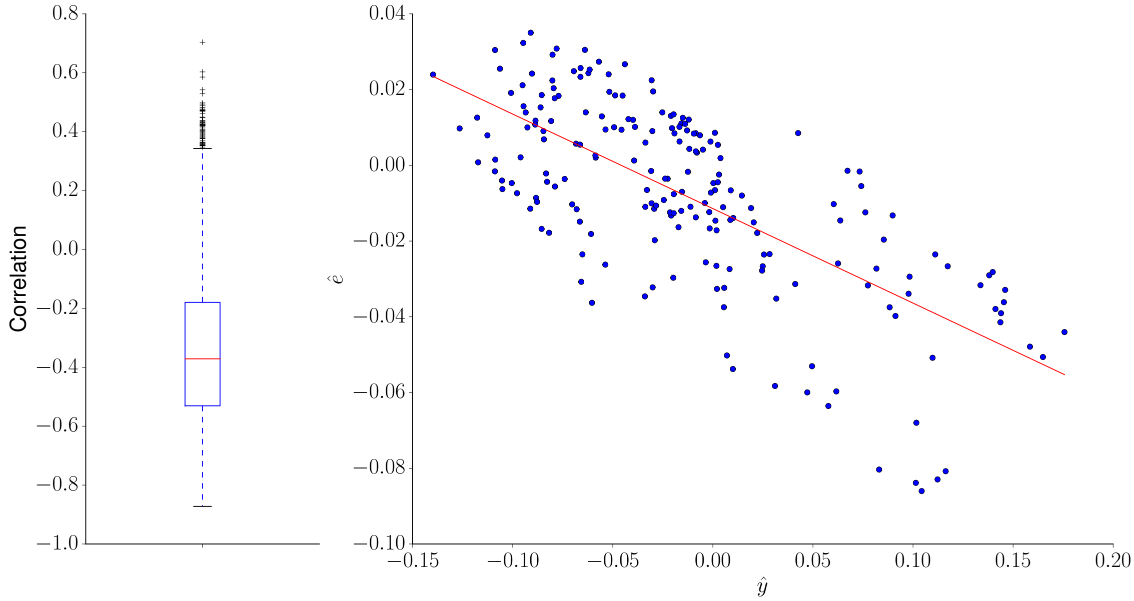
$$\mathbf{Z}^{\mathrm{C,NEW}} = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \\ 0 & 0 & -\frac{u}{1-u} \end{pmatrix} \mathbf{Z}^{\mathrm{C}}.$$

Thus, the covariance matrices for the measurement equations are computed as:

$$\boldsymbol{\Gamma}_{\mathbf{y}}(\tau) = \mathbf{Z}^{\mathrm{C,NEW}}\boldsymbol{\Gamma}_{\boldsymbol{\alpha}}(\tau)\mathbf{Z}^{\mathrm{C,NEW}\prime}, \tau = 0, 1, \ldots. \tag{36}$$

We are interested in the contemporaneous correlation between $\hat{y}_t$ and $\hat{e}_t$, i.e. the last row and first column of $\boldsymbol{\Gamma}_{\mathbf{y}}(0)$, divided by the standard deviations of $\hat{y}_t$ and $\hat{e}_t$. Figure 6 shows the result of this exercise. Consistent with the findings in the literature, education-time cycles are counter-cyclical in our estimated model.

Figure 6: Correlation of $\hat{e}_t$ with $\hat{y}_t$



6.3 Interpolated human capital data

As a further attempt to evaluate the implications and robustness of the estimated model in which human capital was treated as an unobservable, we next interpolate the annual human capital data used above to a quarterly frequency and re-estimate the model. To this end, output and consumption are now defined in per human capital units and hours in per capita units (as in the original estimation). This requires that we modify the measurement equation of the model in eq. (22) as follows:

$$\Delta \ln(\mathbf{Y}_t) - \begin{pmatrix} 1 \\ 1 \\ 0 \end{pmatrix} \ln(\gamma) = \mathbf{Z}^{\mathrm{C}} \boldsymbol{\alpha}_t^{\mathrm{C}} - \mathbf{Z}^{\mathrm{C}} \boldsymbol{\alpha}_{t-1}^{\mathrm{C}} + \boldsymbol{\nu}_t. \tag{37}$$

The results of this new estimation are reported in Appendix E in which we repeat Tables 2 and 3 as well as Figures 2 and 3 of the main text. We also repeat Figures 7 and 8 from Appendix C for the re-estimated model. To the extent that the interpolated data are a reasonable approximation of quarterly human capital, these results suggest that the original estimation is remarkably robust and thus, treating human capital as an unobservable has not biased the estimation in favour of the existence of an aggregate externality to human capital.

7 Conclusions

To more fully understand the quantitative implications of human capital externalites at the aggregate level, this paper econometrically estimated a two-sector model endogenous growth model with physical and human capital. We found that the pre-depreciation private returns to human capital are about 90% of the social returns for the U.S. over the period 1964-2017. To the best of our knowledge, the only other papers providing quantitative evidence supporting aggregate externalites in the U.S. are Choi (2011) and Guo *et al.* (2018) despite that fact that positive pecuniary and non-pecuniary externalities to education provide one of the main economic justifications for public spending on schooling at all levels.

We further find that if the social and private returns to education were equalised, discounted lifetime aggregate welfare, in terms of the compensating consumption supplement, would increase by about 8.7%. This implied that education-time would increase by approximately 14% and the annual growth rate of human capital would increase by about half a percent. We showed that the latter is non-trivial considering its cumulative effects on the per capita levels in the model. For example, we found that these quantities would double rougly 25 years earlier in the model which internalises the social returns.

We conducted several cross-validation exercises to examine whether key model predictions cohered with the data and some stylised facts more broadly. The results of these exercises suggested that (i) the model implied trend for human capital fits well with actual data not used to estimate the model; (ii) consistent with the findings in the literature, education-time cycles are counter-cyclical in the estimated model; and (iii) the original estimation is remarkably robust and thus that treating human capital as an unobservable has not biased the estimation in favour of the aggregate human capital externality.

Thus, our results provide a reasonably robust benchmark of the potential welfare gains associated with internalising the aggregate human capital externality. Of course, if distortionary tax and spending policy are used to publicly provide the inefficiently low investment in education implied by the externality, the welfare gains would be reduced. However, analysis of the cost implications on welfare will be left to future research.

References

Acemoglu, D. and Angrist, J. (2001), "How Large are Human-Capital Externalities? Evidence from Compulsory-Schooling Laws." In: *NBER Macroeconomics Annual 2000*, Vol. 15, 9–74, National Bureau of Economic Research.

