Aging, Secular Stagnation and the Business Cycle

Callum Jones*

New York University

Job Market Paper

Please access the latest version at
sites.google.com/site/jonescallum

Abstract

In 2015, US output per capita was 12.5% below its 1950–2007 trend. This paper uses an estimated New Keynesian model to decompose the sources of this gap. The model features demographic trends, real and monetary shocks, and the occasionally binding zero lower bound on nominal interest rates. I calibrate demographic trends to observed mortality and fertility rates between 1940 and 2015, and estimate the model's business cycle processes on quarterly data from 1984 to 2015. The model is successful in accounting for the post-1990 trends in the real interest rate, the employment-population ratio, and labor productivity growth. I extract the model's structural shocks over the zero lower bound period and find that about half of the gap between output per capita and its long-run trend is due to the aging of the population, 30 per cent is due to real and financial factors, 5 per cent to nominal factors, and 15 per cent to the binding zero lower bound.

*Department of Economics, New York University. Email: callum.jones@nyu.edu. I am particularly thankful to Virgiliu Midrigan for valuable advice. I also wish to thank Jess Benhabib, Jaroslav Borovicka, Katarina Borovickova, Joseph Briggs, Tim Christensen, Mark Colas, Tom Cooley, Tim Cogley, Mark Gertler, Dan Greenwald, Aaditya Iyer, Malika Krishna, Nic Kozeniauskas, Mariano Kulish, Julian Kozlowski, Ricardo Lagos, Steven Pennings, Diego Perez, Klaus Peter-Hellwig, Thomas Philippon, Edouard Schaal, and Venky Venkateswaran for helpful discussions. I also thank participants of NYU macroeconomic seminars, the NYU alumni conference and participants of seminars at Haverford College, NYU Stern, the St. Louis Federal Reserve, and the World Bank for helpful suggestions. I thank the NYU Graduate School of Arts and Sciences for funding through the Dean's Dissertation Fellowship. All remaining errors are mine alone. Date: December 28, 2016.

1 Introduction

In 2015, US output per capita was 12.5% below its 1950–2007 trend. Furthermore, the Federal Reserve has held the Federal Funds rate at or near its lower bound since 2009, while the real interest rate declined from about 4% in 1995 to 1% in 2015 and the employment-population ratio has fallen from 64% in 2000 to just under 60% in 2016. Figure 1 illustrates these series.

Economists have proposed several theories for these observations. Summers (2014), for example, argues that the US is in a period of *secular stagnation* characterized by low real interest rates and constrained monetary policy. Gordon (2016) suggests that supply-side factors are key, as productivity growth has slowed over time. Rogoff (2015) argues that slow output growth is driven by financial factors and a protracted period of deleveraging, and Midrigan and Philippon (2016) show that household deleveraging shocks are highly contractionary when monetary policy is constrained.

The aging of the US population can explain, in theory, why real interest rates are low, as an individual's savings changes with their age, and an economy with a higher fraction of older people has more savings and lower interest rates (Eggertsson and Mehrotra, 2014). The aging of the US population can also explain why productivity growth and the employment-population ratio are low—younger workers have faster productivity growth and older workers supply fewer hours of labor.

This paper develops a model to quantitatively evaluate the relative contribution of the aging of the population, the zero lower bound on nominal interest rates, real and financial factors, and nominal factors in accounting for the large reduction in US output relative to trend. This question is important—if monetary policy does account for a large fraction of the drop in output, then there is scope for additional unconventional policy (Kocherlakota, 2016).

To answer the question, I use a dynamic stochastic New Keynesian model of the US with a rich demographic structure. Individuals in overlapping generations in the model can live for up to 80 years. They choose their labor supply and savings each period, borrowing when young and consuming their savings when old. Monetary policy has a role in the model through nominal rigidities, a monetary policy rule, and the zero lower bound on the nominal interest rate. There are aggregate shocks to household discounts, TFP, markups, and the policy rule. By imposing the occasionally-binding zero lower bound constraint, my model is able to capture the highly nonlinear interactions between demographic trends and a constrained monetary policy. This is important, because the aging of the population causes real and nominal interest rates to decline, so that the zero lower bound is more

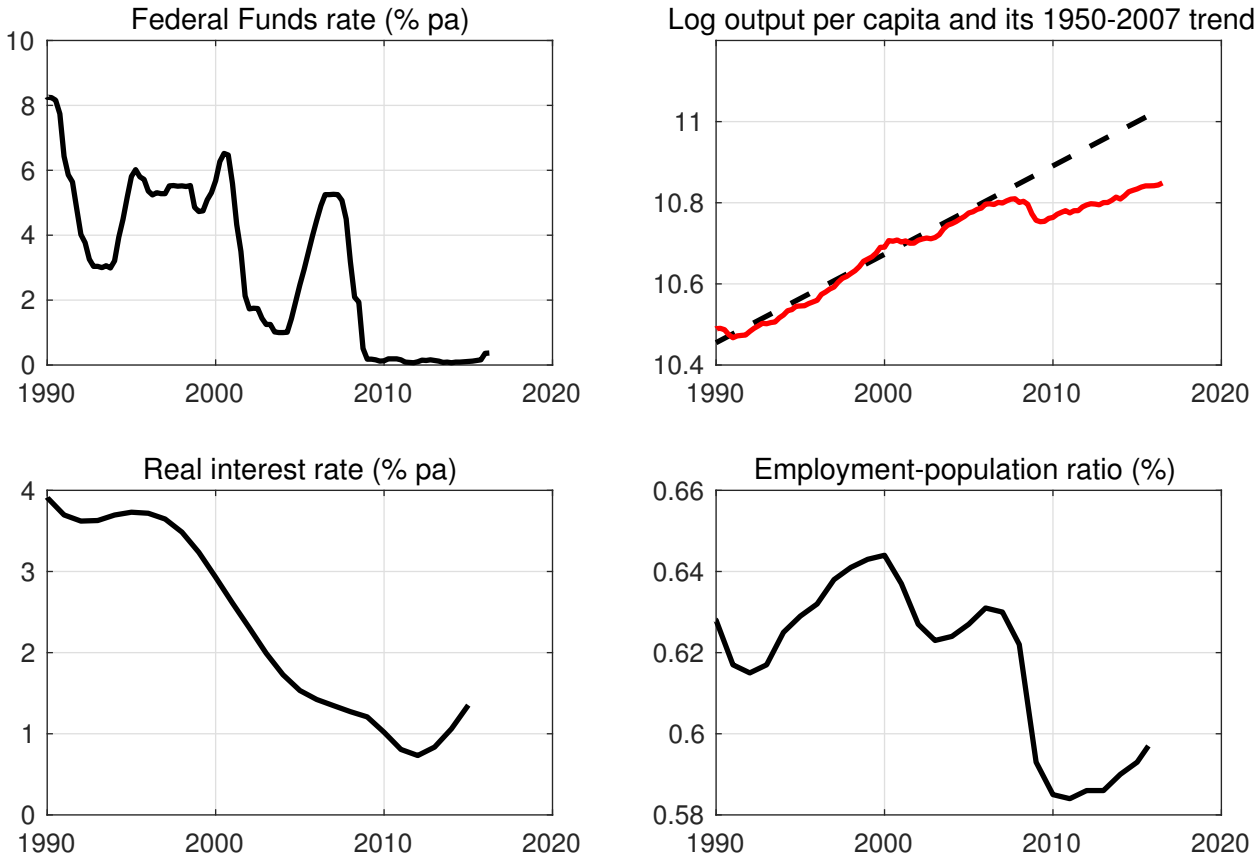

Figure 1: Macroeconomic trends. The real interest rate is constructed using the 10-year treasury yield and CPI inflation and is HP-filtered.

likely to bind for a given shock.

Solving the dynamic stochastic model with the zero lower bound and anticipated demographic trends raises computational challenges, particularly as my model has a large state-space for empirical realism. To overcome these difficulties, I use a piecewise-linear algorithm that I develop in which the monetary policy regime switches between one in which the zero lower bound binds and a regime in which normal Taylor-rule policy prevails (see Jones, 2015a; Guerrieri and Iacoviello, 2015). In a rational expectations solution, agents in the model foresee when each regime arises. The method can handle forward guidance policies, which are identified in this framework when the expected zero lower bound duration at a point in time is longer than the duration implied by shocks and the monetary policy rule. If this is the case, the central bank is using credible announcements to guide expectations in a way that stimulates consumption. In addition to being a computationally efficient way to approximate the zero lower bound, the piecewise-linear approach allows me to use standard maximum likelihood techniques to estimate the model's shocks with quarterly data.

The model's lifecycle parameters are calibrated to individual level data. Workers accumulate human capital over the lifecycle and gradually withdraw from the labor market as they age. There are two exogenous demographic trends in the model. First, workers have age-specific and time-varying mortality rates, which are matched to observed and projected mortality profiles. Second, the size of the initial cohort each period is chosen to match observed changes in the relative size of the the 16-year-old cohort. The model is calibrated to anticipated paths for these two demographic trends. I show that these demographic changes capture well the observed dynamics of the increase in the average age of the population since 1940.

In the model, the unintentional bequests of those who die between periods are redistributed to the remaining living members of each generation, following Yaari (1965) and Blanchard (1985). I show that this insurance against mortality risk means that that the consequences of demographic changes are approximated very well by slow-moving, anticipated wedges to the preferences of a representative agent and to aggregate output. These wedges vary over time with the aging population and can be readily constructed from population data and lifeycle parameters—in particular, the age-productivity profile and the disutility of labor supply by age. Approximating the model with these anticipated wedges reduces the model's state-space substantially. In the solution, I treat the paths of these wedges as anticipated changes in the parameters of the model.

Using this efficient representation, I apply likelihood methods to estimate the parameters of the four transitory shocks that drive fluctuations around the model's demographic trend using quarterly data on output, consumption, the Federal Funds rate, and inflation. During the zero lower bound period, I use expected durations implied by Federal Funds futures data as an observable to discipline the anticipated zero lower bound durations.

With the estimated model, I extract the set of structural shocks that hit the economy over the post-crisis period to decompose the gap between output and its 1950–2007 pre-crisis trend into demographic trends, real and nominal factors, and the zero lower bound, for 2015Q1. I find that removing the zero lower bound has a surprisingly small role, explaining about 15 percent of the gap. This is because the estimated model assigns a stimulatory role to forward guidance policy, mitigating the effect of the zero lower bound constraint during this period. This conclusion is consistent with the empirical evidence on the stance of monetary policy since the Great Recession (see Wu and Xia, 2016; Swanson and Williams, 2014). By turning off the shocks, I find that about half of the gap between output per capita and its trend is accounted for by demographic changes, caused by

the slowing growth rates of labor, capital, and productivity induced by the aging of the population. About 30 percent of the gap can be explained by shocks to productivity and to preferences, while 5 percent can be explained by markup shocks and nominal rigidities.

The results show that demographic trends drive a significant amount of the trend decline in output. I decompose further the effects of demographics on the model's trends in the real interest rate, productivity growth, and the employment-population ratio.

First, I find that demographics caused the real interest rate to decline by 1.6 percentage points from 1980 to 2015. This compares to an observed 1.9 percentage point decline from 1980 to 2015 and is in the range of recent studies (Gagnon et al., 2016; Eggertsson et al., 2016). The gradual fall in mortality rates causes a compositional shift towards an older population: longer expected lifetimes generate an increase in the demand for savings and capital for retirement consumption, contributing about a 1 percentage point decline in the real interest rate. Fertility shocks cause a large oscillation around the path implied by the fall in mortality rates because they change, over time, the relative size of the workforce to the capital stock, and therefore the marginal product of capital.

In studying the trend decline in the real interest rate, my focus on the US as a closed economy is shared by Gagnon et al. (2016). However, conclusions about the fall in real interest rates can, in theory, be sensitive to international factors. Gagnon et al. (2016) discuss reasons why the relationship between an aging population and a falling real interest rate is robust to open-economy considerations, noting, first, that most advanced economies are experiencing similar demographic changes, and second, that the observed magnitude and direction of international capital flows were opposite to the outflows that the aging of the US population should imply.[1]

Second, I find that my model predicts that the *growth rate* of labor productivity—output per hour worked—has declined by 0.46 percentage points from 1990 to 2015. This compares to the observed decline of slightly over 1 percentage point over the same period. While much of the change is driven by changes to the total amount of labor and capital, I find that changes in *labor quality* have been an important factor since the 1990s. In particular, when the baby boomer generation enters the workforce, there is an increase, and subsequent decrease, in the rate of human capital accumulation and therefore in output and productivity growth.

Third, I find that the employment-population ratio has declined by 1.2 percentage points in the

[1] Eggertsson et al. (2016) consider open-economy factors in their study of the real interest rate from the 1970s, finding that a decline in the real interest rate caused by capital inflows were roughly the same as an increase caused by the growth in US public debt–both factors are absent in my model.

model from 1990 to 2015. This compares to the observed period-to-period decline of 3 percentage points from 1990 to 2015. While general equilibrium wage effects do change the labor supply decision, the trend decline is primarily driven by changes in the composition of the workforce as workers live for longer and the labor force participation rate declines with age.

Related literature

This paper relates to a number of studies on the secular stagnation debate (Hamilton et al., 2015; Rogoff, 2015; Rachel and Smith, 2015; Bernanke, 2015; Summers, 2014; Eggertsson and Mehrotra, 2014). My emphasis on the importance of demographics in explaining the decline in the real interest rate is shared by the quantitative work of Gagnon et al. (2016), Eggertsson et al. (2016) and Carvalho et al. (2015). These papers study the decline in real interest rates by looking at nominal rates adjusted by the rate of inflation. In my work, I compare the decline in the real interest rate that my model predicts to the marginal product of capital, which I construct from observed capital-output ratios and which is closer to the model's object. Other papers tie to the secular stagnation literature by studying the decline in long-run growth and its consequences (Gordon, 2016; Antolin-Diaz et al., 2014; Fernald, 2015; Christiano et al., 2015). I complement this body of work by considering how demographics relates to measures of output and productivity growth through changes in aggregate hours and capital and a distributional channel by which the age-composition of the workforce interacts with the age-productivity curve.

This paper also relates to research that studies the macroeconomic implications of demographic trends (see, for example, Aksoy et al., 2015; Karahan et al., 2015; Fujita and Fujiwara, 2014; Backus et al., 2014; Bloom et al., 2001; Auerbach and Kotlikoff, 1987). There are also a number of empirical studies of the macroeconomic implications of demographic changes (see, for example, Bloom et al., 2011). My paper builds on the insights of the recent empirical work connecting demographic trends to output growth (Maestas et al., 2016), productivity growth (Feyrer, 2007), and the labor force participation rate (Aaronson et al., 2014), showing, in a lifecycle model, that demographics can jointly explain the trends in these variables.

Finally, this paper relates to studies of how the propagation of transitory shocks can change as an economy undergoes structural changes, either anticipated or unanticipated (Kulish and Pagan, 2016; Canova et al., 2015; Wong, 2015; Jaimovich and Siu, 2012; Fernández-Villaverde et al., 2007). I build on this literature by providing a methodology for explictly accounting for demographics in the

estimation of a business cycle model. These demographic trends are modeled as anticipated shocks to the parameters of a business cycle model. I focus also on modeling the nonlinear interactions that take place between the decline in the real interest rate and the zero lower bound.

2 Motivating evidence

This section presents motivating evidence on the aggregate effects of the aging of the population. I first show output growth and the age composition are correlated. I then study how the age-composition affects the labor force participation rate.

2.1 Output growth

How is output growth correlated with a country's age structure? Figure 2 presents the coefficients of a panel regression of real GDP growth per capita on the share of the population in 10-year binds for a broad set of countries. These results illustrate that countries with young populations tend to grow faster, and that the decline in growth is higher for older societies. Interpreting the magnitudes, the regression results suggest that a one percentage point increase in the share of the population made up of those aged between 30 and 39, on average increases per capita output growth by 0.37 percentage points. By contrast, an increase in the 70 to 79 population share by one percentage point decreases output growth by 0.88 percentage points. The pattern can be driven by how demographics changes the supply of labor and the supply of capital (Maestas et al., 2016) and by productivity (Feyrer, 2007). I use a structural model to decompose exactly how output growth changes over the demographic transition that is observed in the US.

2.2 Labor force participation

In this exercise, I fix the age-specific labor force participation rates at their values observed in a time period τ and then vary the population-shares of each cohort as observed to compute a counterfactual series for the aggregate participation rate PR_t:

$$\text{PR}_t = \sum_{s=16}^{95} \text{PR}_\tau^s \frac{n_t^s}{n_t}, \tag{1}$$

where PR_τ^s is the age s participation rate measured in period τ, and obtained from the Current Population Survey. The decomposition is repeated for PR_τ^s profiles observed between 1980 and 2010.

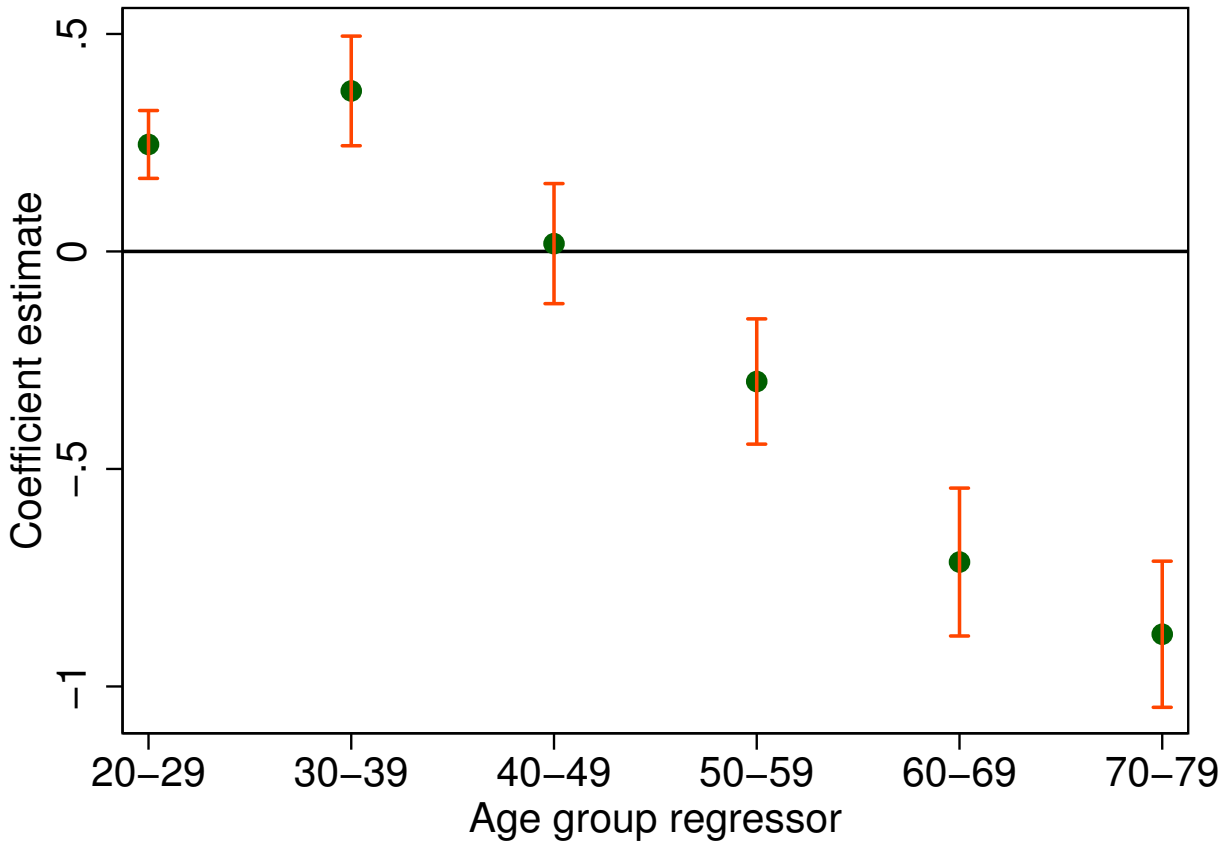

Figure 2: Regression coefficients of real GDP growth per capita on age shares. This figure shows the regression coefficients plus two standard error bands for regressions of real GDP growth per capita on the fraction of the population in different age bins. The countries included in the regressions are: Australia, Canada, France, Italy, Japan, New Zealand, the United Kingdom, and the United States, over the years 1950 to 2015. The data are from the United Nations Population Division and the Federal Reserve Economic Database.