An, S. and Schorfheide, F. (2007), "Bayesian Analysis of DSGE Models." *Econometric Reviews* **26**, 113–172.

Anzoategui, D., Comin, D., Gertler, M., and Martinez, J. (2017), "Endogenous Technology Adoption and R&D as Sources of Business Cycle Persistence." *NBER Working Paper Series (22005)* .

Bastable, C. F. (1892), *Public Finance*. London: Macmillan & Co.

Bastiat, F. (1851), *Les Harmonies Économiques*. Second ed., Paris: Gillaumin.

Becker, G. S. (1964), *Human Capital*. University of Chicago Press.

Bianchi, F., Kung, H., and G., M. (2018), "Growth, Slowdowns, and Recoveries." *Journal of Monetary Economics* (in press).

Brown, J. and Hoxby, C. (2015), *How the Financial Crisis and Great Recession Affected Higher Education*. NBER Conference Report.

Buchanan, J. M. and Tullock, G. (1962), *The Calculus of Consent: Logical Foundations of Constitutional Democracy*. Ann Arbor, Michigan: University of Michigan Press.

Canova, F. (2007), *Methods for Applied Macroeconomic Research*. Princeton University Press.

Canova, F. (2014), "Bridging DSGE Models and the Raw Data." *Journal of Monetary Economics* **67**, 1–15.

Canova, F. and Ferroni, F. (2011), "Multiple Fltering Devices for the Estimation of Cyclical DSGE Models." *Quantitative Economics* **2**, 73–98.

Černý, V. (1985), "Thermodynamical Approach to the Traveling Salesman Problem: An Efficient Simulation Algorithm." *Journal of Optimization Theory and Applications* **45**, 41–51.

Chang, Y., Doh, T., and Schorfheide, F. (2007), "Non-stationary Hours in a DSGE Model." *Journal of Money, Credit and Banking* **39**, 1357–1373.

Chib, S. and Greenberg, E. (1994), "Bayes Inference in Regression Models with ARMA(p,q) Errors." *Journal of Econometrics* **64**, 183–206.

Chib, S. and Greenberg, E. (1995), "Understanding the Metropolis-Hastings Algorithm." *American Statistician* **49**, 327–335.

Chib, S. and Ramamurthy, S. (2010), "Tailored Randomized-Block MCMC Methods for Analysis of DSGE Models." *Journal of Econometrics* **155**, 19–38, forthcoming.

Choi, S. M. (2011), "How large are learning externalities." *International Economic Review* **52**, 1077–1103.

Ciccone, A. and Peri, G. (2006), "Identifying Human-Capital Externalities: Theory with Applications." *Review of Economic Studies* **73**, 381–412.

Davies, J. (2003), "Empirical Evidence on Human Capital Externalities.", University of Western Ontario Working Paper No. 2003-5.

DeJong, D. and Dave, C. (2011), *Structural Macroeconometrics*. Princeton: Princeton University Press.

DeJong, D. and Ingram, B. (2001), "The Cyclical Behavior of Skill Acquisition." *Review of Economic Dynamics* **4**, 536–561.

Del Negro, M. and Schorfheide, F. (2007), "Bayesian Macroeconometrics." In: J. Geweke, G. Koop, and H. van Dijk (Eds.), *Handbook of Bayesian Econometrics*, Oxford.

Dellas, H. and Koubi, V. (2003), "Business Cycles and Schooling." *European Journal of Political Economy* **19**, 843–859.

Dellas, H. and Sakellaris, P. (2003), "On the Cyclicality of Schooling: Theory and Evidence." *Oxford Economic Papers* **55**, 148–172.

Feenstra, R. C., Inklaar, R., and Timmer, M. P. (2015), "The Next Generation of the Penn World Table." *American Economic Review* **105**, 3150–3182.

Fernández-Villaverde, J., Rubio-Ramírez, J., and Schorfheide, F. (2016), "Solution and Estimation Methods for DSGE Models." In: J. B. Taylor and H. Uhlig (Eds.), *Handbook of Macroeconomics*, Vol. 2A, first ed., chap. 9, 530–724, Elsevier.

Ferroni, F. (2011), "Trend Agnostic One-Step Estimation of DSGE Models." *The B.E. Journal of Macroeconomics* **11**, Article 25.

Geweke, J. (1992), "Evaluating the Accuracy of Sampling-Based Aproaches to the Calculation of Posterior Moments." In: J. Bernardo, J. Berger, A. P. Dawid, and A. F. M. Smith (Eds.), *Bayesian Statistics*, 641–649, Oxford: Oxford University Press.

Gill, P. E. and Murray, W. (1974), "Newton-Type Methods for Unconstrained and Linearly Constrained Optimization." *Mathematical Programming* **7**, 311–350.

Guo, J., Roy, N., and Seshadri, A. (2018), "Estimating Aggregate Human Capital Externalities.", University of Wisconsin-Madison.

Hamilton, J. D. (1994), *Time Series Analysis*. Princeton, New Jersey: Princeton University Press.

Harvey, A. C. (1992), *Forecasting, Structural Time Series Models and the Kalman Filter*. Cambridge: Cambridge University Press.

Herbst, E. and Schorfheide, F. (2016), *Bayesian Estimation of DSGE Models*. Princeton: Princeton University Press.

Ireland, P. (2004), "A Method of Taking the Model to the Data." *Journal of Economic Dynamics & Control* **28**, 1205–1226.

Jones, L., Manuelli, R., and Siu, H. (2005), "Fluctuations in Convex Models of Endogenous Growth, II: Business Cycle Properties." *Review of Economic Dynamics* **8**, 805–828.

Jorgenson, D. and Fraumeni, B. M. (1989), "The Accumulation of Human and Nonhuman Capital, 1948-1984." In: R. E. Lipsey and H. S. Tice (Eds.), *The Measurement of Saving, Investment and Wealth*, Vol. 52 of *Studies in Income and Wealth*, Chicago: University of Chicago Press.

Kim, D. and Lee, C. (2007), "On-the-Job Human Capital Accumulation in a Real Business Cycle Model: Implications for Intertemporal Substitution Elasticity and Labor Hoarding." *Review of Economic Dynamics* **10**, 494–518.

Kirkpatrick, S., Gelatt, C. D., and Vecchi, M. P. (1983), "Optimization by Simulated Annealing." *Science* **220**, 671–680.

Klein, P. (2000), "Using the Generalized Schur Form to Solve a Multivariate Linear Rational Expectations Model." *Journal of Economic Dynamics and Control* **24**, 1405–1423.

Lange, F. and Topel, R. (2006), "The Social Value of Education and Human Capital." In: E. Hanushek and F. Welch (Eds.), *The Handbook of Economics of Education*, Vol. 1, 459–509, Elsevier.

Lucas, R. E. (1988), "On the Mechanics of Economic Development." *Journal of Monetary Economics* **22**, 3–42.

Lucas, R. E. (1990), "Supply-Side Economics: An Analytical Review." *Oxford Economic Papers* **42**, 293–316.

Malley, J. and Woitek, U. (2010), "Technology Shocks and Aggegate Fluctuations in an Estimated Hybrid RBC Model." *Journal of Economic Dynamics and Control* **34**, 1214–1232.

Marshall, A. (1890), *Principles of Economics*. London: Macmillan & Co.

Metropolis, N., Rosenbluth, A. W., Rosenbluth, M. N., Teller, A. H., and Teller, E. (1953), "Equations of State Calculations by Fast Computing Machines." *Journal of Chemical Physics* **21**, 1087–1092.

Moretti, E. (2004a), "Estimating the Social Return to Higher Education: Evidence from Longitudinal and Repeated Cross-Sectional Data." *Journal of Econometrics* **121**, 175–212.

Moretti, E. (2004b), "Human Capital Externalities In Cites." In: J. Henderson and J. Thisse (Eds.), *Handbook of Regional and Urban Economics*, 2243–2291, Elsevier.

Moretti, E. (2004c), "Workers' Education, Spillovers, and Productivity: Evidence from Plant-Level Production Functions." *American Economic Review* **94**, 656–690.

Perli, R. and Sakellaris, P. (1998), "Human Capital Formation and Business Cycle Persistence." *Journal of Monetary Economics* **42**, 67–92.

Press, W. H., Teukolsky, S. A., Vetterling, W. T., and Flannery, B. P. (1992), *Numerical Recipes in C. The Art of Scientific Computing*. Second ed., Cambridge: Cambridge University Press.

Psacharopoulos, G. and Patrinos, H. A. (2004), "Returns to Investment in Education: a Further Update." *Education Economics* **12**, 111–134.

Rathke, A., Straumann, T., and Woitek, U. (2017), "Overvalued: Swedish Monetary Policy in the 1930s." *International Economic Review,* **58**, forthcoming.

Romer, P. M. (1986), "Increasing Returns and Long-Run Growth." *Journal of Political Economy* **94**, 1002–1037.

Sakellaris, P. and Spilimbergo, A. (2000), "Business Cycles and Investment in Human Capital: International Evidence on Higher Education." *Carnegie-Rochester Conference Series on Public Policy* **52**, 221–256.

Schmitt-Grohé, S. and Uribe, M. (2004), "Solving Dynamic General Equilibrium Models Using a second-order approximation to the policy function." *Journal of Economic Dynamics and Control* **28**, 755–775.

Smets, F. and Wouters, R. (2007), "Shocks and Frictions in US Business Cycles: A Bayesian DSGE Approach." *American Economic Review* **97**, 586–606.

Smith, A. (1776), *An Inquiry into the Nature and Causes of the Wealth of Nations*. London: for W. Strahan and T. Cadell.

Tamura, R. (1991), "Income Convergence in an Endogenous Growth Model." *Journal of Political Economy* **99**, 522–540.

van Laarhoven, P. J. M. and Aarts, E. H. L. (1987), *Simulated Annealing: Theory and Applications*. Amsterdam: Kluwer Academic Publishers.

Appendix A First-order approximation

To solve the model, we take the first-order Taylor series expansion of the non-linear stationary DCE in eq. (15) and the exogenous processes in eq. (16) around the steady-state. After substituting out the log deviations of the Lagrange multipliers and the growth of human capital, $\hat{\gamma}_t = \left(\frac{B\theta e^{\theta}}{n\gamma}\right)\hat{e}_t + \left(\frac{Be^{\theta}}{n\gamma}\right)\hat{b}_t$, the linearised system which we solve to obtain the state-space form in eq. (17) is:

$$\begin{aligned}
-\hat{y}_t - \omega_4\hat{e}_t + \omega_5\hat{k}_t + \hat{a}_t &= 0 \\
\omega_2\hat{c}_t - \hat{y}_t + \omega_1(\omega_6\hat{e}_t + \omega_7\hat{b}_t) - \omega_3\hat{k}_t &= -\omega_1\hat{k}_{t+1} \\
-\omega_8\hat{c}_t + \omega_8(\omega_6\hat{e}_t + \omega_7\hat{b}_t) &= \omega_9 E_t\hat{y}_{t+1} - \omega_8 E_t\hat{c}_{t+1} - \omega_9 E\hat{k}_{t+1} \\
-\omega_8\hat{c}_t + \hat{y}_t + \omega_{23}\hat{e}_t + \omega_{24}\hat{b}_t &= -\omega_{19}E_t\hat{c}_{t+1} + \omega_{20}E_t\hat{y}_{t+1} \\
&\quad -\omega_{21}E_t\hat{b}_{t+1} + \omega_{22}E_t\hat{e}_{t+1} \qquad (38) \\
\rho^a\hat{a}_t &= \hat{a}_{t+1} - \varepsilon^a_{t+1} \\
\rho^b\hat{b}_t &= \hat{b}_{t+1} - \varepsilon^b_{t+1}
\end{aligned}$$

where for any variable x_t, $\hat{x}_t = ln(x_t/x)$; x is the model-consistent steady-state value of x_t; $\hat{y}_t$ is the control variable; $\hat{c}_t$ and $\hat{u}_t$ are jump variables; $\hat{k}_t$ is the state variable; $\hat{a}_t$ and $\hat{b}_t$ are the two exogenous processes; and the ω_i coefficients are defined in the following table.