Panel A of Table 1 presents the counterfactual total participation rate. It shows that demographics can replicate roughly half of the observed changes in the participation rate since 1996. In particular, changes in the size of the population appear to be able to capture some of the increase in the participation rate between 1970 and 1995, with the unexplained part driven by the increase in female labor force participation not captured when age-specific participation rates are held constant. Compositional changes then appear to be able to explain about one third of the decrease in the overall labor force participation rate between 2008 to 2015, which is the period when the first cohorts of the baby boomer generation started to withdraw from the labor market.

This compositional analysis suggests an important role for demographics in explaining changes in the participation rate, particularly for the post-1990 period. The degree to which it explains

Table 1: Participation rate and population composition

	Change, percentage points				Level, %		
PR^s_τ profile	1948 to 1969	1970 to 1995	1996 to 2007	2008 to 2015	1970	1990	2015
$\tau = 1980$	-3.05	-0.34	-0.28	-1.63	62.09	61.96	59.60
$\tau = 1990$	-3.34	-0.16	-0.35	-1.83	65.14	65.09	62.56
$\tau = 2000$	-2.99	-0.37	-0.15	-1.67	63.80	63.53	61.41
$\tau = 2010$	-3.47	0.08	-0.03	-1.38	65.11	65.31	63.59
Data	0.90	6.20	-0.80	-3.40	59.60	65.90	63.30

the decline varies with the profile for the age-specific participation rates. Using the age-specific participation rates observed in 1990 assigns about a 30% larger explanation for the decline in the total participation rate to changes in the age-composition of the workforce, as compared to the decline computed off the 2010 age-specific participation rate. This suggests that workers' labor supply decisions are offsetting some of the purely compositional effect of population changes on the aggregate labor force participation rate. The model I use in the next section endogenizes the labor supply decision at each age, accounting for how labor supply reacts to equilibrium changes in wages and asset holdings.

3 Model

I develop a model in which the macroeconomic trends are driven by the aging of the population. The focus on demographic changes as a driver of long-run trends is motivated by the results of the previous section. The model features business cycle fluctuations and the zero lower bound. In this section, I introduce the demographic structure by describing the overlapping generations setup of the model. I then show in Section 4 that these demographic trends can be represented as aggregate wedges to a representative agent's preferences and to technology. In Section 5, I add nominal frictions, monetary policy, and shocks to the representative agent model.

3.1 Households

The economy is populated by T overlapping generations. Each generation lives for a maximum of T periods, so that the age range of an individual is 0 to $T-1$. Generation s is of mass n^s_t, comprised of a continuum of identical members of age s. The size n^s_t is measured at the start of period t. I abstract from trend population growth, so that in the absence of exogenous increases in the size of

the incoming population, the initial population size is normalized to $n_t^0 = 1$. The total size of the population at time t is therefore:

$$n_t = \sum_{s=0}^{T-1} n_t^s.$$

Each period, a fraction of each age s cohort die with an exogenous age-specific mortality rate γ_t^s:

$$n_{t+1}^{s+1} = (1 - \gamma_t^s) n_t^s.$$

These mortality rates vary exogenously over time. In particular, permanent decreases in mortality rates reflect increases in longevity due to, for example, improvements in health.

An individual of age s has the period utility function $u(c_t^s, \ell_t^s)$ and chooses consumption c_t^s, labor supply ℓ_t^s, and savings a_t^s—claims to the aggregate capital stock—to maximize lifetime utility. The value function of the individual of age s at period t is therefore:

$$V_t^s(a_{t-1}^{s-1}) = \max_{\{c_t^s, \ell_t^s, a_t^s\}} \left\{ u(c_t^s, \ell_t^s) + \beta(1 - \gamma_t^s) V_{t+1}^{s+1}(a_t^s) \right\},$$

where β is the discount factor common to all individuals. Their age-specific mortality rate features in their discounting of the future. A maximum lifespan of T implies $\gamma_t^{T-1} = 1$.

The unintentional bequests made by individuals of a household who die between periods are aggregated and redistributed as an annuity to the remaining living households of the same generation, following Yaari (1965) and Blanchard (1985), eliminating mortality risk.[2]

Individuals have age-specific productivities z^s, receives a transfer from the government ξ_t^s which are described in more detail below, receive a gross return R_t on last period's savings, and receives d_t^s for the redistributed unintentional bequest.[3] The period budget constraint of the individual is:

$$c_t^s + a_t^s \leq z^s w_t \ell_t^s (1 - \tau_t) + \xi_t^s + R_t a_{t-1}^{s-1} + d_t^s.$$

Consumption in the last period of life equals the return on an individual's remaining assets:

$$c_t^T \leq R_t a_{t-1}^{T-1}.$$

[2]For an in-depth discussion of this point, see Hansen and Imrohoroglu (2008). The assumption of annuities markets in quantitative lifecycle models is common. See, for example, Backus et al. (2014).

[3]As these unintentional bequests are redistributed equally to members of the same generation, this can be also be written as scaling the return on savings a_{t-1}^{s-1} by $1/(1 - \gamma_{t-1}^{s-1})$, as $d_t^s = (n_{t-1}^{s-1}(1 - \gamma_{t-1}^{s-1}))/(n_{t-1}^{s-1}\gamma_{t-1}^{s-1}) R_t a_{t-1}^{s-1}$ (also see Hansen and Imrohoroglu, 2008).

By assumption, the individual retires fully from the labor market in her last period of life.

3.2 Production

Firms hire labor and capital from individuals and operate a Cobb-Douglas production technology: $y_t = k_{t-1}^{\alpha} \ell_t^{1-\alpha}$. Aggregate capital is the sum of each cohorts's savings: $k_t = \sum_s n_t^s a_t^s$, while aggregate labor hired by the firm is in efficiency units of labor $\ell_t = \sum_s z^s n_t^s \ell_t^s$. Firms use the capital of the deceased in production. The firm pays the marginal product of capital $r_t = \alpha \frac{y_t}{k_{t-1}}$ and the marginal product of labor for the wage: $w_t = (1-\alpha)\frac{y_t}{\ell_t}$.

3.3 Government

The government taxes labor income at the rate τ_t to fund a pay-as-you-go social security system. I follow Attanasio et al. (2007) in specifying that the benefit paid each period above an eligibility age T^* depends on three elements: the expected remaining life of the recepient, the accumulated pre-tax labor income of the worker, and a parameter λ governing the replacement rate of past earnings. Denote by W_t^s accumulated gross lifetime earnings, defined recursively as:

$$W_t^s = \begin{cases} w_t z^s \ell_t^s + W_{t-1}^{s-1}, & \text{if } s < T^* \\ W_{t-1}^{s-1}, & \text{if } s \geq T^*. \end{cases}$$

The amount ξ_t^s redistributed to an agent of age $s \geq T^*$ depends on W_t^s:

$$\xi_t^s = \lambda \frac{W_t^s}{(T^* - 1)},$$

where the denominator reflects the amount of time that W_t^s is accumulated over. For those younger than the eligibility age T^*, the transfer $\xi_t^s = 0$. Individuals aged above T^* can work for labor income and earn social security payments without penalty.[4] The government budget constraint is:

$$\sum_s n_t^s \xi_t^s = \sum_s n_t^s z^s w_t \ell_t^s \tau_t.$$

The tax rate τ_t adjusts to equalize social security outlays and tax revenues.

[4] I abstract from a reduction in retirement benefits resulting from taking social security benefits early, which would require a choice made between the ages of 62 and 67. I also abstract from a number of questions about the sustainability of pension systems in an aging society: an in-depth analysis of these issues is in Attanasio et al. (2007).

3.4 Equilibrium

In equilibrium, households and firms optimize their choices of consumption, savings, and labor. Each individual's optimal choice of consumption and savings delivers an Euler equation for consumption $u_1(c_{t+1}^{s+1}, \ell_{t+1}^{s+1})/u_1(c_t^s, \ell_t^s) = \beta R_{t+1}$. Labor supply reflects the marginal rate of substitution between consumption and leisure: $u_1(c_t^s, \ell_t^s)/u_2(c_t^s, \ell_t^s) = z^s w_t$. Firms set wages to the marginal product of labor: $w_t = (1-\alpha)\frac{y_t}{\ell_t}$, and the gross return on capital to its marginal product $r_t = \alpha \frac{y_t}{k_{t-1}}$. Asset market clearing requires that aggregate savings equals the firm's capital demand:

$$k_t = \sum_s n_t^s a_t^s.$$

Labor market clearing sets the total amount of labor supplied equal to firm's demand for labor:

$$\ell_t = \sum_s z^s n_t^s \ell_t^s.$$

Goods market clearing pins down the real allocations of consumption and savings:

$$y_t = c_t + k_t - (1-\delta)k_{t-1}.$$

Finally, the government budget constraint is balanced.

4 Aggregate representation with demographic wedges

In this section, I discuss how heterogeneity in the lifecycle framework at a point in time can be approximately aggregated using wedges that attach to the preferences of a representative agent and to the level of technology. For convenience, I call these wedges *demographic adjustment* factors (DAs). This representation is useful for the computations and for understanding the effect of demographics.

4.1 Derivation

I follow Constantinides (1982) and Maliar and Maliar (2003) and proceed in two steps.[5] In the first step, I consider the problem of a social planner who maximizes a weighted sum of each individual's utility function. In the solution to this problem, the planner distributes aggregate consumption

[5] A full derivation is given in the appendix.

and aggregate labor supply between individuals alive in each period. In the second step, I solve the problem where the planner then maximizes aggregate consumption, capital, and labor supply subject to the economy's resource constraint.

An important assumption needed for the derivation is the redistribution scheme assumed for the unintentional bequests of those who die between periods. Because of the insurance provided by this scheme, the individual faces no uncertainty along the demographic path. With claims to capital, they smooth consumption along that path. The result of this complete markets setup is that the ratio of marginal utilities of wealth across individuals is the same at any point in time, which implies the existence of the welfare weights that attach to each individual's utility function in the planner's problem. Those weights equate the planner's solution with the decentralized equilibrium.

The solution in the appendix shows that when individuals of age s have the period utility function over consumption c_t^s and hours ℓ_t^s of the type $\frac{(c_t^s)^{1-\rho}}{1-\rho} - v^s \frac{(\ell_t^s)^{1+\varphi}}{1+\varphi}$, the social planner—the representative agent—has preferences over aggregate consumption c_t and aggregate *efficiency units* of labor ℓ_t with the period utility function taking the form:

$$U(c_t, \ell_t) = \phi_t \frac{c_t^{1-\sigma}}{1-\sigma} - v_t \frac{\ell_t^{1+\varphi}}{1+\varphi}. \tag{2}$$

The representative agent's problem will be to maximize this utility function over time with discount factor β by choosing aggregate consumption c_t, efficiency units of labor ℓ_t and capital k_t subject to the economy's resource constraint and its production function $y_t = \theta_t^{1-\alpha} k_t^{\alpha} \ell_t^{1-\alpha}$, and where the relationship between the efficiency units of labor and aggregate hours is $\ell_t = A_t h_t$.

The setup shows how the demographic state affects the aggregate economy through four wedges: ϕ_t, v_t, θ_t, and A_t. First, demographics affects aggregate output through θ_t. The second is the process A_t that affects the labor-input relationship. The third is the wedge to the marginal utility of consumption ϕ_t. The fourth is affects the aggregate disutility of providing labor v_t, which, together with the preference shock ϕ_t, attaches to the labor wedge, affecting the incentive to supply labor. I discuss each DA wedge and its microfoundations in turn.

4.2 Productivity wedge

The first wedges are to productivity, shifting total output θ_t and shifting labor-input A_t. Substituting the definition of efficiency units of labor $\ell_t = A_t h_t$ into the production function gives the production

function:

$$y_t = \theta_t^{1-\alpha} k_t^{\alpha} (A_t h_t)^{1-\alpha}.$$

In the appendix, I show that θ_t and A_t are:

$$\theta_t = \sum_s n_t^s z^s, \qquad \text{and} \qquad A_t = \frac{\sum_s n_t^s (\hat{z}^s)^{1+1/\varphi} (v^s \lambda^s)^{-1/\varphi}}{\sum_s n_t^s (\hat{z}^s)^{1/\varphi} (v^s \lambda^s)^{-1/\varphi}}, \tag{3}$$

where the value $\hat{z}^s = z^s/\theta_t$ denotes the individual s's skill level relative to the average skill level in the economy θ_t, and the λ^s parameters are the Pareto weights attached to an individual of age s. The shock θ_t encodes the increase in output caused by changes in the size and composition of the workforce over idiosyncratic skill levels. In particular, larger populations have higher levels of θ_t, as do populations whose composition is more heavily weighted towards more productive workers.

The wedge A_t has the interpretation as a population-weighted average of individual labor and skills supplied, reflecting the amount of hours needed to obtain an effective unit of labor. It incorporates both relative productivities and the disutility of providing labor. When labor supply is inelastic, θ_t and A_t combine to generate a total productivity adjustment factor equal to $\frac{\sum_s n_t^s z^s}{\sum_s n_t^s}$: the productivity adjustment in this case reflects only the age-composition of the population. In principle, θ_t and A_t are straightforward to compute, requiring only hours by age and assumptions about the age-productivity profile, the age-disutility of work profile, Pareto weights that attach to each generation, and a value for the inverse Frisch elasticity of labor supply φ.

4.3 Consumption preference wedge

The second demographic adjustment wedge attaches to the marginal utility of consumption:

$$\phi_t c_t^{-\sigma} = \lambda_t.$$

where λ_t is the Lagrange multiplier on the economy's resource constraint. In the appendix, I show that this wedge has a simple expression:

$$\phi_t = \left[\sum_s n_t^s \lambda^s \right]^{\sigma}. \tag{4}$$

The aggregate preference wedge simply maps to the size of the population at each point in time, and is increasing in the curvature of the utility function σ. A larger population scales the marginal

utility of consumption and aggregate consumption.

4.4 Marginal disutility of labor wedge

The third wedge, v_t, is an aggregate trend in the marginal disutility of labor. Equating the marginal utility of consumption and the marginal disutility of labor, and substituting in for hours worked:

$$\frac{w_t}{\ell_t^{\varphi}/c_t^{-\sigma}} = \frac{v_t}{\phi_t}.$$

This says that a component of the labor wedge measured with efficiency units of labor is the ratio of the two demographic shocks v_t/ϕ_t, where v_t is a population-weighted average of individual disutilities of providing labor, shown in the appendix to be:

$$v_t = \left[\sum_s n_t^s (\hat{z}^s)^{\frac{1}{\varphi}+1} (v^s \lambda^s)^{-\frac{1}{\varphi}}\right]^{-\varphi}. \tag{5}$$

An increase in the average disutility of labor increases the labor wedge, so that a level increase in v^s scales linearly v_t. Furthermore, changes in the distribution of the population towards younger workers who have lower disutility of providing labor reduces v_t.

4.5 Approximation for longevity and proportional taxes

Two additional trends are needed in the computations to ensure the aggregate representation approximates the full lifecycle solution well. The first is a slight trend in the aggregate resource constraint to account for variation in the average mortality rate over time. The second is one to proportional taxes. I take the path of labor income taxes as given and anticipated.

To verify that these shocks recover the paths of the real variables from the global solution to the lifecycle model, I take the aggregate series implied by the lifecycle model and compare it to the linear approximation that is computed under anticipated shocks, under the assumption that the Pareto weights attached to each individual are the same.[6] The two series are plotted in Figure C.1 in the appendix, showing that the aggregate representation generates paths for aggregate variables which are very close to those generated by the path computed from the full nonlinear lifecycle model and perfectly foreseen paths for the aging of the population.

[6] The solution method is described in more detail in Section 7.

5 Nominal rigidities, monetary policy and shocks

In this section, I add New Keynesian frictions and aggregate business cycle shocks to the model. I add quadratic capital adjustment costs, intermediate goods-producers who produce a differentiated good, quadratic costs of adjustment over those intermediate goods' prices, and a monetary policy rule that responds to inflation, the growth rate of output, and the gap between the level of output and its steady-state value. The nominal interest rate is additionally subject to the zero lower bound.

In addition to the nominal frictions and monetary policy, I add to the model four transitory, unanticipated shocks: to the discount factor, to the level of technology, to the elasticity of substitution between intermediate goods, and to the Taylor rule. These four shocks will drive business cycle fluctuations around the model's demographic trend.

5.1 Firms

I enrich the firm side of the model with a continuum of intermediate goods firms i who access the economy's Cobb-Douglas technology, hiring capital and labor from households to supply a substitutable good $y_t(i)$ to final goods producers, who in turn sell the final good to consumers at price P_t. The elasticity of substitution between intermediate goods ξ_t is subject to stochastic shocks, which generates time-varying markups over marginal costs. The production function is:

$$y_t(i) = \theta_t^{1-\alpha} \mu_t^{1-\alpha} \left(k_{t-1}(i)\right)^{\alpha} \left(A_t h_t(i)\right)^{1-\alpha}, \tag{6}$$

where there is a stationary economy-wide technology process μ_t:

$$\ln \mu_t = (1-\rho_\mu) \ln \mu + \rho_\mu \ln \mu_{t-1} + \sigma_\mu \varepsilon_{\mu,t}, \tag{7}$$

which is subject to standard normal shocks $\varepsilon_{\mu,t}$. Intermediate goods-producing firms face a Rotemberg quadratic cost of price adjustment, parameterized by ϕ_p. Denoting mc_t as the Lagrange multiplier on the cost minimization problem of the firm, the rental rate on capital is:

$$r_t = \alpha \text{mc}_t \frac{y_t}{k_{t-1}}, \tag{8}$$

and the wage is:

$$w_t = (1-\alpha)\text{mc}_t \frac{y_t}{\ell_t}. \tag{9}$$

The first order condition on the optimal choice of price resetting is:

$$\beta\phi_p \mathbb{E}_t \frac{\lambda_{t+1}}{\lambda_t}\frac{y_{t+1}}{y_t}\left[\frac{\Pi_{t+1}}{\Pi_t}-1\right]\left[\frac{\Pi_{t+1}}{\Pi_t}\right] = \xi_t - 1 - \xi_t \text{mc}_t + \phi_p\left[\frac{\Pi_t}{\Pi_{t-1}}-1\right]\left[\frac{\Pi_t}{\Pi_{t-1}}\right], \tag{10}$$

where $\Pi_t = P_t/P_{t-1}$ is the rate of inflation and where ξ_t is an autoregressive process for the elasticity of substitution between intermediate goods:

$$\ln \xi_t = (1-\rho_\xi)\ln\xi + \rho_\xi \ln \xi_{t-1} + \sigma_\xi \varepsilon_{\xi,t}. \tag{11}$$

Log-linearizing (10) gives rise to a standard forward-looking Philips curve.

5.2 Monetary policy

The estimation is specified to allow for calendar-based forward guidance policy to be used by the central bank. In Jones (2015a), I discuss how the additional state variable that is needed to describe the state of the economy at each point in time that the zero lower bound binds is the number of quarters that the constraint is expected to bind.[7] This duration can, in principle, be any length. There is a duration which is consistent with the structural shocks that prevail at each point in time. I call this duration the *endogenous duration*. If the endogenous duration is the one that is expected by all agents in the economy, the central bank is acting passively in response to those structural shocks and expects to raise the nominal interest rate off its constraint as the policy rule prescribes. If, instead, the expected zero lower bound duration is longer than the endogenous duration, then agents in the economy believe the monetary authority is making a commitment to hold its policy rate at zero beyond that implied by the endogenous duration. This credible *forward guidance* announcement guides expectations about the path of interest rates, stimulating consumption (see Werning, 2012).