Table 5: Parameter convolutions

$\omega_1 \equiv \frac{n\gamma k}{y}$	$\omega_9 \equiv \frac{\alpha y}{\beta^{-1}\gamma^{\sigma}k}$	$\omega_{17} \equiv \omega_8$
$\omega_2 \equiv \frac{c}{y}$	$\omega_{10} \equiv \frac{e}{1-e} - \theta + 1$	$\omega_{18} \equiv \frac{\omega_{11}}{\omega_{11}+\omega_{12}}$
$\omega_3 \equiv \frac{\left(1-\delta^k\right)k}{y}$	$\omega_{11} \equiv c^{-\sigma}(1-\alpha)y$	$\omega_{19} = \omega_8\omega_{13} + \omega_{16}$
$\omega_4 \equiv \frac{(1-\alpha)e}{1-e}$	$\omega_{12} = \lambda^b(1-\delta^h) + B\theta e^{\theta}$	$\omega_{20} = \omega_{13} + \omega_{18}$
$\omega_5 \equiv \alpha$	$\omega_{13} \equiv \frac{\lambda^b\left[1-\delta^h+B\theta e^{\theta}\right]}{\omega_{11}+\omega_{12}}$	$\omega_{21} = \omega_{13} - \omega_{14}$
$\omega_6 \equiv \frac{B\theta e^{\theta}}{n\gamma}$	$\omega_{14} \equiv \frac{B\lambda^b\theta e^{\theta}}{\omega_{11}+\omega_{12}}$	$\omega_{22} = \omega_{10}\omega_{13} + \omega_{15}$
$\omega_7 \equiv \frac{Be^{\theta}}{n\gamma}$	$\omega_{15} \equiv \frac{B\lambda^b\theta^2 e^{\theta}}{\omega_{11}+\omega_{12}}$	$\omega_{23} = \omega_{10} + \omega_6\omega_{17}$
$\omega_8 \equiv \sigma$	$\omega_{16} \equiv \frac{\sigma\omega_{11}}{\omega_{11}+\omega_{12}}$	$\omega_{24} = \omega_7\omega_{17} - 1$

Appendix B Estimation procedure

B.1 Kalman filter

We use the Kalman filter to calculate the likelihood $p(\hat{\mathbf{y}}_1, \ldots, \hat{\mathbf{y}}_T|\boldsymbol{\psi})$, which combined with the prior $p(\boldsymbol{\psi})$, gives the part of the posterior distribution $p(\boldsymbol{\psi}|\hat{\mathbf{y}}_1, \ldots, \hat{\mathbf{y}}_T)$ relevant for the Metropolis-Hastings steps in the procedure outlined in Section B.2. For given initial estimates of the state vector, $\mathbf{a}_0$, i.e. $\mathbf{a}_0 = E(\boldsymbol{\alpha}_0)$ and the covariance matrix, $\mathbf{P}_0$, the filter consists of the following steps:

Prediction step

$$\begin{aligned} \mathbf{a}_{t|t-1} &= \mathbf{T}\mathbf{a}_{t-1}; \\ \mathbf{P}_{t|t-1} &= \mathbf{T}\mathbf{P}_{t-1}\mathbf{T}' + \mathbf{R}\mathbf{Q}\mathbf{R}'. \end{aligned} \tag{39}$$

In this step, a prediction $\mathbf{a}_{t|t-1}$ of the state vector and its variance-covariance matrix $P_{t|t-1}$ is generated, based on information available at period $t-1$.

Updating step

$$\begin{aligned} \boldsymbol{v}_t &= \mathbf{y}_t - \mathbf{Z}\mathbf{a}_{t|t-1}; \\ \mathbf{F}_t &= \mathbf{Z}\mathbf{P}_{t|t-1}\mathbf{Z}'; \\ \mathbf{K}_t &= \mathbf{T}\mathbf{P}_{t|t-1}\mathbf{Z}'\mathbf{F}_t^{-1}; \\ \mathbf{a}_t &= \mathbf{T}\mathbf{a}_{t|t-1} + \mathbf{K}_t\boldsymbol{v}_t; \\ \mathbf{P}_t &= (\mathbf{T} - \mathbf{K}_t\mathbf{Z})\,\mathbf{P}_{t|t-1}\,(\mathbf{T} - \mathbf{K}_t\mathbf{Z})' + \mathbf{R}\mathbf{Q}\mathbf{R}', \end{aligned} \tag{40}$$

where $\boldsymbol{v}_t$ are the model's forecast errors. The new information from the errors at time t is used to generate the updates $\mathbf{a}_t$ and $\mathbf{P}_t$. The remaining vector and matrices have either been defined above or, in the case of $\mathbf{F}_t$, the variance-covariance matrix of $\boldsymbol{v}_t$, and the Kalman gain $\mathbf{K}_t$, are simply transformations of previously defined matrices.[20] In our case, we divide the vector of unknowns into two blocks: the state variables $\boldsymbol{\alpha}_t, t = 1, \ldots, T$ and the vector with hyper-parameters $\boldsymbol{\psi}$. Thus, the model's likelihood function is:

$$p(\hat{\mathbf{y}}_1, \ldots, \hat{\mathbf{y}}_T|\boldsymbol{\psi}) = \prod_{t=1}^{\mathrm{T}} (2\pi)^{-0.5n} |\mathbf{F}_t|^{-0.5} \exp\left(-0.5\boldsymbol{v}_t'\mathbf{F}_t^{-1}\boldsymbol{v}_t\right). \tag{41}$$

[20]See Hamilton (1994) or Harvey (1992) for further details regarding the Kalman filter.

B.2 Chib and Ramamurthy (2010)

Using the stationary data $\hat{\mathbf{y}}_t, \ldots, \hat{\mathbf{y}}_T$, we estimate the vector of model hyper-parameters $\boldsymbol{\psi}$ using the tailored multiple-block Metropolis-Hastings (MH) algorithm proposed by Chib and Ramamurthy (2010), see also (Chib and Greenberg, 1994, 1995). This method separates parameters into different groups and updates them block-wise in an MH step, conditional on the remaining groups. Usually, parameter blocks are generated by searching for groups of correlated parameters, but this is difficult in a DSGE framework, since the parameters of the linear state-space representation are non-linear combinations of the underlying parameters in $\boldsymbol{\psi}$. Instead, Chib and Ramamurthy (2010) randomize the formation of the parameter blocks since it helps to avoid poor *a priori* choices. This framework also allows parameter groupings to change, which is preferable if there are irregularities such as changes in the shapes of the posterior parameter distributions.

To generate these blocks, we permute the index of the parameters randomly. The first parameter initializes the first block. As in Chib and Ramamurthy (2010), the next parameter is included into this block with probability $\tau = 0.8$, and starts a new block with probability $1 - \tau$. Note that in simulation step k, the above algorithm generates p_k blocks $\boldsymbol{\psi}_{k,1}, \ldots, \boldsymbol{\psi}_{k,p_k}$. To find the maximum of the posterior with respect to block j, we keep all the other blocks constant and calculate:

$$\boldsymbol{\psi}^{\star}_{k,j} = \operatorname{argmax}\big(f(\hat{\mathbf{y}}_1, \ldots, \hat{\mathbf{y}}_T | \boldsymbol{\psi}_{k,1}, \ldots, \boldsymbol{\psi}_{k,j}, \ldots, \boldsymbol{\psi}_{k,p_k})\pi(\boldsymbol{\psi})\big), \tag{42}$$

where $\pi(\boldsymbol{\psi})$ is the prior parameter distribution given in Table 1 below.

We use simulated annealing to calculate (42).[21] The negative inverse Hessian $\mathbf{V}_{k,j}$ of the target posterior distribution is calculated at $\boldsymbol{\psi}^{\star}_{k,j}$. If it is not positive definite, a modified Cholesky decomposition (Gill and Murray, 1974) is applied to the negative Hessian to find the matrix $\mathbf{P}$, and $\mathbf{V}_{k,j} = (\mathbf{P}\mathbf{P}')^{-1}$. As in Chib and Ramamurthy (2010), the proposal density is a multivariate t-distribution with $\nu > 2$ degrees of freedom. Drawing from this distribution, a candidate $\boldsymbol{\psi}^{1}_{k,j}$ is generated, and accepted if $a \geq u, u \sim$

[21] The algorithm is a generalization of the Metropolis algorithm (Metropolis *et al.*, 1953) developed by Kirkpatrick *et al.* (1983) and Černý (1985). For an overview, see e.g. van Laarhoven and Aarts (1987) or Press *et al.* (1992, Section 10.9). Following Rathke *et al.* (2017), we use the parameterisation found in Chib and Ramamurthy (2010) and set the parameters for the algorithm as follows: scaling factor for parameter proposal: 0.02; cooling constant: 0.4; stage expansion factor: 8; initial temperature: 5; number of stages: 8; initial stage length: 4.

$U(0,1)$, where

$$
\begin{aligned}
a(\boldsymbol{\psi}_{k,j}^{1}, \boldsymbol{\psi}_{k,j}^{0}) = & \\
= \min & \left(\frac{p(\hat{\mathbf{y}}_1, \ldots, \hat{\mathbf{y}}_T | \boldsymbol{\psi}_{k,1}^{0}, \ldots, \boldsymbol{\psi}_{k,j}^{1}, \ldots \boldsymbol{\psi}_{k,p_k}^{0})}{p(\hat{\mathbf{y}}_1, \ldots, \hat{\mathbf{y}}_T | \boldsymbol{\psi}_{k,1}^{0}, \ldots, \boldsymbol{\psi}_{k,j}^{0}, \ldots, \boldsymbol{\psi}_{k,p_k}^{0})} \times \right. \\
& \left. \times \frac{g(\boldsymbol{\psi}_{k,j}^{1}) t(\boldsymbol{\psi}_{k,j}^{0} | \boldsymbol{\psi}_{k,j}^{\star}, \mathbf{V}_{k,j}, \nu)}{g(\boldsymbol{\psi}_{k,j}^{0}) t(\boldsymbol{\psi}_{k,j}^{1} | \boldsymbol{\psi}_{k,j}^{\star}, \mathbf{V}_{k,j}, \nu)}, 1 \right).
\end{aligned} \tag{43}
$$

Appendix C Convergence

Figure 7: Trace Plots of Structural Parameters

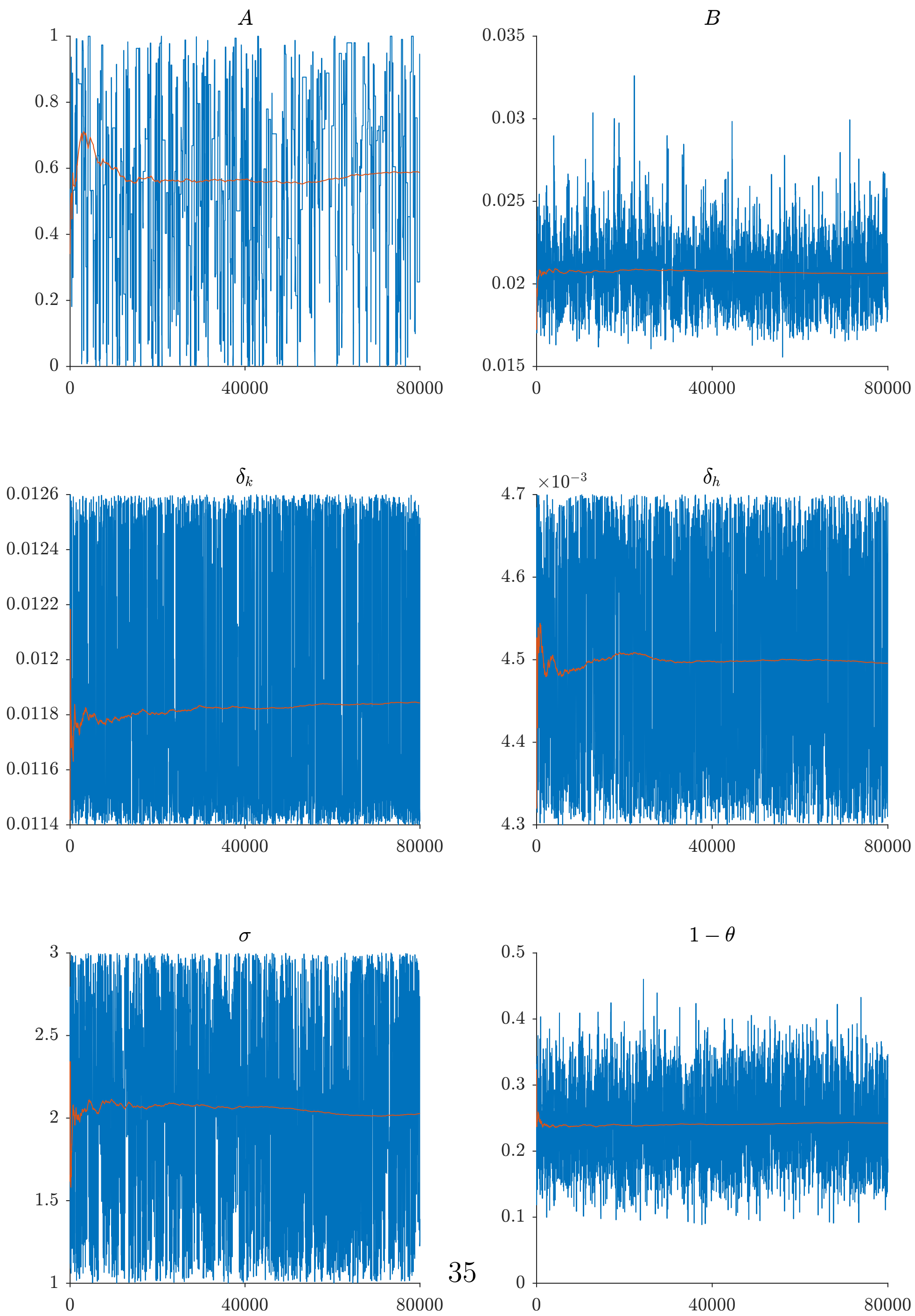
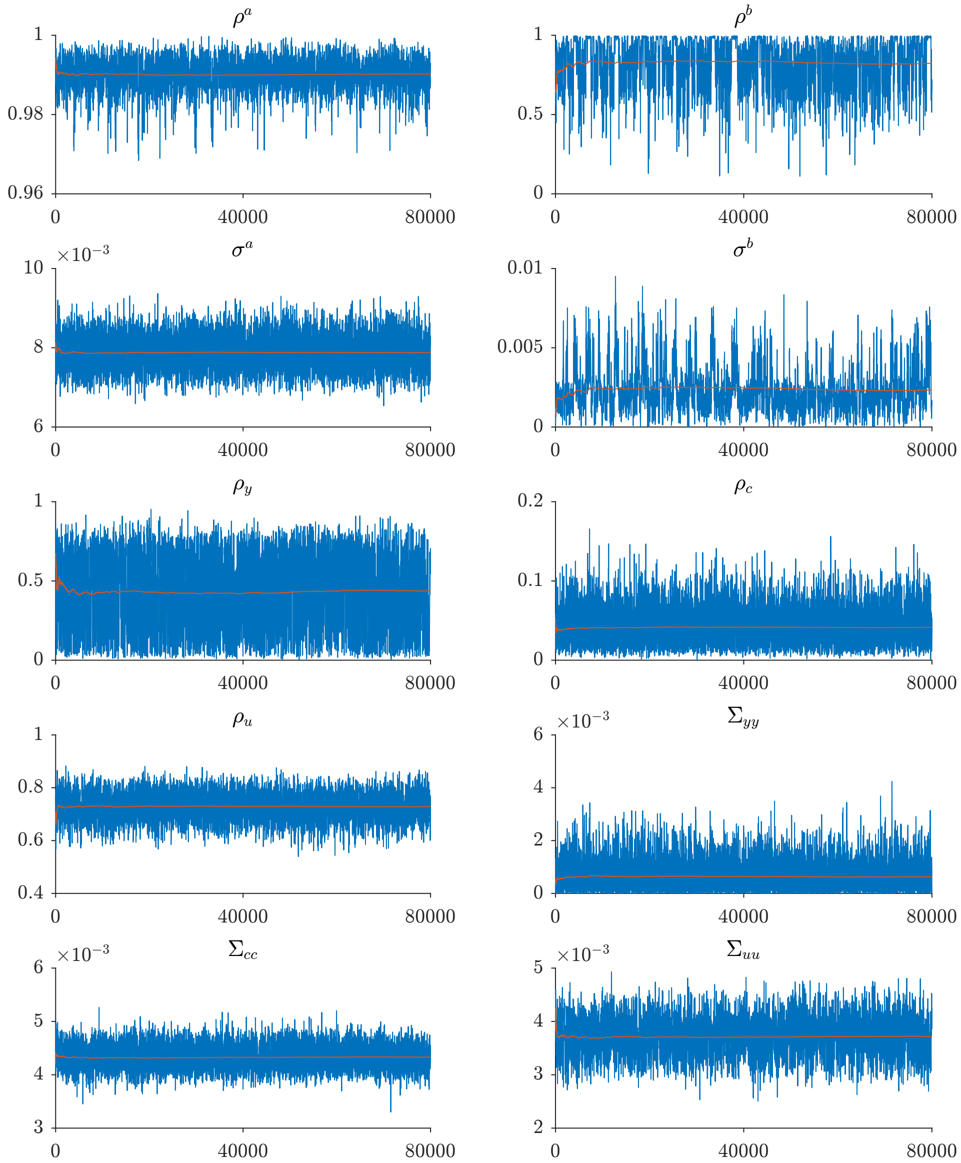