Let the variable FG_t indicate whether forward guidance is in use. If forward guidance is in use, $\text{FG}_t = 1$, and the nominal interest rate remains at its lower bound. When forward guidance is not in use, monetary policy follows a standard Taylor rule, with the nominal interest rate responding to

[7] See also the application in Midrigan and Philippon (2016).

deviations in inflation from a target rate Π^*, deviations in output from its steady-state level, and the growth rate of output, and is subject to the zero lower bound:

$$R_t = \begin{cases} \max\left(1, R_{t-1}^{\phi_r}\left(\frac{\Pi_t}{\Pi^*}\right)^{\phi_\pi}\left(\frac{y_t}{y}\right)^{\phi_y}\left(\frac{y_t}{y_{t-1}}\right)^{\phi_g}\exp(\varepsilon_{R,t})\right), & \text{if } \mathrm{FG}_t = 0 \\ 1, & \text{if } \mathrm{FG}_t > 0, \end{cases} \tag{12}$$

where y is the steady-state value of output. The monetary policy shock is $\varepsilon_{R,t}$.

5.3 Representative household and equilibrium

In addition to the firms' equilibrium conditions and monetary policy, the representative household with preferences (2) has optimal choices over consumption:

$$\lambda_t = \phi_t \exp(\chi_t) c_t^{-\sigma}, \tag{13}$$

where the exogenous processes for the discount factor shock χ_t is autoregressive:

$$\chi_t = (1-\rho_\chi)\ln\chi + \rho_\chi \chi_{t-1} + \varepsilon_{\chi,t}. \tag{14}$$

The choice of labor supply gives:

$$\lambda_t = \frac{v_t \exp(\chi_t)\left(A_t h_t\right)^{\varphi}}{w_t(1-\tau_t)}. \tag{15}$$

There is an Euler equation associated with the choice of nominal bonds:

$$\mathbb{E}_t \frac{\lambda_{t+1}}{\lambda_t} = \mathbb{E}_t \frac{1}{\beta}\frac{\Pi_{t+1}}{R_t}, \tag{16}$$

and an Euler equation associated with the choice of capital:

$$\mathbb{E}_t \frac{\lambda_{t+1}}{\lambda_t} = \mathbb{E}_t \frac{1}{\beta}\left[1+\phi_k\left(\frac{k_t}{k_{t-1}}-1\right)\right]\left[\frac{\phi_k}{2}\left(\frac{k_{t+1}^2}{k_t^2}-1\right)+1-\delta+r_{t+1}\right]^{-1}, \tag{17}$$

where quadratic costs of capital adjustment are parameterized by ϕ_k. Finally, the goods market clears:

$$y_t = c_t + k_t - (1-\delta)k_{t-1} + \tfrac{\phi_k}{2}\left(\tfrac{k_t}{k_{t-1}}-1\right)^2 k_{t-1} + \tfrac{\phi_p}{2}\left[\tfrac{\Pi_t}{\Pi_{t-1}}-1\right]^2 y_t. \tag{18}$$

To solve the system (6) to (18) in a symmetric equilibrium, I linearize them around the steady-state at each point in time and use the solution method described in the next section.

6 Estimation

This section discusses the calibration of the lifecycle parameters of the model, the identification of the transitory and permanent demographic changes which enter the model, and the estimation of the transitory processes that drive fluctuations around the model's demographic trends.

6.1 Calibrated parameters and demographic paths

Before estimation, I calibrate the demographic trends and a subset of the parameters.

6.1.1 Lifecycle parameters

I first calibrate the lifecycle parameters and demographic trends driving the aging of the population. Households are assumed to enter the model at 16 years of age and live for a maximum of 80 additional years, up to age 95.[8] Full retirement is only imposed in the last period of life.[9]

The disutility of providing labor v^s is calibrated following Kulish et al. (2010) who use the functional form of a scaled cumulative density function of a normal distribution.[10] Panel C of Figure 3 plots v^s over age. The disutility of labor supply is increasing in s, which is motivated by studies which link the disutility of work to deteriorating health.[11] The parameters of the function for v^s are chosen so that the participation rate by age broadly matches observed participation rates in 2000. In the appendix, I plot in Figure B.1, age-specific participation rates in the data against the model's predictions during the transition path.

I calibrate the age-productivity parameters z^s to the age-experience earnings profile. I follow Elsby and Shapiro (2012) in constructing the log experience-earnings profile. In constructing this profile, I use deflated data on full-time, full-year workers. The data is decennial Census data from

[8] During the transition, the proportion of people who live past 95 is very small, at roughly 0.5% of the population.

[9] Given the low choice of labor supply at older ages, this choice is not too important. Kulish et al. (2010) show how *unanticipated* changes in life expectancy can change labor supply decisions deep into the period of retirement, as older workers with very few assets remaining return to the workforce to fund consumption during their unanticipated increase in the lifespan. In contrast, in my framework, all demographic changes are perfectly foreseen.

[10] The formula is left to the appendix.

[11] Kulish et al. (2010) also choose to make the function time-varying with increases in life-expectancy, with the result that the disutility of labor from employment declines in the gap between age and life-expectancy (see also, Bloom et al., 2011). I keep it constant to ensure age-participation rates do not vary significantly over time.

1960 to 2000, and annual American Community Survey data from 2001 to 2007.[12] In constructing the log experience-earnings profile, I pool high school dropouts, high school graduates, those with some college education, and those who have completed college or higher education.[13] Panel A of Figure 3 plots the earnings-profile over age. The estimates imply a peak increase in wages of about 134% at age 45, before gradually declining around the age of 50. This profile is in line with the estimate of Guvenen et al. (2015) who find an increase in the earnings of the mean worker of 127%.[14] Given the lack of reliable experience-earnings data on the productivity of older workers, after age 65, I calibrate the productivity of workers to decay by 20% a year.[15]

For the social security system, I set the replacement ratio of accumulated earnings λ to 46.7%, the same value that is used in Attanasio et al. (2007), which in turn was based on the study of social security systems by Whitehouse (2003).

6.1.2 Mortality profiles

I calibrate the mortality probabilities of each generation during the 80 years they could possibly live, γ_t^s, to match the actuarial probabilities reported by the Social Security Administration.[16] Calibrating to these probabilities also matches changes in the life expectancy of each generation over time, conditional on an individual reaching 16 years of age. The values used are the cohort-specific survival rates computed for the cohort year of birth. These profiles include both observed survival rates of cohorts up to their current age, and extrapolated survival rates based on the Social Security Administrations's forecasts of life expectancy. I assume that all changes to these actuarial probabilities are exogenous and perfectly foreseen. For the initial γ_t^s profile, I use the survival probabilities reported for those born in 1900 onwards. For those cohorts born before 1900 but who are alive in 1940, I use extrapolated values of the survival probabilities.[17] Panel B of Figure 3 plots

[12] Computed off IPUMS-USA extracts. A full description is given in the appendix.

[13] Clearly, a finer calibration would distinguish between workers of different education levels. In robustness exercises reported in the appendix, I distinguish between these education groups and analyze how anticipated changes in the earnings profile map to labor supply decisions. In particular, Kong et al. (2016) find flattening experience-earnings curves in synthetic cohorts constructed off IPUMS-USA data. The key results are quite robust to these findings.

[14] This number corresponds to the mean worker, which I match as matching means is more suitable when studying asset accumulation. Guvenen et al. (2015) report that the median worker has a smaller increase of about 35%.

[15] This assumption is not too strong, given the small number of workers who remain in the labor force beyond 65 years of age. In the appendix I consider a productivity profile which holds productivity, for workers above age 65, fixed at the last observation at age 65, as in Kulish et al. (2010). This arguably aligns more with the profiles derived by Casanova (2013), who finds more or less a discrete level shift in earnings for older workers.

[16] These probabilities were sourced from Table 7 from the Cohort Life Tables for the Social Security Area by Calendar Year, in Actuarial Study No. 120 by Felicitie C. Bell and Michael L. Miller, available at: `https://www.ssa.gov/oact/STATS/table4c6.html`. A full description is given in the appendix.

[17] Because the survival probabilities are quite low for those years, the results are robust to alternative specifications and, in any event, are not that important for the model outcomes beyond 1970, which is the period I am interested in.

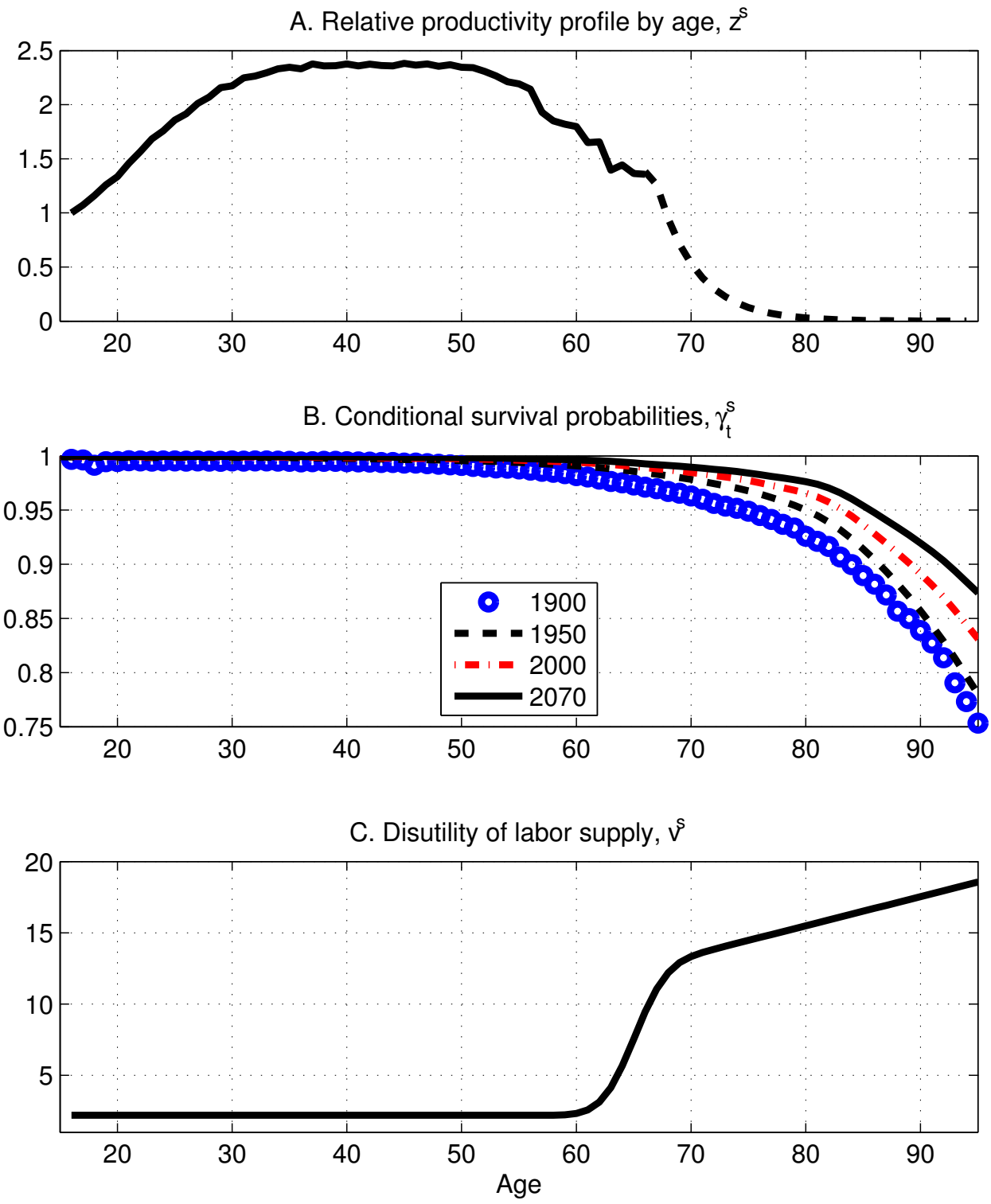

Figure 3: Calibrated lifecycle profiles. This figure shows the calibrated lifecycle profiles for the age-productivity parameters, the conditional survival probabilities and the disutility of labor supply.

the survival probabilities for cohorts born between 1900 and 2070.

6.1.3 Incoming cohort-size

To capture the demographic dynamics of the baby boom, I choose anticipated shocks to the size of the incoming cohort so that the change in the observed cohort share is the same as the change in the model cohort share.[18] This approach imperfectly captures changes in the population distribution due to, for example, immigration.[19] Panel C of Figure 11, discussed later, plots this fraction over time, clearly showing the effect of the increase in fertility on the fraction of young workers in the economy. I assume that changes to the incoming population beyond 2015 decay to zero, so that the population distribution converges to the steady-state implied by the mortality profile that is constant from 2070.

6.1.4 Demographic adjustment wedges

What does the calibration imply for the paths of the demographic adjustment trends? The paths of these are given in Figure 4.[20] The two shocks in Panel A and Panel B summarize how the productivity demographic adjustment (DA) wegdes expressed in (3) vary during the demographic transition. There is both a level effect and a transitory effect. To understand why, consider the value of the wedge absent the population shocks to the 16 year-old cohorts. Because of the larger size of the population which occurs as individuals live for longer, the wedge to θ_t gradually increases over time. The labor-input wedge, however, slowly decreases over time absent changes in fertility. Over the full demographic transition, the total productivity wedge increases with the size of the population, which swells as the baby boomer generations enter the workforce. The increase in aggregate productivity peaks in the year 2015, after which it declines as workers leave the labor market.

The demographic trends imply a large increase in the DA which attaches to the marginal utility of consumption, the process expressed in (4). This wedge increases significantly in the 1960s reflecting the increasing size of the population with the entrance of the baby boomers. Over the

[18] Choosing initial population shocks to matching the *changes* is necessary because the implied steady-state cohort sizes under the γ_t^s profiles do not match the actual cohort sizes at each point in time, and the model is initialized at the 1940 steady-state. In practice, these choices are quantitatively minor and unimportant for the results.

[19] To understand how important migration is to changing the population structure, I considered statistics on the characteristics of immigrants, reported by The Migration Policy Institute, at www.migrationpolicy.org. The median age of immigrants in the US has increased from the 1980s/1990s to 2010. In 2014, the median age of immigrants was 43 years, compared to the 36 years in the general population.

[20] In computing these trends, I set the Pareto weights attached to each individual to be equal. This gives a very accurate path of variables of the model, plotted in Figure C.1.

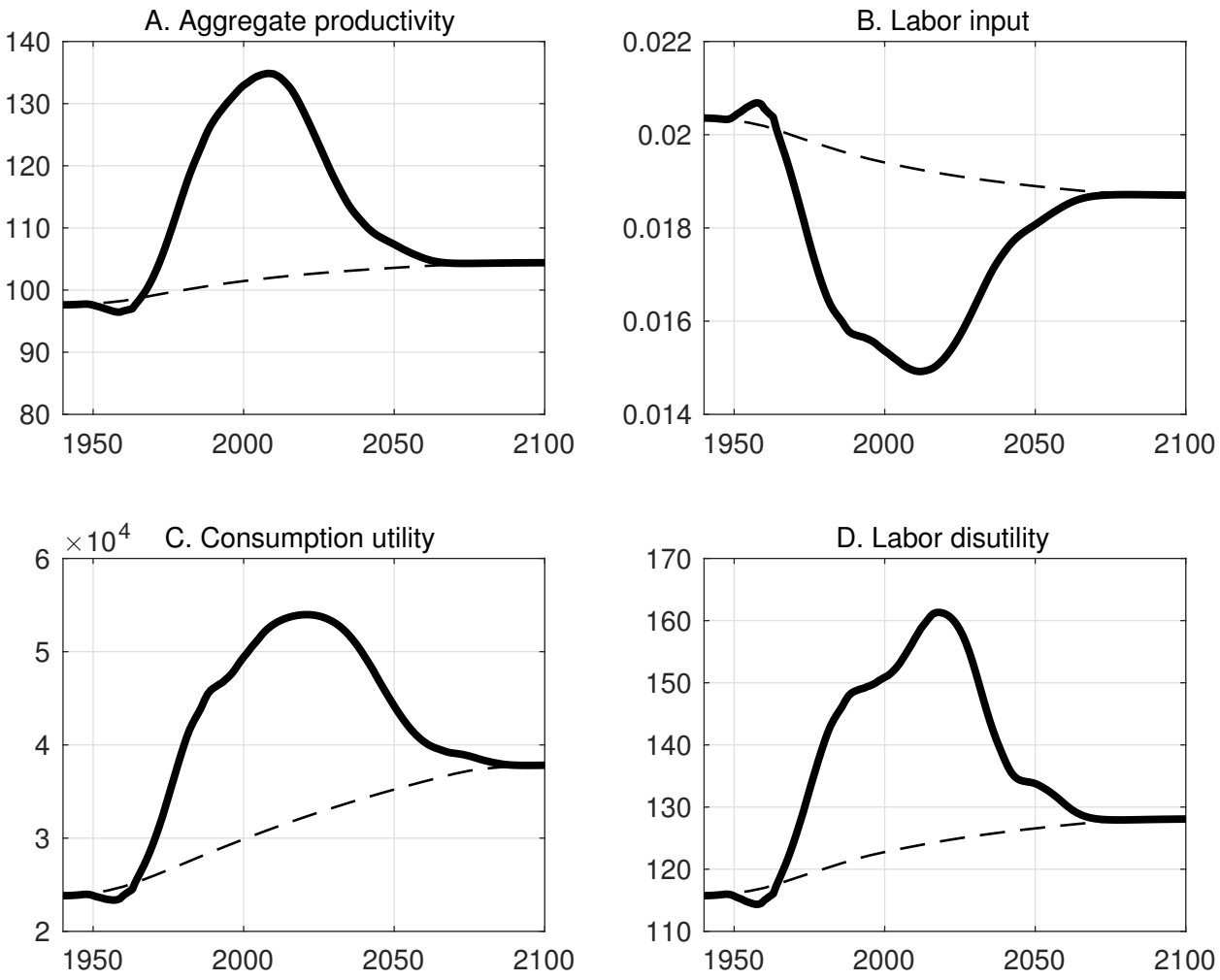

Figure 4: Demographic adjustment wedges. This figure shows the paths of the aggregate anticipated wedges that summarize the aggregate effects of heterogeneity in the lifecycle model. Aggregate productivity increases with the population size, peaking around 2010. The consumption utility wedge also increases with the population size, peaking around 2025.

full demographic transition, the consumption demand wedge peaks around 2020, which is after the peak in the productivity wedge. This reflects consumption smoothing on the part of households, with their consumption staying high after they exit the labor force.[21] The preference wedge is at an elevated level around the period of the Great Recession. This suggests that a given percentage change in the aggregate discount factor scales with the value of the aggregate demand wedge, providing a mechanism to justify a large negative preference shock around the time of the recession, consistent with a number of papers that use large discount factor shocks to model the Great Recession and ensure the nominal interest rate hits the zero lower bound (see, for example, Eggertsson and

[21]This observation suggests implications for asset prices more generally. In particular, the consumption utility wedge at this point is less pronounced than the labor productivity wedge, suggesting strong demand for an asset like housing in the run-up to the financial crisis.

Woodford, 2003).

Over the demographic transition, the value of v_t, expressed in (5), increases substantially to peak in the year 2000, before decreasing. With no fertility shocks, v_t trends up, as more weight is placed on higher disutilities by age so that, on average, there is a higher disutility of supplying labor.

6.1.5 Preference and nominal parameters

The remaining parameters are set to values which imply steady-state capital-output ratios that align with those in the Bureau of Labor Studies' Multifactor Productivity (BLS-MFP) program, which were also used by Fernald (2012).[22] I set capital to depreciate by $\delta = 10\%$ a year. The quarterly discount factor β is set to 0.9938. The capital share α is set to 1/3, the average of the capital share reported by Fernald (2012) over 1948 to 2015. The intertemporal elasticity of substitution ρ is set to 2.5, and the inverse Frisch elasticity of labor supply φ is set to 2.5, the central estimate of Reichling and Whalen (2012) and in line with the analysis of Rios-Rull et al. (2012). Both values are consistent with the range cited in Auerbach and Kotlikoff (1987). These parameters imply a capital-output ratio in 2000 of about 2.7, which is the observed capital-output ratio in the BLS-MFP in 2000.

I calibrate a small set of the parameters describing the nominal side of the economy to values commonly used in the literature. The steady-state value of ξ is set at 8, which implies a steady-state markup over marginal costs of $\xi/(\xi - 1)$ of 14%. Applying the analysis of Keen and Wang (2005), who compare the Calvo price adjustment parameterization across different values of Rotemberg quadratic cost of price adjustment, I set ϕ_p to 100, which in a Calvo pricing model would imply that with the calibrated steady-state markup, about 25% of firms each quarter reset their prices. With ξ calibrated, the choice of ϕ_p implies that the coefficient on marginal costs in the Phillips curve is 0.07, which is consistent with low estimates of the slope of the Phillips curve in the New Keynesian literature (see, for example, Ireland, 2004). The choice of ϕ_p also implies a slope on output in the Phillips curve which is consistent with the estimates of Aruoba et al. (2016).

I assume that the parameters of the monetary policy rule takes standard values close to those estimated by Smets and Wouters (2007), with the persistence of the nominal interest rate set to $\rho_r = 0.65$, the weight on inflation being $\phi_\pi = 2$, the weight on the deviation of output from its steady-state value being $\phi_y = 0.05$, and the weight on output growth being $\phi_g = 0.2$. The inflation target Π^* is set to an annual 2%.

[22] These observed capital-output ratios vary between 2 and 2.7 over the period 1950 to 2013. A full description of the dataset used is given in the appendix.

Table 2: Steady-state over longevity profiles

Quantity	Longevity γ_t^s profiles				
	1940	1970	2000	2030	2070
Capital-output ratio, k/y	2.564	2.597	2.627	2.657	2.686
Consumption-output ratio, c/y	0.724	0.721	0.717	0.714	0.711
Labor force participation rate, h/n	0.619	0.621	0.620	0.618	0.615
Real interest rate, $1 + r - \delta$	1.024	1.023	1.021	1.019	1.018

6.1.6 Steady-state

This section discusses the steady-state effects of changes in life expectancy. Changes in longevity affect steady-state quantities by changing the distribution of the population by age. Table 2 documents what the calibration implies for the capital-output ratio, the consumption-output ratio, the employment-population ratio, and the real interest rate under different mortality profiles observed at different points in the data, assuming that the economy is at a steady-state under those profiles.

As longevity increases with declining mortality rates, aggregate savings increases to fund longer expected retirements. As a result, the capital-output ratio increases, the investment-output ratio declines (so that the consumption-output ratio increases) and the real interest rate falls as the marginal product of capital declines. These forces are familiar in lifecycle models with the demographic changes considered here (see, for example, the discussion in Carvalho et al., 2015).

As people expect to live longer, the employment-population ratio is roughly constant. There are two opposing forces on the employment-population ratio. The first is a compositional effect similar to the analysis in Section 2: as longevity increases, there are simply more older workers. Because of lower productivity and higher disutility of providing labor at older ages, those workers supply fewer hours, so that the overall employment-population ratio declines. Acting against this is the second force: an increase in wages. As retirees accumulate capital, the marginal return to labor increases, incentivizing more labor supply. In the appendix, I consider the solution with exogenous labor supply and find that the participation rate falls more when workers cannot adjust their labor supply in response to wage movements induced by demographics. The equilibrium responses to wages can change the expected labor supply by age.

6.2 Bayesian estimation of the shocks

With the demographic parameters and population dynamics calibrated, I next discuss how Bayesian techniques are used to estimate the model's shocks that drive business cycle fluctuations around the demographic trend. I first discuss the solution method needed to efficiently incorporate demographic trends as anticipated parameter changes and to impose the zero lower bound.

6.2.1 Solution method

The standard linear, time-invariant solution set out by Sims (2002) cannot be used for two reasons: first, because the demographic trends are changes to the wedges attaching to the equations of the model which are fully anticipated by agents in the economy, and second, because the zero lower bound is an occasionally binding, nonlinear constraint. In this case, I adapt a methodology based on anticipated structural changes to the parameters of the model, where the sequences of demographic wedges are taken as an anticipated path of the structural parameters of the model and where the zero lower bound is handled with a regime-switching procedure.[23] The final structure of the economy is the one that arises at the expected completion of the demographic transition and under Taylor-rule policy. Under my calibration, this final demographic structure applies from the year 2070 onwards.

To describe the full time-varying approximation, I first outline the time-invariant approximation of a rational-expectations model of the form $x_t = \Psi(x_{t-1}, \mathbb{E}_t x_{t+1}, \varepsilon_t)$ where x_t is the vector of model variables (state and jump), and ε_t is a vector of exogenous unanticipated shocks whose stochastic properties are known. The rational-expectations approximation of the model, linearized around its steady-state, is written as:

$$\mathbf{A}x_t = \mathbf{C} + \mathbf{B}x_{t-1} + \mathbf{D}\mathbb{E}_t x_{t+1} + \mathbf{F}\varepsilon_t, \tag{19}$$

where $\mathbf{A}$, $\mathbf{B}$, $\mathbf{C}$, $\mathbf{D}$, and $\mathbf{F}$ are matrices that encode the structural equations of the model. A solution to (19), following Binder and Pesaran (1995) or Sims (2002), is written as:

$$x_t = \mathbf{J} + \mathbf{Q}x_{t-1} + \mathbf{G}\varepsilon_t,$$

[23] See also Canova et al. (2015), Kulish and Pagan (2016), and Fernández-Villaverde et al. (2007). More generally, Jones (2015a) discusses how the zero lower bound is a change in the structural parameters of the monetary policy rule that applies for a state-contingent period, or a period that is governed by a forward guidance motive. As is recognized from the work of Christiano et al. (2015) on government spending multipliers, the interpretation of transitory structural shocks can differ when the zero lower bound binds.

where $\mathbf{J}$, $\mathbf{Q}$, and $\mathbf{G}$ are conformable matrices which are functions of $\mathbf{A}$, $\mathbf{B}$, $\mathbf{C}$, $\mathbf{D}$, and $\mathbf{F}$.[24]

In the case where the agents in the model have time-varying beliefs about the evolution of the model's structural parameters, we have: $x_t = \Psi_t(x_{t-1}, \mathbb{E}_t x_{t+1}, \varepsilon_t)$. Denote the corresponding structural matrices for the model linearized at each point in time around the steady-state corresponding to the time t structural parameters by $\mathbf{A}_t$, $\mathbf{B}_t$, $\mathbf{C}_t$, $\mathbf{D}_t$, and $\mathbf{F}_t$.[25] A solution to the problem with time-varying structural matrices exists if agents in the model expect the structural matrices to be fixed at a future point in time at values which are consistent with a time-invariant equilibrium (Kulish and Pagan, 2016).[26] In this case, the solution has a time-varying VAR representation:

$$x_t = \mathbf{J}_t + \mathbf{Q}_t x_{t-1} + \mathbf{G}_t \varepsilon_t, \tag{20}$$

where $\mathbf{J}_t$, $\mathbf{Q}_t$, and $\mathbf{G}_t$ are conformable matrices which are functions of the evolution of beliefs about the time-varying structural matrices $\mathbf{A}_t$, $\mathbf{B}_t$, $\mathbf{C}_t$, $\mathbf{D}_t$, and $\mathbf{F}_t$. The law of motion for the model's state variables at a time period t depends on the full anticipated path of the structural matrices.[27]

6.2.2 The zero lower bound

The occasionally binding zero lower bound constraint is implemented using solution (20) with a regime-switching algorithm, where the two regimes are the zero lower bound regime and a Taylor-rule policy regime (see Jones, 2015a; Guerrieri and Iacoviello, 2015). Agents have rational expectations over which of the two regimes will apply at each point in time. The algorithm iterates on the time periods that the zero lower bound regime applies until it accords with agents' expectations. To obtain the time-varying representation (20) that reflects an expected duration of the zero lower bound at each point in time, the method iterates backwards through the model's structural equations from the system (19) that arises at the expected exit from the zero lower bound regime.

The zero lower bound duration that agents expect is not constrained to be the same duration as

[24] See the appendix for a description of the full time-invariant solution.

[25] One can instead linearize the model around its original steady-state, the steady-state associated with the time-varying system's final structure, or the steady-state implied by the structure at each point in time. Given the somewhat large movements in the steady-state induced by demographic changes, I chose the latter approach, linearizing each set of structural matrices around the steady-state implied by that structure.

[26] Also see Jones (2015a) and Guerrieri and Iacoviello (2015), who apply this procedure to approximating models with occasionally binding constraints quickly and efficiently.

[27] The appendix derives in full the time-varying solution. The methodology is very general and applicable to many other problems of interest. The zero lower bound, forward guidance and changes to the steady-state level of inflation and output growth are anticipated changes to the model's structural parameters which can be handled by solution (20) (Jones, 2015a).

that implied by structural shocks. In this case, the central bank has actively extended the zero lower bound duration through a policy of calendar-based forward guidance. In the estimation, I set these expected zero lower bound durations to those implied by Federal Funds futures data which ensures forward guidance policy over the post-2009 period is taken into account.

6.2.3 Estimation with demographic trends and the zero lower bound

The solution expressed in equation (20) has a state-space representation, so that likelihood methods are straightforward to adapt to estimate the parameters of the transitory shocks in the model, such as Bayesian methods typically applied in the New Keynesian literature (see An and Schorfheide, 2007). The data series used is:

$$\text{Data} = \left\{ \log R_t, \log\left(\frac{\Pi_t}{\Pi^*}\right), \log\frac{y_t}{y_{t-1}}, \log\frac{c_t}{c_{t-1}}, T_t \right\}_t,$$

over the time period 1984Q1 to 2015Q1, and where T_t is the vector of zero lower bound durations that arise each period. When $T_t > 0$, the nominal interest rate $R_t = 1$, so that the observation equation in the state-space representation of the model is time-varying. I use, as time series, the growth rate of output per capita, of consumption per capita, the GDP deflator, the Federal Funds rate, and Morgan Stanley's measure of the length of time until the first rate hike constructed from Federal Funds futures prices. I use a Bayesian likelihood estimation with priors that are common to the New Keynesian estimation literature. I evaluate the convergence of the estimated posterior distributions in the usual way. The full estimation procedure and results are discussed in the appendix.

7 Demographic trends and the business cycle

With the estimated model, I study what drives the gap in output from its pre-crisis trend, how the inclusion of demographic trends affects the interpretation of shocks that drive the business cycle over the post-2009 period, and how they change the probability that the zero lower bound binds during the demographic transition.

7.1 Output since the Great Recession

I extract the structural business cycle shocks and use those shocks and the demographic trends to decompose the gap in output per capita from its long run trend over the period since the Great

Recession. Figure 5 plots the results of the decomposition. Focusing on the most recent period in the sample, 2015Q1, I find that the level of output per capita was about 12% below its 1950–2007 trend. Removing the zero lower bound reduces the gap by only about 2 percentage points. The relatively small contribution of the zero lower bound to the gap in output is because I find that forward guidance was actively used during this period, mitigating much of the contractionary effect of the zero lower bound, as I discuss next. On top of removing the zero lower bound, if the model's nominal markup shocks and frictions on price setting were omitted, the output gap would close by a further 1 percentage point. Next, removing the model's real shocks to productivity and to the discount factor lowers the gap by a further 4 percentage points. The remaining 5 percentage point gap is due to the role of demographics in slowing output growth over the post-recession period. I discuss the factors driving this demographic trend in more detail in Section 7.3.

To illustrate the contribution of forward guidance, in Figure 6, I decompose each quarter's zero lower bound duration into a component due to forward guidance and a residual due to structural shocks and the constraint alone. To compute this decomposition, I calculate the zero lower bound durations implied by the structural shocks, using the algorithm of Jones (2015a). The difference between those computed endogenous durations and the durations used in the estimation is the contribution of forward guidance, or the extension of the zero lower bound regime that, together with the structural shocks, will generate the observed series. The decomposition shows that between 2009Q1 and 2011Q2, the identified forward guidance duration is small, being at most two quarters. From 2011Q3, the forward guidance duration expands notably, increasing to 7 quarters in 2012Q3. These findings are consistent with the results in Swanson and Williams (2014), who find that between 2009 and 2011, longer-term yields were relatively unconstrained, and that after 2011, longer-term yields tightened significantly towards their lower bounds, in line with the Federal Reserve announcing expansive unconventional monetary policies. In particular, around 2011, the Federal Reserve's FOMC announced its "to mid-2013" guidance announcement, the first of many subsequent extensions of the lower bound regime.[28] These results are also consistent with those of Carvalho et al. (2016).

In Figure 7, I compute what the structural shocks would imply for the nominal interest rate in the absence of the zero lower bound and where the Federal Reserve is following its Taylor rule. The

[28] See the FOMC press release, August 9 2011, in which the FOMC announced:

> The Committee currently anticipates that economic conditions–including low rates of resource utilization and a subdued outlook for inflation over the medium run–are likely to warrant exceptionally low levels for the federal funds rate at least through mid-2013.

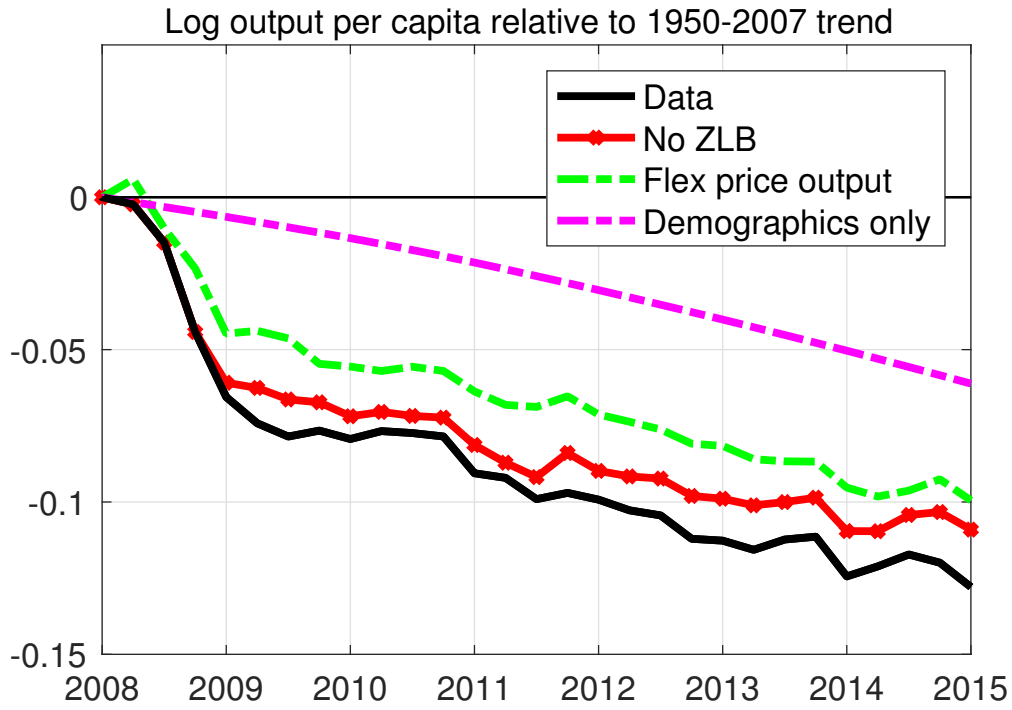

Figure 5: Decomposition of output per capita relative to its pre-crisis trend. Normalized to 2008Q1.

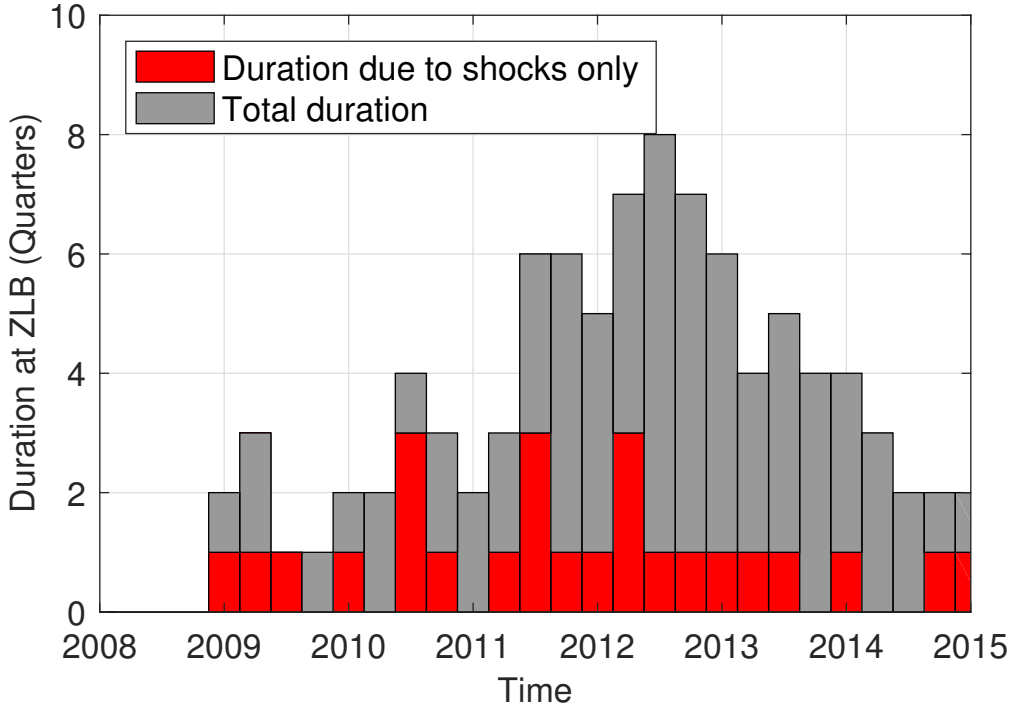

Figure 6: Decomposition of ZLB durations. The difference between the duration due to shocks–the endogenous duration–and the total duration is the identified forward guidance extension.

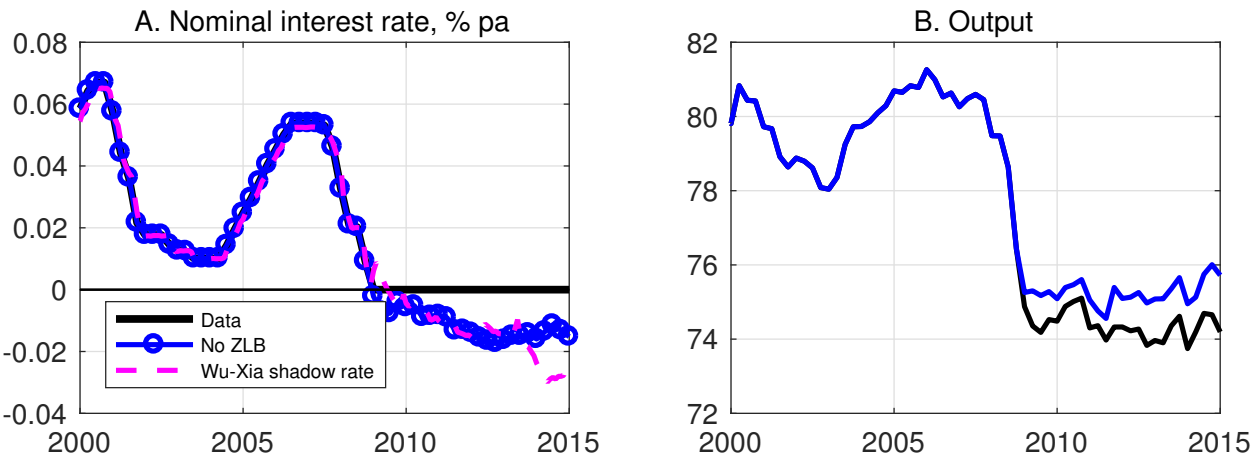

Figure 7: No-ZLB counterfactual.

nominal interest rate in this counterfactual declines gradually to fall to just above -2% in 2013, corresponding to the longest zero lower bound duration implied by the Federal Funds futures data. I compare, in Panel A, this counterfactual interest rate to the shadow rate computed by Wu and Xia (2016). Their shadow Federal Funds rate is estimated using a nonlinear model of the term structure and is intended to measure the stance of monetary policy during the zero lower bound period. As the figure shows, their measure is very close to the one implied by my estimated model, deviating only around 2014. Panel B shows that the level of output is higher in the counterfactual without the zero lower bound, but not by much.