Figure 8: Trace Plots of AR Processes and Measurement/Specification Errors



Appendix D Lifetime welfare on BGP

The instantaneous utility function in eq. (2) can be expressed in per capita terms as:

$$\overline{U}_t = \frac{(\overline{C}_t)^{1-\sigma}}{1-\sigma}, \tag{44}$$

where $\overline{C}_t = \frac{C_t}{N_t}$. Using our notation for stationary variables we can write (44) equivalently as:

$$\overline{U}_t = \frac{(c_t \overline{H}_t)^{1-\sigma}}{1-\sigma}, \tag{45}$$

where $c_t \equiv \frac{\overline{C}_t}{\overline{H}_t}$ is stationary per capita consumption and $\overline{H}_t$ is the beginning-of-period per capita human capital stock. Since $\gamma_t \equiv \overline{H}_{t+1}/\overline{H}_t$, we have for $t \geq 1$:

$$\overline{H}_t = \overline{H}_0 \left(\prod_{s=0}^{t-1} \gamma_s \right), \tag{46}$$

where $\overline{H}_0$ is given from initial conditions.

Substituting (46) into (45) gives:

$$\overline{U}_t = \frac{\left[\overline{H}_0 c_t \left(\prod_{s=0}^{t-1} \gamma_s \right) \right]^{1-\sigma}}{1-\sigma} \text{for } t \geq 1. \tag{47}$$

Equation (47) can be rewritten equivalently as:

$$\overline{U}_t = \frac{\gamma^{t(1-\sigma)} \left(\overline{H}_0 c_t \widetilde{\gamma}_t \right)^{1-\sigma}}{1-\sigma} \text{ for } t \geq 1, \tag{48}$$

where $\widetilde{\gamma}_t = \left(\prod_{s=0}^{t-1} \gamma_s \right) / \gamma^t$ is a scaled measure of the cumulated growth rate. In the deterministic steady-state $\widetilde{\gamma} = 1$.

Expected discounted lifetime utility in eq. (1), can be written in recursive form for $t \geq 1$ as follows:

$$\begin{aligned} \overline{W}_t &= E_t \sum_{j=2}^{\infty} \beta^{j-1} \overline{U}_{t+j-1}; \\ &= \overline{U}_t + \beta E_t \sum_{j=1}^{\infty} \beta^{j-1} \overline{U}_{t+j}; \\ &= \overline{U}_t + \beta E_t \overline{W}_{t+1}. \end{aligned} \tag{49}$$

Substituting (48) into (49) and dividing by sides of the resulting expression by $\overline{\gamma}^{t(1-\sigma)}$ gives lifetime utility:

$$\overline{V}_t = \frac{\left(\overline{H}_0 c_t \widetilde{\gamma}_t\right)^{1-\sigma}}{1-\sigma} + \beta\overline{\gamma}^{(1-\sigma)} E_t \overline{V}_{t+1} \text{ for } t \geq 1 \tag{50}$$

where $\overline{V}_t = \frac{\overline{W}_t}{\overline{\gamma}^{t(1-\sigma)}}$ and $\overline{V}_{t+1} = \frac{\overline{W}_{t+1}}{\overline{\gamma}^{t(1-\sigma)}}$.