Wu and Xia (2016) also show that the correlations between a number of macroeconomic variables and this shadow interest rate are similar to the correlations between the Federal Funds rate and those macroeconomic variables in the pre-2009 period. Accounting for Federal Reserve policy through forward guidance as I model it can therefore rationalize why shocks which impact during a zero lower bound episode generate, empirically, the same correlations as those observed before 2009. This stands in contrast to theoretical predictions of shocks that impact at the zero lower bound–for example, for the effect of supply shocks, see Wieland (2014), and for the effect of expansionary fiscal policy, see Christiano et al. (2011).

To illustrate the quantitative role of forward guidance, I compute a generalized impulse response to a contraction in the forward guidance duration from 3 quarters to zero quarters in 2011Q3, the period of the Federal Reserve's "to mid-2013" forward guidance announcement. I compute the generalized impulse response because of the dependence of the change in the duration on the history of shocks leading up to the forward guidance announcement (see Kulish and Pagan, 2016). The

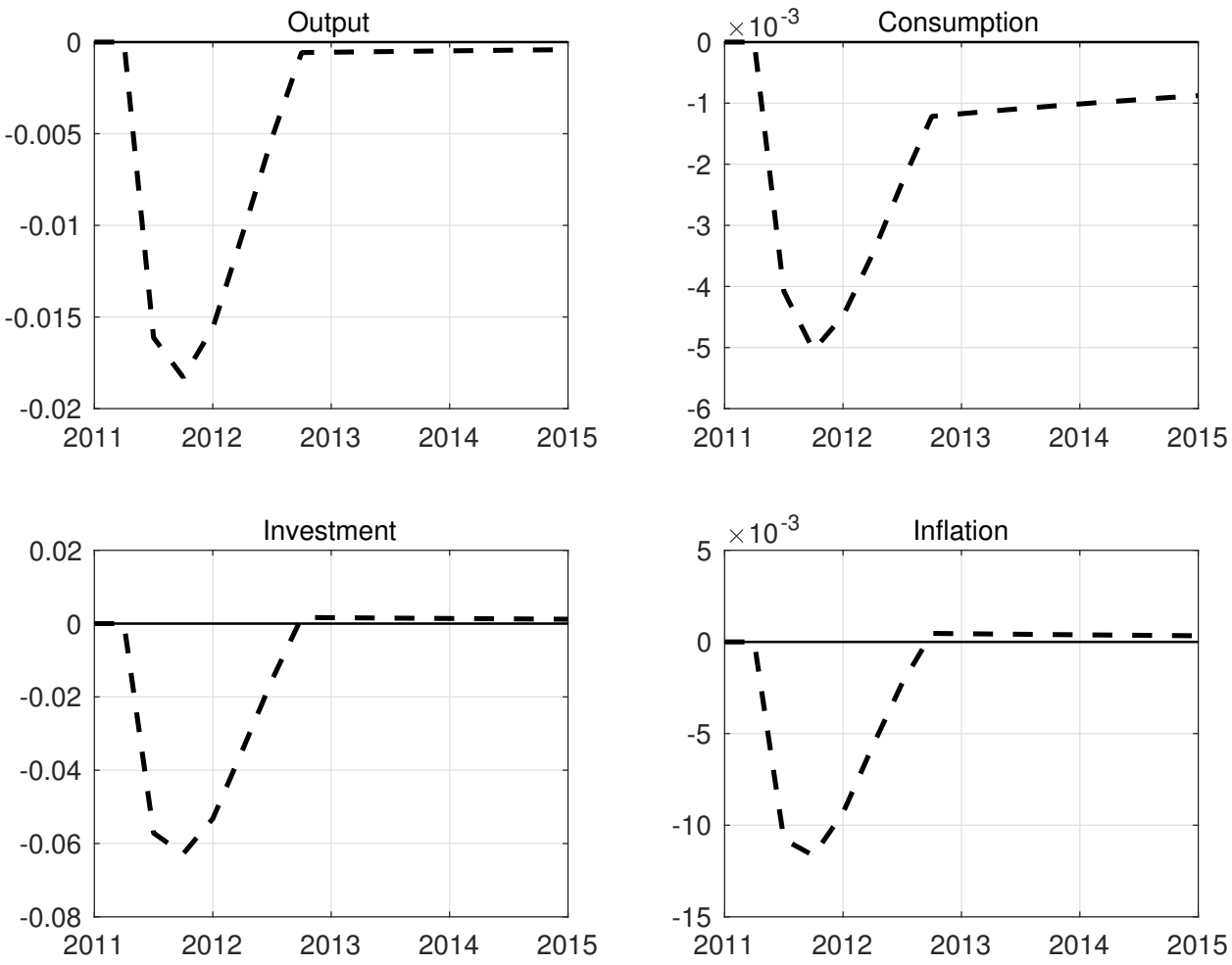

Figure 8: Generalized impulse response function. In this exercise, the forward guidance duration is eliminated in 2011Q3, so that the total expected duration falls from 6Q to 3Q.

3 quarter forward guidance duration was the size of the announcement identified by my approach as illustrated in Figure 6. After 2011Q3, I turn off the structural shocks and plot in Figure 8 the difference between the path computed under the forward guidance announcement and the path without forward guidance. The impulse response shows that the extension of the zero lower bound by 3 quarters longer than what would be implied by the monetary policy rule caused quarterly output to be about 1.75 percentage points higher and quarterly inflation to be about 1 percentage point higher. The magnitude of this effect on inflation and output is roughly equivalent to a three standard deviation innovation to monetary policy, which is a reasonable magnitude in light of the forward guidance puzzle (McKay et al., 2016).

Next, I compute a counterfactual asking what would happen if the demographic state was maintained at its 1984 level. This allows me to examine the extent to which demographics has

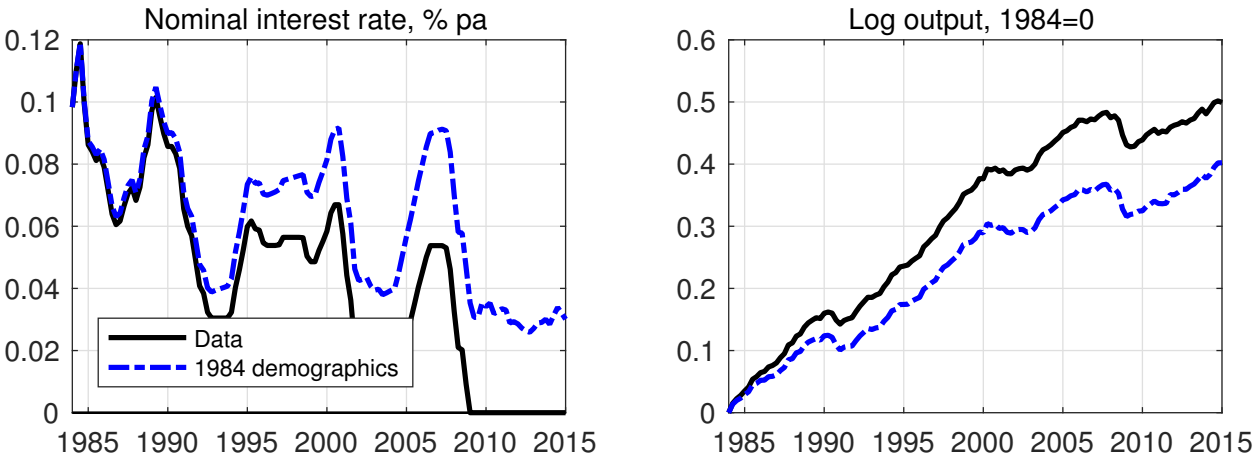

Figure 9: Counterfactual series. I add the mean of the output growth series to generate the output indexes. The 1984 demographic state is the one associated with the demographic adjustment parameters observed in 1984.

constrained the nominal interest rate since the Great Recession, and to isolate the effect of the zero lower bound in the post-recession path of output. This counterfactual is presented in Figure 9. I find that the *level* of output would have been lower over the full period, as the size of the workforce, higher levels of physical capital and lower rates of human capital growth lowers output relative to what was observed. In addition, the counterfactual series predicts a fairly rapid recovery in the level of output, essentially eliminating the stagnation relative to a pre-crisis trend. Under this counterfactual scenario, the level of output recovers its 2008Q1 counterfactual level in 2013Q1, as compared to 2014Q2, illustrating that the recession would still have been long-lived, but that the recovery would have been faster. These results accord with the predictions of the model for the trend path of output, discussed in Section 6, where the aging trends predict sluggish growth post-2010 when compared to the 1990s/2000s.

7.2 Simulations of the economy during the demographic transition

To what extent do demographic changes make the zero lower bound a more binding constraint? I examine this interaction between demographic trends and the zero lower bound by simulating the estimated New Keynesian model repeatedly and studying the distribution of paths of output and the nominal interest rate from those simulations. The distribution of the simulated paths of the real interest rate, the nominal interest rate, output, and consumption are presented in Figure 10. I augment the path of the nominal interest rate with expected inflation from the University of Michigan's consumer inflation expectations survey, available from 1978 to 2015, to proxy for changes

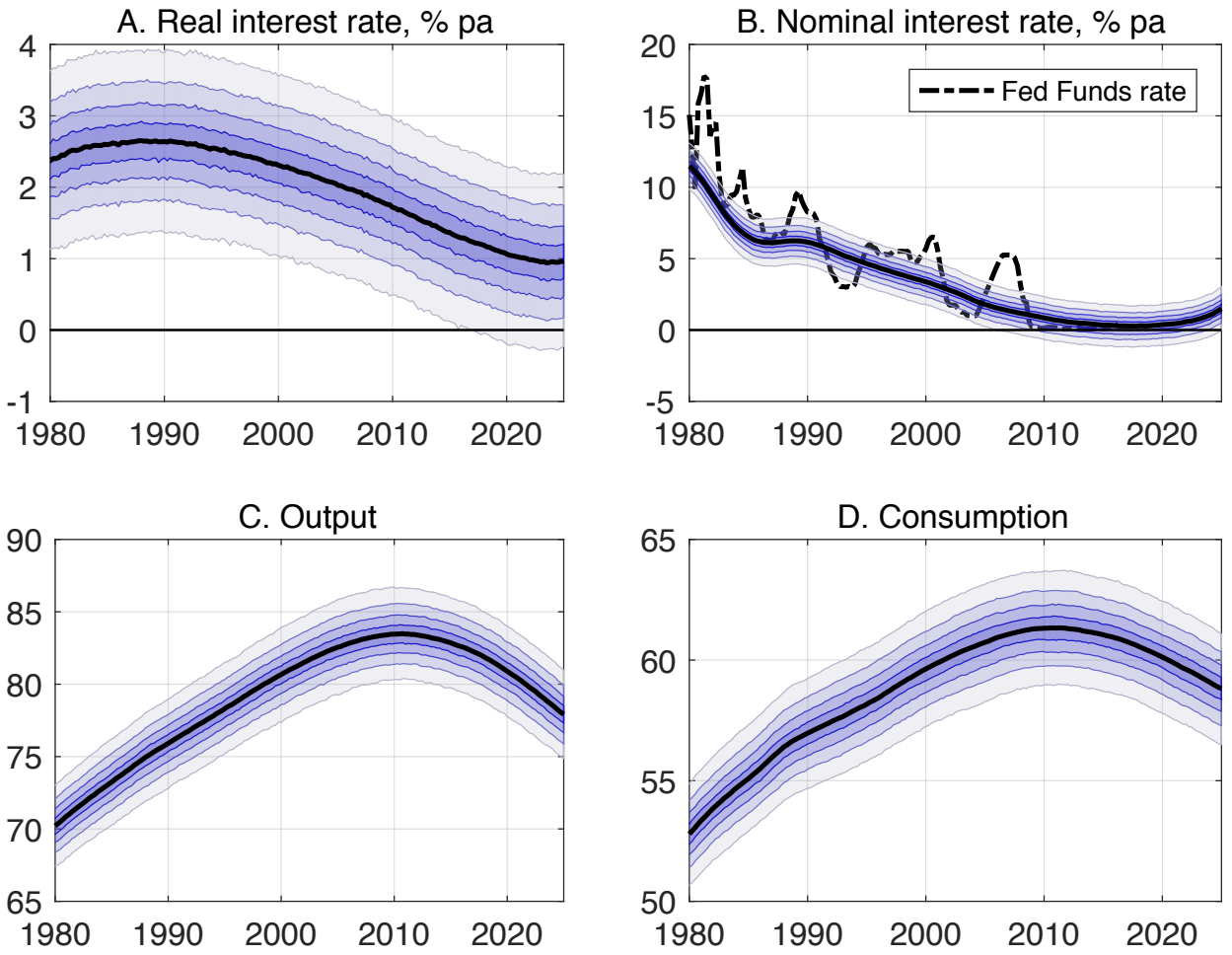

Figure 10: Stochastic simulations of the business cycle model during the demographic transition. This figure shows fancharts of simulated paths around the demographic trends. I augment the nominal interest rate with the University of Michigan's consumer inflation expectations series from the start of that series 1978 to 2004, after which I set inflation expectations to be the Federal Reserve's implicit 2% target for inflation.

in the Federal Reserve's implicit inflation target. From 2004 onwards, I set inflation expectations to a constant annual rate of 2%, consistent with the Federal Reserve's implicit inflation target during this period (see Ireland, 2007), and which was made explicit in 2012.

The distribution of simulated paths of the real interest rate show how demographic trends cause a strong trend decline from highs in the 1980s and 1990s. That trend pushes real interest rates low enough so that, in some simulations, business cycle shocks push the real interest rate negative. Real interest rates as measured from ex-post Treasury interest rates and inflation have been negative at points following the financial crisis in 2009, and have been a feature of many studies of the liquidity trap (see, for example, Eggertsson and Mehrotra, 2014).

The simulated paths of the nominal interest rate show that the trend decline in the real interest rate generates a trend decline in the nominal interest rate, to the extent that it drives about 30% to 40% of the density of nominal interest rate paths below zero between 2010 and 2025. By contrast, the zero lower bound was not visited in any simulation before 2000. This illustrates the important nonlinear interactions between demographics and the zero lower bound that motivates a model with both factors.

7.3 The role of demographics

In this section, I discuss the model's trend predictions under the demographic transition to understand why demographics is the primary cause of the decline in output relative to trend. I discuss the key results on trends in the real interest rate, total output and productivity growth, and in the employment-population ratio. For this, I use the full nonlinear solution of the overlapping generations framework where the demographic trends are perfectly foreseen by agents in the model.[29]

7.3.1 Demographic variables

I first present in Figure 11 summary measures of demographics from the model. Panel A plots life expectancy when individuals enter the model, Panel B plots the median age of those in the model, and Panel C presents the fraction of the population in the model at the entering age of 16. Life expectancy conditional on surviving to age 16 during the transition increases from about 77 years in 1950 to about 85 years in 2020. This profile matches exactly that observed in the data, because it is computed off observed conditional survival probabilities. The median age of those 16 and above from the model tracks well the median age observed over the transition. The slight difference between the two profiles is due to the initial population distribution and other population changes not fully captured in the model, such as immigration. In the model, the initial population distribution reflects the steady-state associated with the mortality distribution of those born in 1940.

7.3.2 Marginal product of capital and the real interest rate

The demographic trends imply a steadily declining path for the real interest rate in the model. The decline can be decomposed into a component due to increased longevity, and one due to changes in the composition of the workforce. This decomposition shows that the increase in longevity causes an

[29] The nonlinear solution is described in the Appendix.

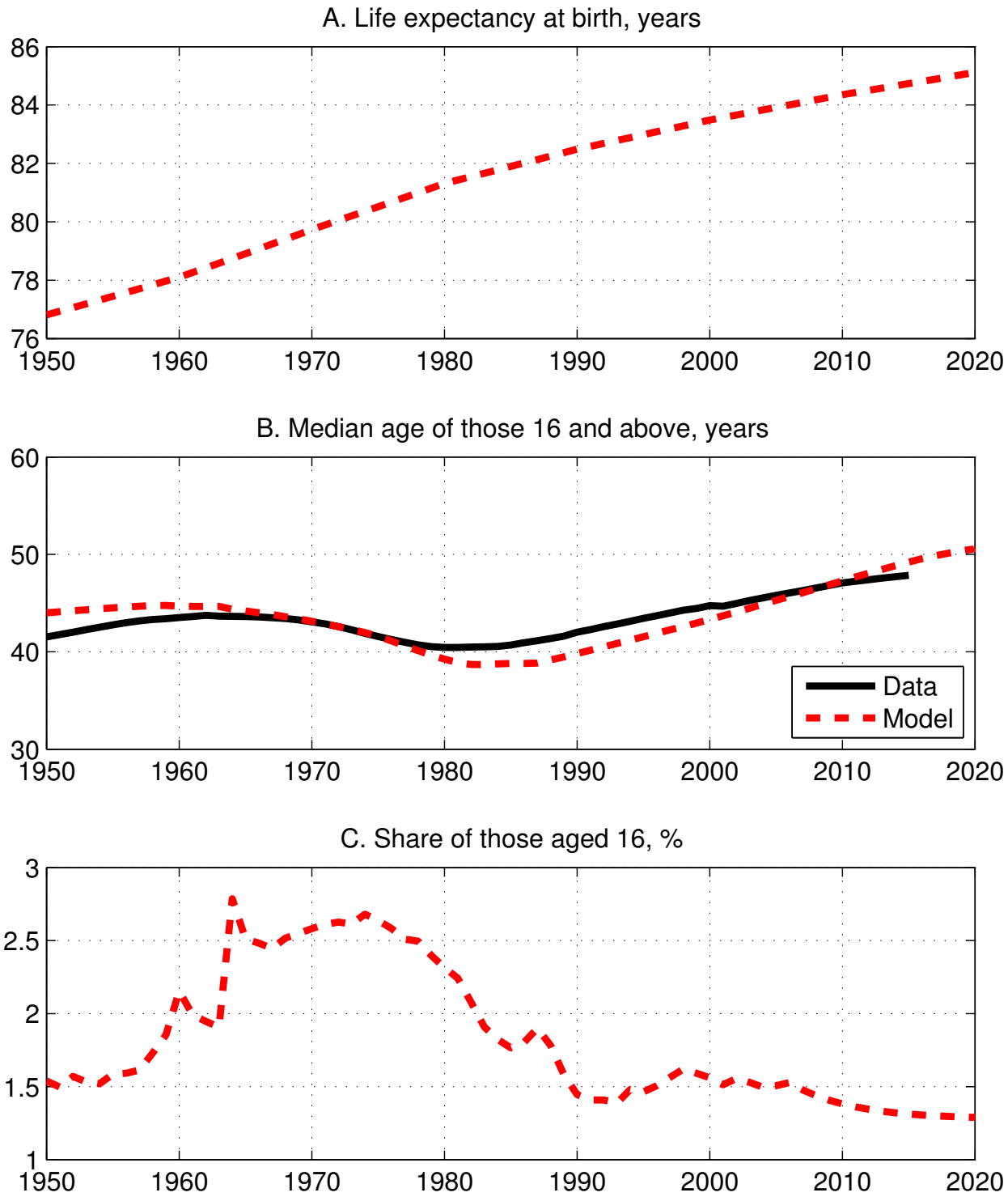

Figure 11: Demographic trends in model. This figure shows three summary measures the exogenous demographic variables. Life expectancy conditional on being 16 rises by about 10 years while the median age of those 16 and above decreases from 1960 to 1980 and rapidly increases thereafter. The share of those age 16 is used to calibrate the initial population each period.

approximately 0.6 percentage point decline in the real interest rate. This decrease is driven by the increase in savings used to fund consumption during longer expected retirement periods.

More notably, the aging of the baby boomer cohorts generates a large oscillation around the path implied by increasing longevity, falling from peak-to-trough between 1985 and 2015 by just over 2 percentage points. This pattern matches very closely, over the same period, the decline in the real interest rate computed from the observed capital-output ratio provided by the Bureau of Labor Studies Multifactor Productivity Program.[30] The oscillation is driven by changes in the relative size of the workforce. The workforce is relatively young as the baby-boomers enter the labor market in the 1960s to 1980s, so that aggregate hours supplied is high relative to capital, thereby increasing the marginal return to capital. As the baby-boomer cohort ages and accumulate savings for retirement, the marginal return to capital and the real interest rate decline. This decline is then reinforced by the withdrawal of the baby-boomer cohort from the labor market, rapidly decreasing the marginal return to capital and staying low beyond 2030.

7.3.3 Growth rates of aggregate quantities

What does the model say about how demographic trends affect the growth rates of aggregate quantities and productivity growth? Panel A of Figure 13 presents the growth rates of total output, output per capita and output per worker. Total output growth due to demographics peaks at 1 percentage point just before 1980. The growth rate then steadily declines, until demographics becomes a drag on total output growth, which occurs in 2012. Demographic changes are then a substantial drag on total output growth, with the contribution to overall growth from demographics staying negative throughout the forecast horizon to 2070. In total, output growth declines from peak-to-trough by about 1.5 percentage points. From 1980 to 2015, I find that demographics predicts that output growth declines by about 1.25 percentage points, which is the same decline reported by Gagnon et al. (2016) in their study of the macroeconomic consequences of demographics.

Output per capita and output per worker—labor productivity—growth rates show a very different pattern to the total growth rate. This difference is due to labor supply, in particular, the entrance of the baby boomer generation into the workforce in the 1960s. Per worker and per capita output growth due to demographic changes is initially negative between 1960 and 1980, before becoming positive between 1980 and 2010. From then on, demographics causes per capita output growth to

[30] In computing this observed series, I use the marginal product of capital and parameters of the model.

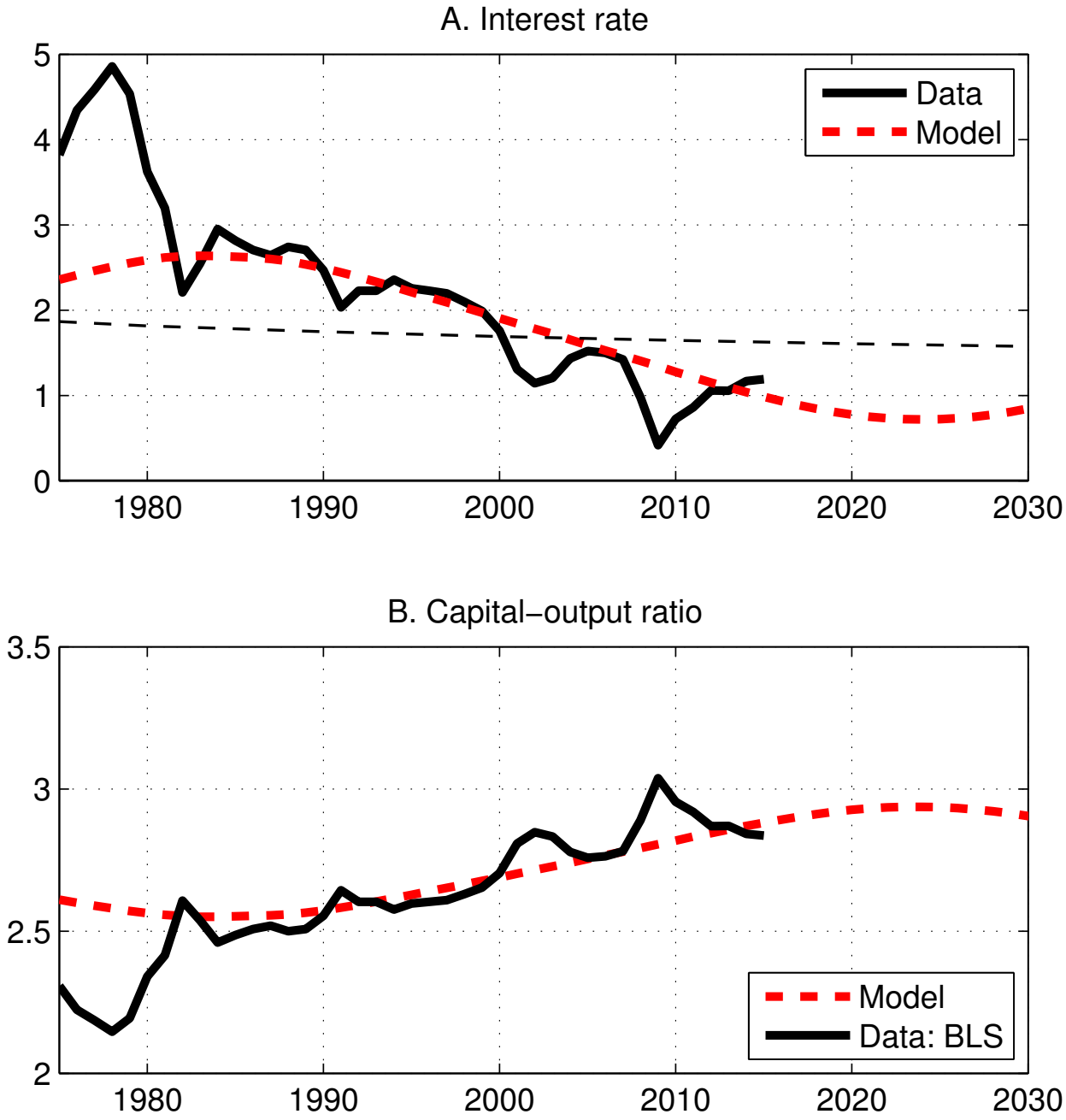

Figure 12: Capital-output ratio and real interest rate. Panel A shows changes in the output-capital ratio in the model against the ratio of measure from the Multifactor Productivity Program. Panel B plots the model's interest rate in the full demographic transition. Peak-to-trough between 1985 and 2010, the interest rate falls about 1.5% points.

Table 3: Average annual growth rates

	Output		Capital		Consumption		y/h		y/n	
Period	Model	Data	Model	Data	Model	Data	Model	Data	Model	Data
A. Model Against Raw Data Averages										
1990 to 1999	3.46	3.70	3.97	3.88	3.30	3.43	2.76	2.19	2.23	1.98
2000 to 2009	3.19	1.73	3.74	3.23	3.13	2.40	2.41	2.52	1.93	0.87
2010 to 2015	2.75	2.84	3.28	1.64	2.81	2.22	2.30	1.15	1.66	1.30
B. Model Against HP-filtered Data Averages										
1990 to 1999	3.46	3.73	3.97	3.97	3.30	3.57	2.76	2.24	2.23	2.14
2000 to 2009	3.19	2.48	3.74	3.19	3.13	2.70	2.41	2.49	1.93	1.31
2010 to 2015	2.75	2.04	3.28	1.75	2.81	1.87	2.30	1.06	1.66	0.82

C. Model Forecasts

Period	Output	Capital	Consumption	y/h	y/n
2016 to 2025	2.31	2.58	2.58	2.17	1.33
2026 to 2035	2.13	1.91	2.42	2.10	1.32
2036 to 2045	2.28	2.01	2.32	2.11	1.70
2046 to 2055	2.44	2.40	2.36	2.20	1.91
2056 to 2065	2.67	2.75	2.62	2.29	1.94

Notes: y/h is output per hour worked. y/n is output per person. To make the growth rates in the model and data comparable, I add to the model growth rate the amount that is needed to equalize the average growth rates in the model and data, computed over the years 1948 to 2015.

turn negative until at least 2040, while per worker output growth stays slightly negative over the forecast horizon. In total, per capita output growth declines from peak-to-trough by just over 1 percentage point between 1990 and 2025, while per worker output growth declines by about 0.7 percentage points over the same period. The magnitude of the decline in per capita growth accords with the results in Antolin-Diaz et al. (2014).

Table 3 compares the model's predictions for growth rates against those in the data. To make the model and data growth rates comparable, I add to the model's growth rate an amount that ensures that the average growth rate from 1948 to 2015 is the same in the model and in the data. The model performs well in explaining the decline in average output growth across the 1990-1999 period and the 2010-2015 period, with the decline in average growth in the model being 0.71% compared to 0.86% in the data. The model generates about 30% of the decline in average capital growth and 40% of the decline in average consumption growth over the same period, but forecasts a steep decline in the growth rate of capital to the average levels observed between 2010-2015 by the mid-2020s.

Turning to per-person measures of output, the model explains almost all of the average decline in per capita output growth between 1990-1999 and 2010-2015, and explains 0.5 percentage points

of the observed 1 percentage point decline in the growth of output per hour. Looking forward, the demographic trends imply a further decline in productivity of another 0.2 percentage points, and 0.3 percentage points in per capita output growth, on average between 2026-2035.

What factors are driving these declines in average growth rates? In the model, there are three ways that these measures of output growth can change over time. Individuals can supply more hours, affecting both output and aggregate labor. There are also changes in physical capital, as individuals save and consume out of accumulated savings in retirement. Consumption smoothing motives ensure that the level of savings changes at a different rate to the supply of labor. Third, the *quality* of labor can change. In particular, changes in the distribution of workers resulting from demographic changes alters the average skill-level of the workforce, which shows up in the productivity decomposition as fluctuations in the quality of labor.

Formally, I decompose the model's predictions for output growth and labor productivity growth into their component parts following a standard growth accounting exercise (Hall and Jones, 1999; Fernald, 2015). The production function in the model is $y_t = k_t^{\alpha} \ell_t^{1-\alpha}$, where ℓ_t is aggregate efficiency units of labor. The total derivative of the production function decomposes the change in output into: $\frac{\mathrm{d}y_t}{y_t} = \alpha \frac{\mathrm{d}k_t}{k_t} + (1-\alpha)\frac{\mathrm{d}\ell_t}{\ell_t}$. In the lifecycle model, growth in output per efficiency unit of labor ℓ_t arises from changes in aggregate labor supply or from changes in the labor quality of the workforce, as individual workers become more productive with age: $\frac{\mathrm{d}\ell_t}{\ell_t} = \frac{\mathrm{d}h_t}{h_t} + \frac{\mathrm{dLQ}_t}{\mathrm{LQ}_t}$, where h_t is aggregate hours and LQ_t is labor quality.

Panel B of Figure 13 plots the decomposition for labor productivity growth, while Panel C plots the decomposition for total growth. Accelerating capital accumulation increases the growth rate of both labor productivity and total output up to 1995, after which the growth rate starts to decline. The change in labor supply has a large negative effect on productivity growth, but a positive effect on total growth, when the baby boomer cohorts enter the labor force around 1960.

A key component of both labor productivity and total growth is the change in the quality of the workforce which arises as the composition of the workforce interacts with the age-productivity profile. The decomposition implies that the contribution of the change in average labor quality to the growth rate of output and output per worker peaks around 1990, adding roughly 0.35 percentage points to total growth and productivity growth. The contribution of labor quality turns negative in 2000 as the mass of workers reaches the peak of the age-productivity profile, exhausting the potential for further growth in average human capital across the workforce. This force remains a drag on

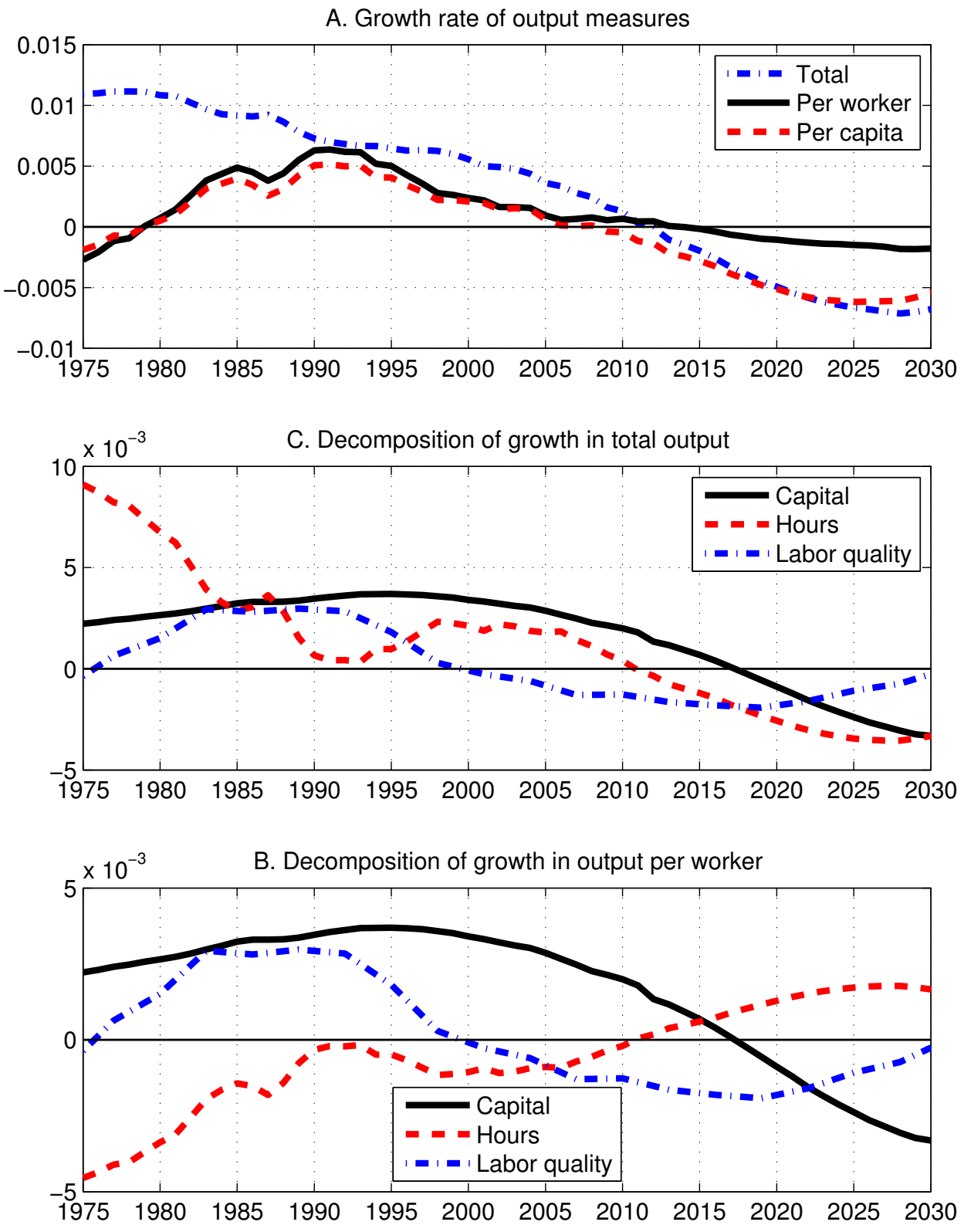

Figure 13: Measures of output growth and their decomposition. This figure plots different measures of output growth and the decomposition of output per worker and total output into changes in factors of production.

productivity growth until 2030.

Based on the model's predictions for how growth rates change with demographics, Figure 14 plots the indexes of output, investment, and consumption in the data against their model counterparts, normalized to their 1990 values. I use the same procedure as for Table 3 in adding the trend growth to the model series, and compare the resulting indexes to an index computed of each series' 1992 to 2007 pre-recession trend.[31] The trends that the model generates show clearly the decelerating profiles over time as compared to the rapid growth rates experienced over the 1992 to 2007 period. Comparing these trends to the data, the plots of the indexes show that the model does a good job in explaining the gap between the observed measures of output, investment and consumption relative to their pre-crisis trends.

A surprisingly strong feature of the data is the divergence between the levels of output per hour worked and output per capita. Demographic changes through the model can generate a gap between these variables because of different profiles for aggregate capital and aggregate hours. As workers withdraw from the labor market when they reach retirement age, their relative supply of capital to firms remains high, keeping total output high. As the workforce contracts, measured output per worker increases relative to the same amount of output per person alive. Panel D of Figure 14 plots the ratio of total hours to the population, normalized to 1990, and shows that the model's predictions match the post-1990 trend in the data well.

7.3.4 Employment-population ratio

I next turn to the predictions of the model for the employment-population ratio. Figure B.1 plots the share of the labor endowment that a member of each cohort chooses over time against participation rates observed in decennial censuses from 1960 to 2000, and annual American Community Surveys thereafter. A combination of the age-productivity profile and the calibration of the disutility of labor parameter at each age implies labor force participation rates that are humped shaped over age. The model is not parameterized to match the trend increase in female labor force participation over the years 1950 to 1990, accounting for its inability to match the substantial increase from 1950 to 1990.[32] The fluctuations in labor supply at each age are driven by equilibrium wages. These changes generate changes in labor force participation rates for older workers that track those observed.

[31] This is the same time period that Blanchard et al. (2015) use in studying the slowdown in output growth from a pre-crisis path.

[32] After 1990, the female labor force participation rate was rougly constant, and has been trending down since 2000.

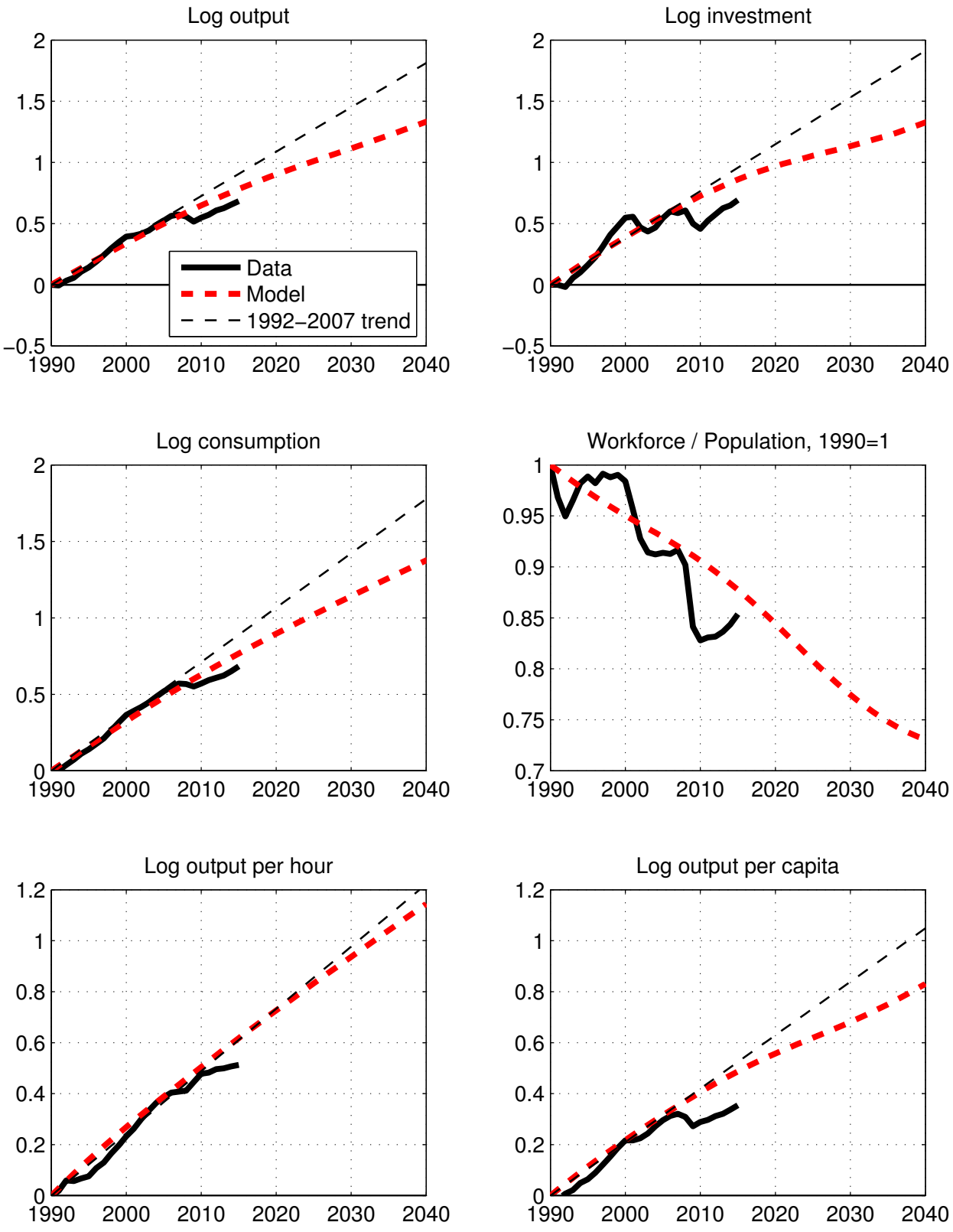

Figure 14: Demographic trends. This figure plots the model's trend paths for some key variables against the observed indexes and a pre-crisis trend path, compared on the average growth rate of each variable from 1992 to 2007.