Finally, steady-state lifetime welfare along the deterministic balanced-growth path is given by:

$$\begin{aligned}\overline{V} &= \frac{\left(\overline{H}_0 c \widetilde{\gamma}\right)^{1-\sigma}}{1-\sigma} + \beta\gamma^{(1-\sigma)}\overline{V} \\ &= \frac{\left(\overline{H}_0 c\right)^{1-\sigma}}{(1-\sigma)\left(1-\beta\gamma^{(1-\sigma)}\right)}, \text{ since } \widetilde{\gamma} = 1.\end{aligned} \tag{51}$$

Appendix E Robustness

In this section, we report the results for the robustness check discussed in Section 6.

Table 6: Posterior distribution of parameters, $\boldsymbol{\psi}$

	Location and Spread			Quantiles		
	mean	s.d.	NSE	%25	%50	%75
A	0.63806	0.25342	0.44402	0.47740	0.66820	0.83802
B	0.01994	0.00115	0.06449	0.01913	0.01988	0.02067
δ_k	0.01182	0.00046	0.04368	0.01146	0.01151	0.01239
δ_h	0.00451	0.00014	0.03466	0.00437	0.00458	0.00464
σ	2.06955	0.57775	0.31210	1.55828	2.05951	2.60624
$1-\theta$	0.26689	0.04914	0.20586	0.23355	0.26631	0.29934
ρ^a	0.99272	0.00174	0.00196	0.99160	0.99281	0.99393
ρ^b	0.84201	0.13956	0.18530	0.74879	0.87822	0.96535
σ^a	0.00801	0.00037	0.05176	0.00775	0.00799	0.00825
σ^b	0.00255	0.00118	0.51845	0.00170	0.00232	0.00318
ρ_y	0.36397	0.21037	0.64615	0.18559	0.35307	0.53970
ρ_c	0.03725	0.01682	0.50479	0.02517	0.03483	0.04672
ρ_u	0.73817	0.04532	0.06863	0.70859	0.73997	0.76937
Σ_{yy}	0.00068	0.00058	0.95041	0.00025	0.00054	0.00096
Σ_{cc}	0.00423	0.00020	0.05337	0.00410	0.00423	0.00437
Σ_{uu}	0.00373	0.00033	0.09732	0.00351	0.00372	0.00395

Table 7: Quantiles of Prior and Posterior Distributions

	Prior			Posterior		
	25%	50%	75%	25%	50%	75%
A	0.25000	0.50000	0.75000	0.47740	0.66820	0.83802
B	0.25000	0.50000	0.75000	0.01913	0.01988	0.02067
δ_k	0.00600	0.01200	0.01800	0.01146	0.01151	0.01239
δ_h	0.00225	0.00450	0.00675	0.00437	0.00458	0.00464
σ	1.00000	2.00000	3.00000	1.55828	2.05951	2.60624
$1-\theta$	0.34934	0.50000	0.65066	0.23355	0.26631	0.29934
ρ^a	0.34934	0.50000	0.65066	0.99160	0.99281	0.99393
ρ^b	0.34934	0.50000	0.65066	0.74879	0.87822	0.96535
σ^a	0.05590	0.11180	0.16771	0.00775	0.00799	0.00825
σ^b	0.05590	0.11180	0.16771	0.00170	0.00232	0.00318
ρ_y	0.34934	0.50000	0.65066	0.18559	0.35307	0.53970
ρ_c	0.34934	0.50000	0.65066	0.02517	0.03483	0.04672
ρ_u	0.34934	0.50000	0.65066	0.70859	0.73997	0.76937
Σ_{yy}				0.00025	0.00054	0.00096
Σ_{cc}		positive		0.00410	0.00423	0.00437
Σ_{uu}				0.00351	0.00372	0.00395