Figure B.2 plots the aggregate employment-population ratio in the model against the data in Panel A, showing that under the calibration which was chosen to generate age-specific labor force participation rates that are consistent with those observed, the labor force participation rate declines in the model at the pace that is roughly as observed, and is predicted to continue to fall by about five percentage points from 2015 to 2030. This result is consistent with the results in Section 2 and other reduced-form studies that assign much of the recent decline to demographic factors and, in particular, the age-compositional change in the population.

7.3.5 Robustness and alternative calibrations

I do a number of experiments to verify that the baseline results are robust to changes in the calibration. The full results of each experiment are presented in the appendix but discussed briefly here. The first check is to see that the model's predictions hold when individuals face a constraint restricting their borrowing early in life. With borrowing constraints, the extent of savings is inflated, pushing up the capital-output ratio. As a consequence, the real interest rate is lower than in the baseline model. The magnitude of the fluctuations of the real interest rate, the participation rate and output growth are very similar to the baseline model.

The remainder of the robustness checks focus on differences in the age-productivity profile. The second robustness check is to incorporate time-varying productivity profiles. The evidence presented in Kong et al. (2016) suggests that the age-productivity profiles have flattened over time. Such a flattening can affect the accumulation of human capital and can affect aggregate productivity measures in two ways: first, with a growth effect by lowering the potential for new workers to accumulate human capital, and second, with a level effect by affecting the productivity level that individuals enter the workforce on. I calibrate the age-productivity profiles by recomputing for each cross-sectional sample, the profile and then interpolating between those points in time. The overall pattern of aggregate labor productivity is much the same as the baseline model, although the magnitude of the amplitude of the change in labor productivity growth is smaller, with demographics contributing the most to labor productivity growth in 1980 and not 1990 as in the baseline results.

From 1985 on, the baseline predictions for the participation rate, aggregate labor productivity growth and the real interest rate are largely unaffected when the age-productivity and labor disutility profiles are calibrated to match female age-earnings profiles and female labor force participation rates from the 1940s to 1990s, after which female labor force participation is roughly constant. As a

final point of comparison, I verify that the directions of the aggregate predictions are robust to a calibration where an additional source of heterogeneity is modeled—where there are two types of workers, low or high skilled, where low skill workers are those with less than college education.

8 Conclusion

This paper studies why, in 2015, the level of US output per capita was 12.5% below its pre-crisis trend. I used a New Keynesian model with demographic trends and the zero lower bound to show that declining mortality rates and changes to the population share of the young can generate trends that match the low frequency movement of output growth, productivity, the real interest rate, and the employment-population ratio.

I estimated the transitory shocks of the model using Bayesian techniques with a sample including the zero lower bound period and under the demographic transition. With the estimated model, I studied stochastic simulations and find strong evidence of time-dependence in the probability that the zero lower bound will bind, with the zero lower bound much more likely to bind between 2010 and 2025. I then extracted the set of structural shocks that drive the business cycle over the 1984 to 2015 period taking into account demographic trends. With those policy-invariant shocks, I found that the aging of the population is responsible for the bulk of the decline in output per capita relative to its pre-crisis trend—about half. By contrast, removing the zero lower bound explains about 20% of the gap between output and its long-run trend.

An important assumption that is made in constructing the model is that the age-productivity profile is held constant over time and over the demographic transition. My model is silent on the extent to which demographic changes might cause the returns to labor to change over time. For instance, it would be a promising avenue of research to study the effects of demographic changes in an environment where there are complementarities between capital and labor.

The results clearly show the importance of demographic trends in explaining macroeconomic trends over time. Additional research on the role of demographics could include formulating a model of housing in an environment where demographic trends are modeled as in this paper. An interesting question would be to ask whether demographics can justify the housing price dynamics observed in the run up to the 2008 financial crisis. The results of this paper motivate this study, as I find that the consumption demand and productivity shocks derived for the aggregate representation peak in

different periods, with consumption demand peaking around 2020, roughly 10 years after the peak in the aggregate productivity shock.

The New Keynesian framework incorporating demographic changes opens up a number of promising avenues for future research. To what extent do demographic trends explain the Great Moderation? How does the demographic state affect the size of the response to transitory shocks and does this differ with the source of the shock? These are questions that can be answered readily using the methodology outlined in this paper. More generally, the methodology shows how long-run trends can be incorporated into estimated business cycle models in a way that captures empirically important nonlinear interactions between the trend, shocks, and the zero lower bound.

Appendix

A Algorithm to solve path of OLG model

I use a perfect foresight, deterministic shooting algorithm to solve for the path of the OLG model under the exogenous demographic forces. I specify the year 2200, well beyond the year 2070 that exogenous demographics stay unchanged, as the period for which the economy is assumed to return to the steady-state associated with the final demographic structure. Computationally, the system is solved by taking the set of model equations at each point in time, and stacking them. Repeated Newton-type iterations are then done on the stacked system. A step in the Newton method is to compute the Jacobian of the full system. The shape of the Jacobian is triangular, and relaxation and block decomposition methods solve the problem efficiently.[33]

B Lifecycle model calibration and robustness

B.1 Disutility of labor calibration

As in Kulish et al. (2010), the equation for the disutility of labor supply is:

$$v^s = \kappa_0 + \left(\kappa_1 \frac{s}{70}\right) \int_{-\infty}^{s} \frac{1}{70\sqrt{2\pi}\kappa_3} \exp\left(-\frac{1}{2}\left[\frac{x - 70\kappa_2}{70\kappa_3}\right]^2\right) \mathrm{d}x.$$

The function is a scaled version of the cumulative density function of a normal distribution. There are four parameters governing v^s: (i) κ_0 implies a baseline level of disutility from labor, (ii) κ_1 scales the entire disutility function, (iii) κ_2 scales the mean of the distribution, or the age at which the disutility from work is increasing the most, and (iv) κ_3 scales the standard deviation of the function, or the slope for which the disutility increases.

B.2 Labor force participation by age

In Figure B.1, I plot the labor force participation rate by age from censuses and the American Community Surveys against their predictions under the demographic changes. The model broadly matches the values of the LFP by age and does a decent job at matching the dynamics of the labor force participation rate, particularly for those workers on the edge of retirement.

B.3 Borrowing constraints

In this robustness exercise, I set $a_t^s \geq 0$ for all t, s. The effect of this change is to cause the young to supply less labor for the periods when borrowing is constrained. The consumption profile is steep for those periods the young are constrained. In the aggregate, with the other calibrated parameters kept constant, there is more aggregate savings and less aggregate labor supplied, resulting in a higher capital-output ratio and lower real interest rate. The movements in the interest rate and labor force participation rate are largely unaffected as compared to the baseline results. In Figure B.3 it is labeled as the A series.

B.4 Time-varying productivity profiles

In this exercise, I calibrate the productivity profiles to be time-varying and fully anticipated, so that there are anticipated changes to the slope and magnitude of the productivity profiles faced

[33] These methods are implemented in Dynare.

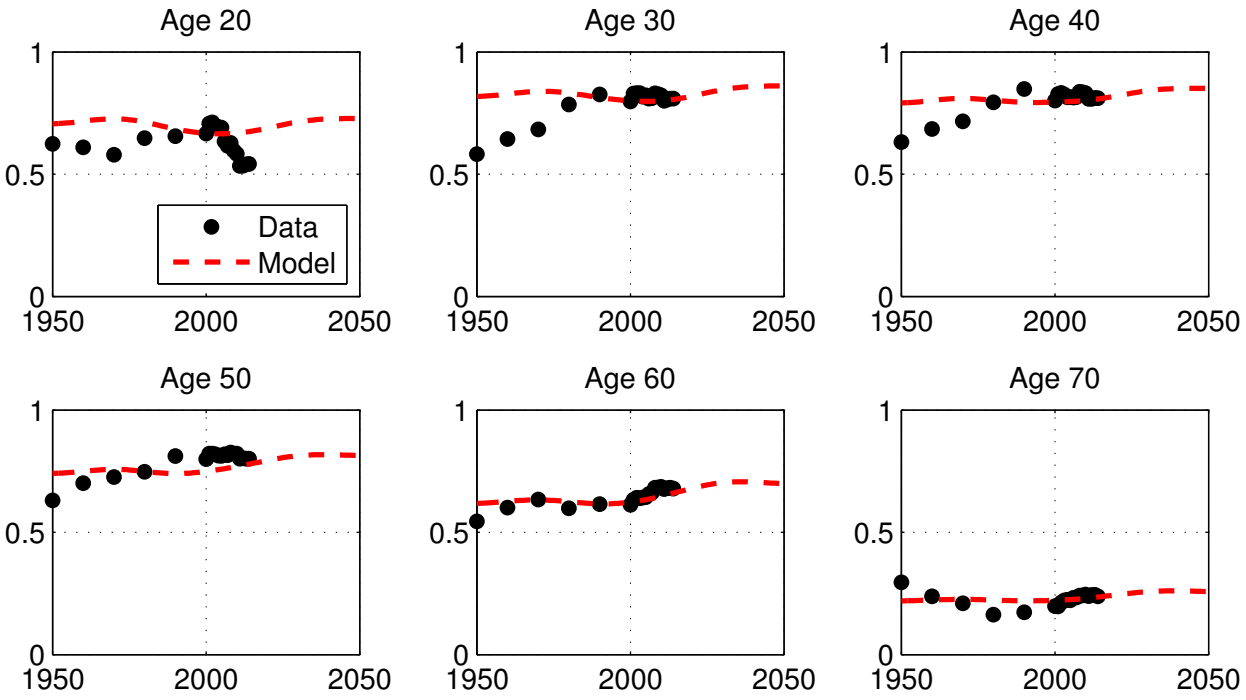

Figure B.1: Labor force participation by age. This figure shows the fraction of the labor endowment chosen by workers at each age in the model against the labor force participation rates observed in censuses and American Community Surveys.

by workers. I recalculate the age-productivity profile by recomputing the Census/ACS age-earning profile for full-time workers and rescaling the profile to those entering the workforce. Earnings are deflated by the GDP deflator. In Figure B.3 it is labeled as the B series.

B.5 Gender-based calibration

In this exercise, I divide workers into genders, recalibrating the age-productivity profile to male workers and female workers separately. I also choose the age-disutilities separately so that female labor force participation increases since the 1950s, pushing up the overall increase in the employment-population ratio. In Figure B.3 it is labeled as the C series.

B.6 Skill-based calibration

In this exercise, I divide workers into two skills, recalibrating the age-productivity profile to those of less than college-education and those with at least some college education. In Figure B.3 it is labeled as the D series.

B.7 Exogenous labor supply

In this exercise, households have no disutility of supplying labor, and are forced to enter retirement full-time at age 65. This exercise changes substantially the profiles for aggregate labor force participation and for the real interest rate.

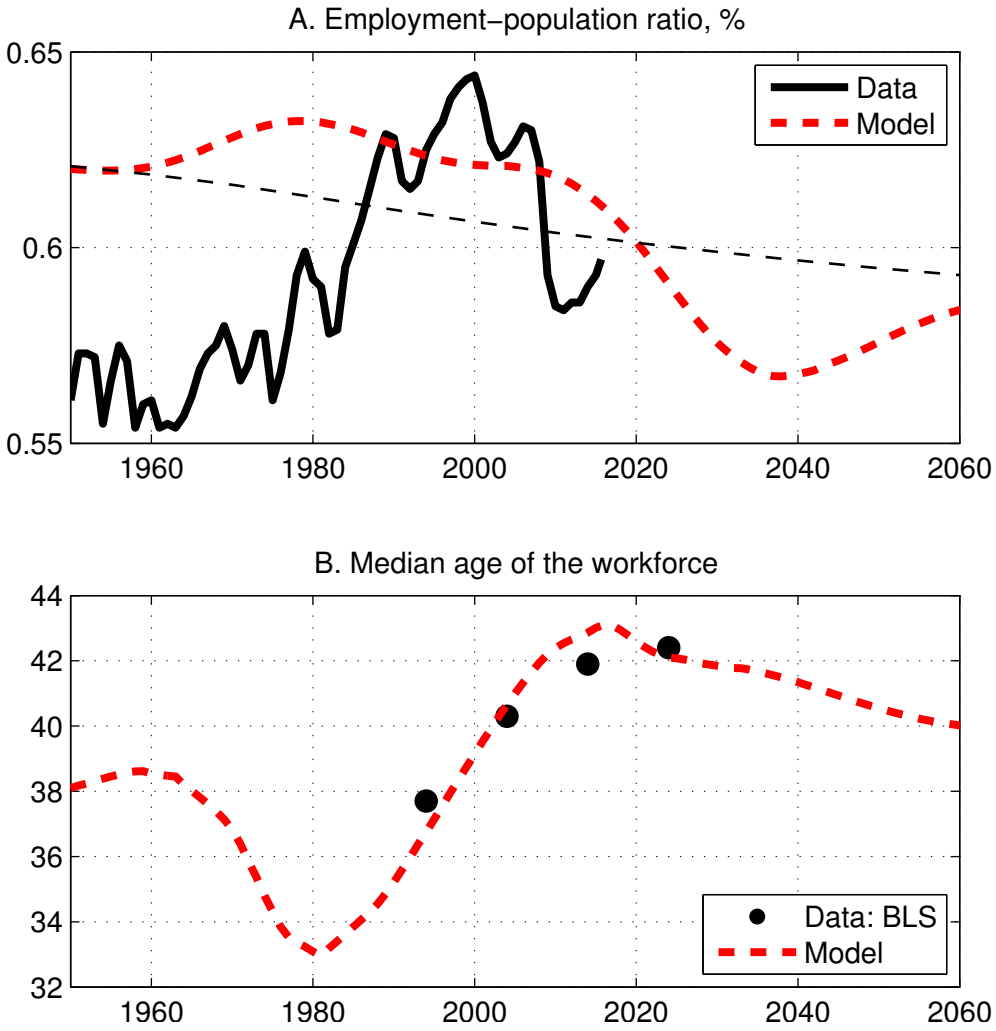

Figure B.2: Employment-population ratio. This figure shows total employment-population ratios in the model and as observed (Panel A), and the median age of the workforce (Panel B).

C Demographic adjustment trends in the lifecycle model

The measure of agents of age s in period t is n_t^s.[34] Take an individual j belonging to the cohort born in period s. Rewrite her lifecycle problem as an infinite horizon problem with a preference process $\phi_t^{j,s}$ that proxies for the lifecycle. Her preferences are therefore:

$$\max_{\{c_t^{j,s},\ \ell_t^{j,s},\ a_t^{j,s}\}} \sum_{\tau=s}^{\infty} \beta^{\tau} \left[\prod_{r=s}^{\tau}(1-\gamma_r^s)\right] \phi_\tau^{j,s} u(c_\tau^{j,s}, \ell_\tau^{j,s}). \tag{21}$$

The individual's sequence of budget constraints is as in Section 3. Let $\lambda_t^{j,s}$ be the marginal utility of wealth, or the Lagrange multiplier on the individual's budget constraint. Because of the redistribution of unintentional bequests, there is full insurance for mortality risk, so that the individual's savings decision implies the Euler equation:

$$\lambda_t^{j,s} = \lambda_{t+1}^{j,s} R_{t+1} \beta. \tag{22}$$

[34]Note, this notation is slightly different from the notation of Section 3, where s refers to the birth date. Where possible, I keep the notation close to Maliar and Maliar (2003).

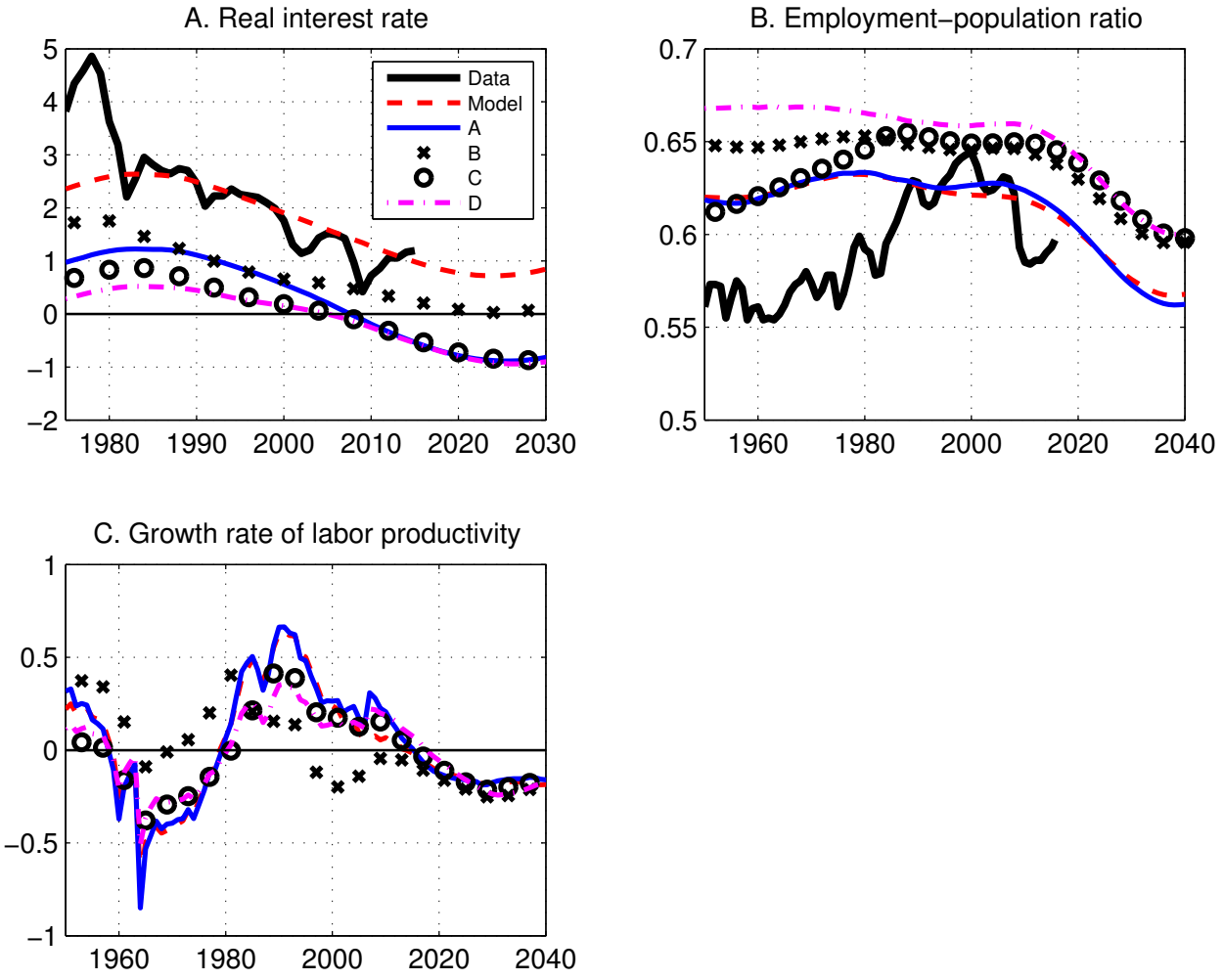

Figure B.3: Robustness exercises. This figure plots real interest rates, employment-population ratios, and the growth rate of labor productivity under different calibrations.