Figure 9: Comparison of Prior and Posterior Distributions

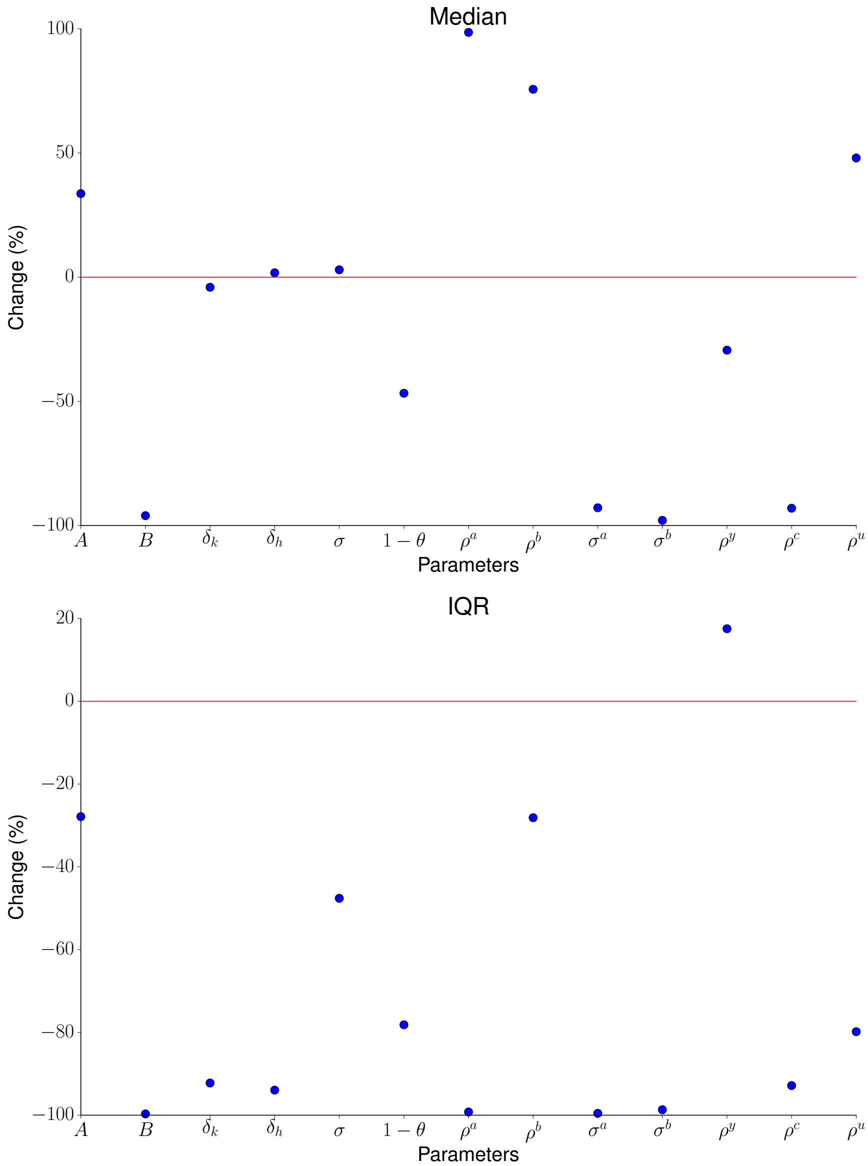


Figure 10: Posterior Plots, $1 - \theta$

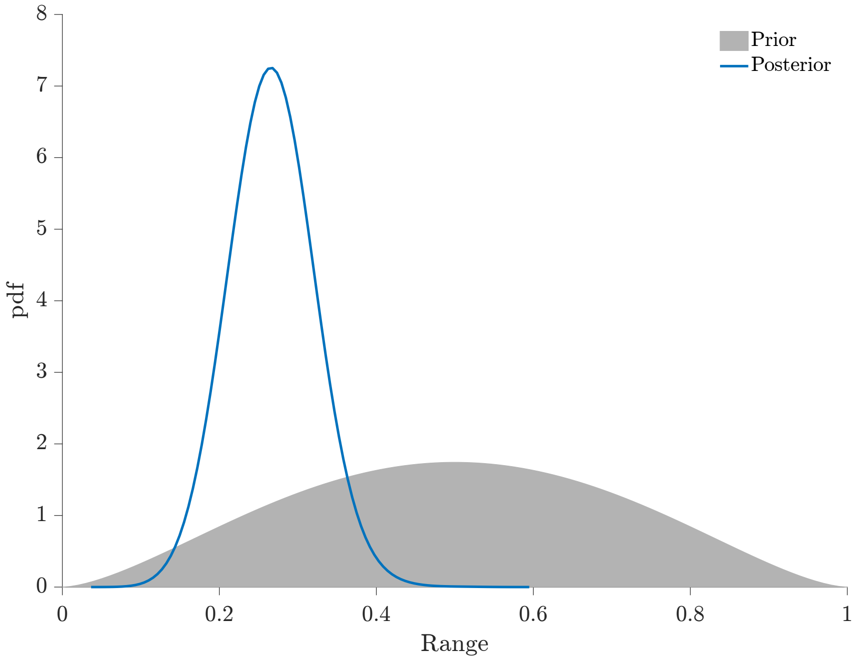


Figure 11: Trace Plots of Structural Parameters

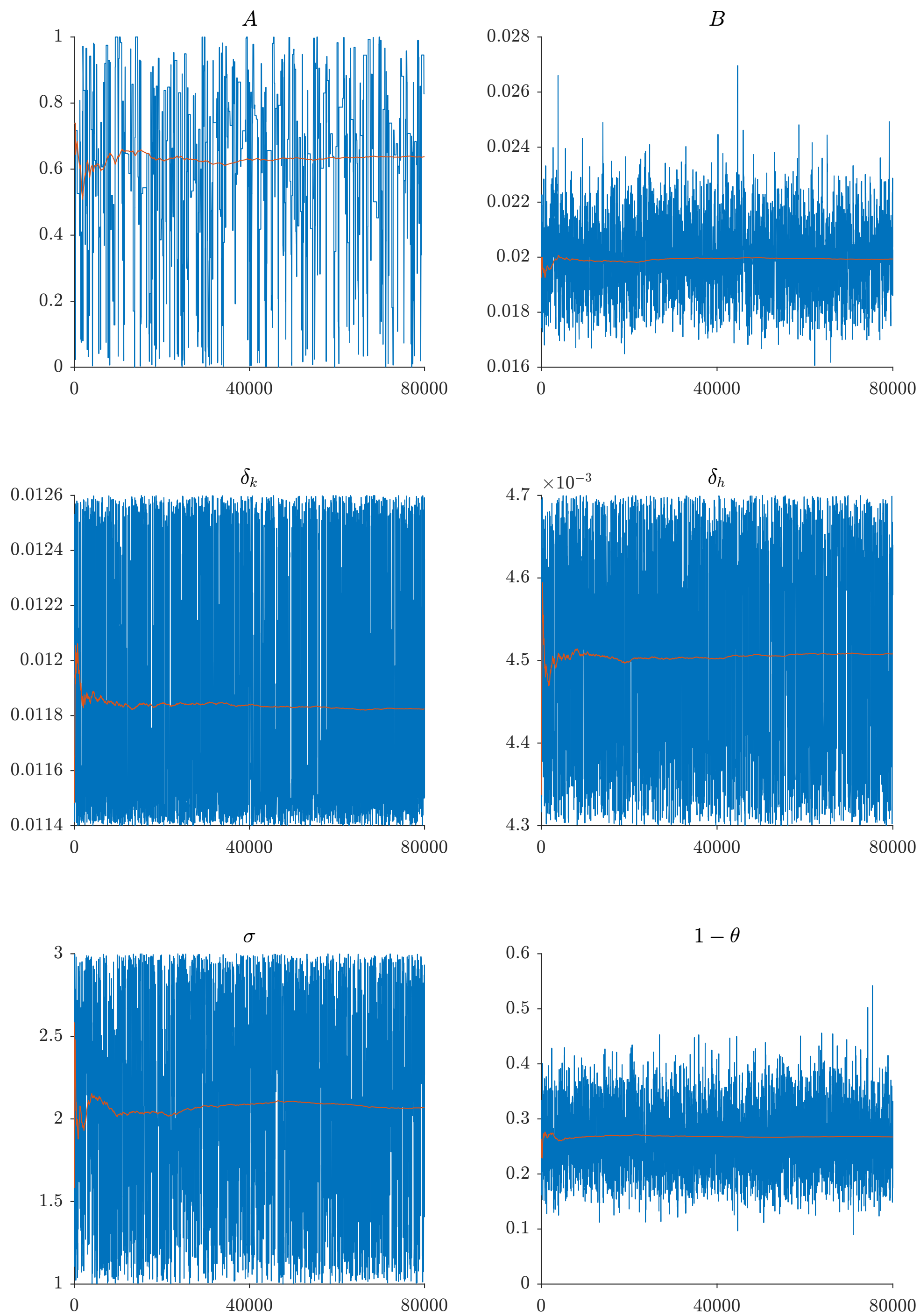


Figure 12: Trace Plots of AR Processes and Measurement/Specification Errors