Optimizing (21) by choices of consumption and labor yields the first order conditions, first for the marginal utility of consumption: $\phi_t^{j,s} u_1(c_t^{j,s}, \ell_t^{j,s}) = \lambda_t^{j,s}$, and second, for the labor supply decision: $\phi_t^{j,s} u_2(c_t^{j,s}, \ell_t^{j,s}) = \lambda_t^{j,s} z^s w_t$. The Euler equation (22) implies that for any two individuals j, i, the ratios of their marginal utilities is constant for all time periods t, t', which implies:

$$\frac{\lambda_t^{j,s}}{\lambda_t^{i,s'}} = \frac{\lambda_{t'}^{j,s}}{\lambda_{t'}^{i,s'}} = \frac{\lambda^{i,s'}}{\lambda^{j,s}},$$

where $\lambda^{j,s} = \frac{\lambda_t}{\lambda_t^{j,s}}$. Using this condition, we can rewrite the first order conditions for the individual's problem as:

$$\lambda^{j,s} \phi_t^{j,s} u_1(c_t^{j,s}, \ell_t^{j,s}) = \lambda_t, \tag{23}$$

and:

$$\lambda^{j,s} \phi_t^{j,s} u_2(c_t^{j,s}, \ell_t^{j,s}) = \lambda_t z^s w_t. \tag{24}$$

These two equations, together with each individual's budget constraints and aggregate definitions, characterize the decentralized economy.

Now consider a period-by-period problem of a social planner who chooses consumption and labor

supply for each individual in each cohort to maximize the sum of individual utilities, weighted by welfare weights $\lambda^{j,s}$:

$$U(c_t, \ell_t) = \max_{\{c_t^{j,s}, \ell_t^{j,s}\}} \left\{ \sum_s \int \lambda^{j,s} \phi_t^{j,s} u(c_t^{j,s}, \ell_t^{j,s}) \, \mathrm{d}j \right\}, \tag{25}$$

subject to the definitions for total consumption ($c_t = \sum_s n_t^s c_t^{j,s}$) and for total labor supplied in efficiency unit terms ($\ell_t = \sum_s n_t^s z^s \ell_t^{j,s}$). Because individuals within a cohort are identical and the measure of individuals within a cohort is given by n_t^s, we can write this problem as:

$$U(c_t, \ell_t) = \max_{\{c_t^{j,s}, \ell_t^{j,s}\}} \left\{ \sum_s n_t^s \lambda^{j,s} \phi_t^{j,s} u(c_t^{j,s}, \ell_t^{j,s}) \right\}.$$

Letting the Lagrange multiplier on the definitions of total consumption and labor be φ_t and ν_t respectively, the first order conditions of this static problem are:

$$n_t^s \lambda^{j,s} \phi_t^{j,s} u_1(c_t^{j,s}, \ell_t^{j,s}) = n_t^s \varphi_t,$$

and:

$$n_t^s \lambda^{j,s} \phi_t^{j,s} u_2(c_t^{j,s}, \ell_t^{j,s}) = n_t^s \nu_t z^s.$$

The envelope conditions for the problem (25) on c_t and ℓ_t are simply φ_t and ν_t, respectively, so that $U_1(c_t, \ell_t) = \varphi_t$ and $U_2(c_t, \ell_t) = \nu_t$.

Now consider the problem where the planner optimizes by choice of total consumption, total efficiency units of labor supplied ℓ_t, and total savings, the social utility function over time:

$$\max_{\{c_t, k_t, \ell_t\}} \sum_{t=0}^{\infty} \beta^t U(c_t, \ell_t),$$

subject to the economy's resource constraint each period. Letting λ_t be the Lagrange multiplier on the resource constraint, the first order conditions of this problem imply the same expressions as those equations that characterize the decentralized economy's problem. In particular, we get for the choice of aggregate savings: $\lambda_t = \lambda_{t+1} R_{t+1} \beta$. For aggregate consumption, the first order condition implies the standard condition: $U_1(c_t, \ell_t) = \lambda_t$, and for the choice of labor, $U_2(c_t, \ell_t) = \lambda_t w_t$, where I have substituted in w_t for the marginal product of labor.

To derive expressions for the shocks, the particular setup of the individual's problem in the lifecycle model has $u(c_t^{j,s}, \ell_t^{j,s}) = \frac{(c_t^{j,s})^{1-\sigma}}{1-\sigma} - v^s \frac{(\ell_t^{j,s})^{1+\varphi}}{1+\varphi}$. The $\phi_t^{j,s}$ are one when the individual is alive and zero otherwise. The skill of each individual z^s is positive in periods the individual is alive and zero otherwise. The first order conditions for the optimization of the social utility function are:

$$\lambda^{j,s} \phi_t^{j,s} (c_t^{j,s})^{-\sigma} = \varphi_t,$$

and:

$$\lambda^{j,s} \phi_t^{j,s} v^s (\ell_t^{j,s})^{\varphi} = \nu_t z^s.$$

This implies:

$$c_t^{j,s} = \varphi_t^{-1/\sigma} (\lambda^{j,s} \phi_t^{j,s})^{1/\sigma},$$

and:

$$z^s \ell_t^{j,s} = \nu_t^{1/\varphi} (z^s)^{1+1/\varphi} (v^s \phi_t^{j,s} \lambda^{j,s})^{-1/\varphi}.$$

Integrating (summing) these expressions with respect to individuals (cohorts) gives aggregate

consumption:

$$c_t = \sum_s n_t^s c_t^{j,s} = \varphi_t^{-1/\sigma} \left(\sum_s n_t^s (\lambda^{j,s} \phi_t^{j,s})^{1/\sigma} \right).$$

For aggregate labor supplied in efficiency units:

$$\ell_t = \sum_s n_t^s z^s \ell_t^{j,s} = \nu_t^{1/\varphi} \left(\sum_s n_t^s (z^s)^{1+1/\varphi} (v^s \phi_t^{j,s} \lambda^{j,s})^{-1/\varphi} \right).$$

These expressions imply that individual consumption and labor supply are fractions of their respective aggregates:

$$c_t^{j,s} = \frac{(\lambda^{j,s} \phi_t^{j,s})^{1/\sigma}}{\sum_s n_t^s (\lambda^{j,s} \phi_t^{j,s})^{1/\sigma}} c_t, \qquad \text{and} \qquad \ell_t^{j,s} = \frac{(z^s)^{1/\varphi} (v^s \phi_t^{j,s} \lambda^{j,s})^{-1/\varphi}}{\sum_s n_t^s (z^s)^{1+1/\varphi} (v^s \phi_t^{j,s} \lambda^{j,s})^{-1/\varphi}} \ell_t.$$

More compactly, $c_t^{j,s} = \chi_1^{j,s} c_t$ and $\ell_t^{j,s} = \chi_2^{j,s} \ell_t$.

Substituting these expressions into the social utility function gives:

$$U = \sum_s n_t^s \lambda^{j,s} \phi_t^{j,s} \left((\chi_1^{j,s})^{1-\sigma} \frac{c_t^{1-\sigma}}{1-\sigma} - v^s (\chi_2^{j,s})^{1+\varphi} \frac{\ell_t^{1+\varphi}}{1+\varphi} \right),$$

which can be rearranged to get an aggregate utility function:

$$U = \phi_t \frac{c_t^{1-\sigma}}{1-\sigma} - v_t \frac{\ell_t^{1+\varphi}}{1+\varphi}.$$

In this representation, aggregate labor ℓ_t is expressed as efficiency units of labor. Reorganizing this gives:

$$U = \phi_t \left[\frac{c_t^{1-\sigma}}{1-\sigma} - \frac{v_t}{\phi_t} \left(\frac{\ell_t^{1+\varphi}}{1+\varphi} \right) \right].$$

The process ϕ_t can be interpreted as a preference shock while v_t/ϕ_t is the labor wedge on the efficiency units of labor supplied. Finally, if we know the welfare weights $\{\lambda^{j,s}\}_{j,s}$, all terms in ϕ_t and v_t are exogenous and can be computed.

The final expression to determine is the term which converts aggregate supply of units of labor, denoted by h_t, into ℓ_t, the aggregate supply of efficiency units of labor which enters the firm's production function. To get this shock, we integrate $\ell_t^{j,s}$ over individuals and cohorts:

$$\sum_s \int \ell_t^{j,s} \, \mathrm{d}j = \sum_s n_t^s \ell_t^{j,s},$$

and compute $A_t = \ell_t / \ell_t = \sum_s n_t^s z_t^s \ell_t^{j,s} / \sum_s n_t^s \ell_t^{j,s}$, which becomes:

$$A_t = \frac{\sum_s n_t^s (z^s)^{1+1/\varphi} (v_t^{j,s} \phi_t^{j,s} \lambda^{j,s})^{-1/\varphi}}{\sum_s n_t^s (z^s)^{1/\varphi} (v_t^{j,s} \phi_t^{j,s} \lambda^{j,s})^{-1/\varphi}}.$$

The derivations show that the effect of heterogeneity by age can be summarized by four exogenous and foreseen paths of aggregate productivity θ_t, to labor input A_t, to the consumption utility shifter ϕ_t, and to the labor disutility shifter v_t.

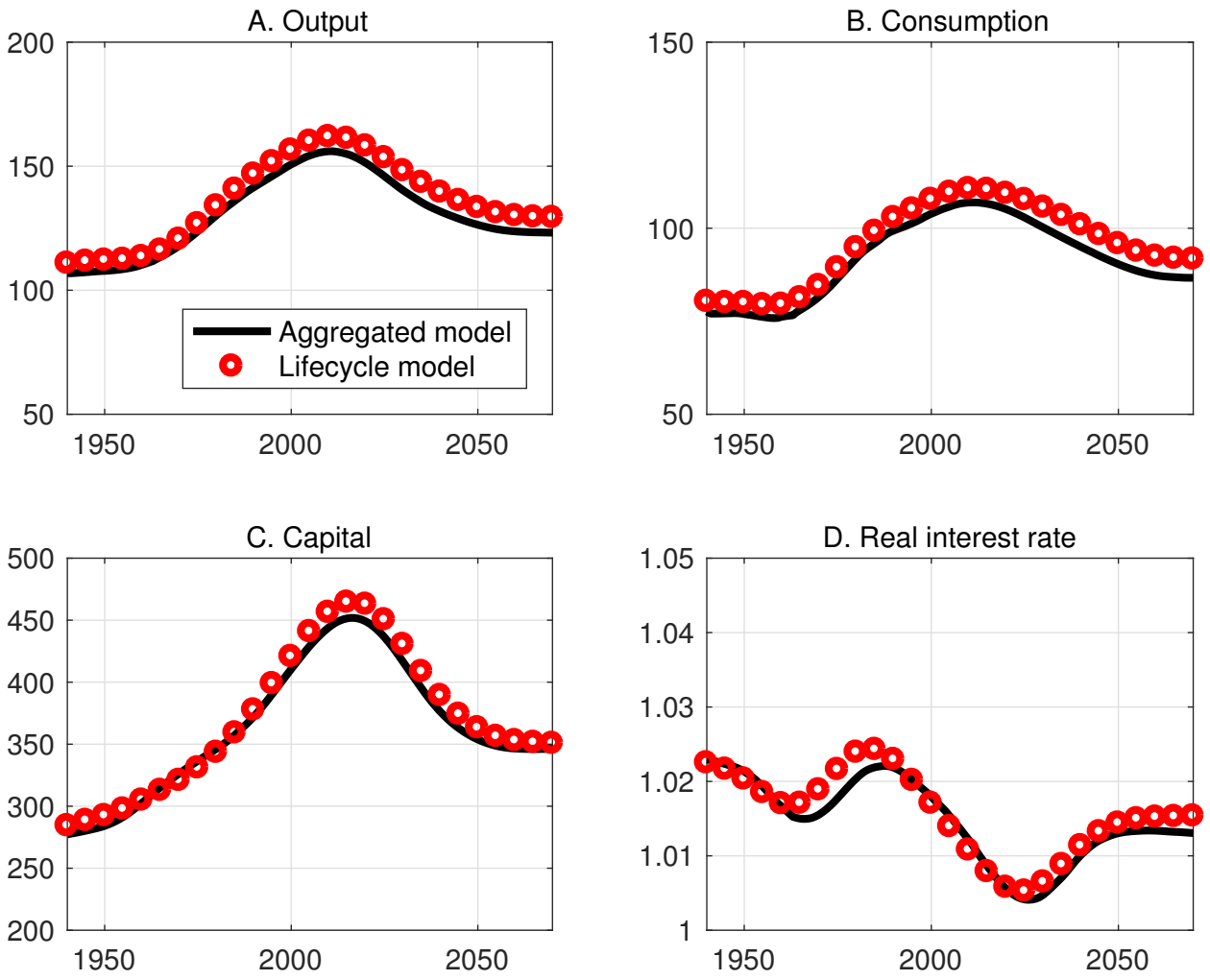

Figure C.1: Comparison of aggregated and lifecycle model. This figure compares the paths for aggregate variables from the full lifecycle model solution and the approximation under the linearized framework and the fully anticipated paths of the average mortality rate and labor income tax rate.

C.1 Accuracy of the approximation

Figure C.1 plots the full nonlinear global solution of the lifecycle model with the calibrated demographic changes against the piecewise-linear solution with anticipated wedges arising from demographic changes. The two methods give very similar paths for the key variables in the model, including for the real interest rate.

D Estimation with trend structural changes and the ZLB

This section details the estimation strategy tailored to my application with anticipated demographic shocks and where the zero lower bound is accounted for over the period 2009Q1 to 2015Q1.

D.1 Solution method

A linear rational-expectations model can be written as:

$$\mathbf{A}x_t = \mathbf{C} + \mathbf{B}x_{t-1} + \mathbf{D}\mathbb{E}_t x_{t+1} + \mathbf{F}w_t, \tag{26}$$

where x_t is a $n \times 1$ vector of state and jump variables and w_t is a $l \times 1$ vector of exogenous variables. A solution to (26), following Binder and Pesaran (1995), is:

$$x_t = \mathbf{J} + \mathbf{Q}x_{t-1} + \mathbf{G}\varepsilon_t.$$

As in Binder and Pesaran (1995), $\mathbf{Q}$ is solved by iterating on the quadratic expression:

$$\mathbf{Q} = [\mathbf{A} - \mathbf{D}\mathbf{Q}]^{-1}\mathbf{B}.$$

With $\mathbf{Q}$ in hand, compute $\mathbf{J}$ and $\mathbf{G}$ with:

$$\begin{aligned}\mathbf{J} &= [\mathbf{A} - \mathbf{D}\mathbf{Q}]^{-1}(\mathbf{C} + \mathbf{D}\mathbf{J}) \\ \mathbf{G} &= [\mathbf{A} - \mathbf{D}\mathbf{Q}]^{-1}\mathbf{F}.\end{aligned}$$

That is, $\mathbf{J}$, $\mathbf{Q}$ and $\mathbf{G}$ are conformable matrices which are functions of the structural matrices $\mathbf{A}$, $\mathbf{B}$, $\mathbf{C}$, $\mathbf{D}$ and $\mathbf{E}$.

In a model where agents have time-varying beliefs about the evolution of the model's structural parameters $\mathbf{A}_t$, $\mathbf{B}_t$, $\mathbf{C}_t$, $\mathbf{D}_t$ and $\mathbf{F}_t$, the solution becomes:

$$x_t = \mathbf{J}_t + \mathbf{Q}_t x_{t-1} + \mathbf{G}_t w_t, \tag{27}$$

where $\mathbf{J}_t$, $\mathbf{Q}_t$ and $\mathbf{G}_t$ are conformable matrices which are functions of the evolution of beliefs about the time-varying structural matrices $\mathbf{A}_t$, $\mathbf{B}_t$, $\mathbf{C}_t$, $\mathbf{D}_t$ and $\mathbf{F}_t$ (Kulish and Pagan, 2016). They satisfy the recursion:

$$\begin{aligned}\mathbf{Q}_t &= [\mathbf{A}_t - \mathbf{D}_t\mathbf{Q}_{t+1}]^{-1}\mathbf{B}_t \\ \mathbf{J}_t &= [\mathbf{A}_t - \mathbf{D}_t\mathbf{Q}_{t+1}]^{-1}(\mathbf{C}_t + \mathbf{D}_t\mathbf{J}_{t+1}) \\ \mathbf{G}_t &= [\mathbf{A}_t - \mathbf{D}_t\mathbf{Q}_{t+1}]^{-1}\mathbf{E}_t,\end{aligned}$$

where the final structures $\mathbf{Q}_T$ and $\mathbf{J}_T$ are known and computed from the time invariant structure above under the terminal period's structural parameters.

Anticipated changes in the path of demographic shocks and the zero lower bound are anticipated changes to the model's structural parameters which can be handled by solution (27) (see Jones, 2015a; Kulish and Pagan, 2016, for details and a discussion).

D.2 Kalman filter

Likelihood methods are used to estimate the parameters of the monetary policy rule and the parameters of the transitory shocks. For that, we need to filter the data, and owing to the linear structure of (27), we can use the Kalman filter, and exploit its computational advantages.

The model in its state space representation is:

$$x_t = \mathbf{J}_t + \mathbf{Q}_t x_{t-1} + \mathbf{G}_t \varepsilon_t \tag{28}$$

$$z_t = \mathbf{H}_t x_t. \tag{29}$$

The error is distributed $\varepsilon_t \sim N(0, \Omega)$ where Ω is the covariance matrix of ε_t. By assumption, there is

no observation error of the data z_t. The Kalman filter recursion is given by the following equations, conceptualized as the predict and update steps. The state of the system is $(\hat{x}_t, \mathbf{P}_{t-1})$. In the predict step, the structural matrices $\mathbf{J}_t$, $\mathbf{Q}_t$ and $\mathbf{G}_t$ are used to compute a forecast of the state $\hat{x}_{t|t-1}$ and the forecast covariance matrix $\mathbf{P}_{t|t-1}$ as:

$$\begin{aligned}\hat{x}_{t|t-1} &= \mathbf{J}_t + \mathbf{Q}_t \hat{x}_t \\ \mathbf{P}_{t|t-1} &= \mathbf{Q}_t \mathbf{P}_{t-1} \mathbf{Q}_{t|t-1}^\top + \mathbf{G}_t \Omega \mathbf{G}_t^\top.\end{aligned}$$

This formulation differs from the time-invariant Kalman filter because in the forecast stage the matrices $\mathbf{J}_t$, $\mathbf{Q}_t$ and $\mathbf{G}_t$ can vary over time. We update these forecasts with imperfect observations of the state vector. This update step involves computing forecast errors $\widetilde{y}_t$ and its associated covariance matrix $\mathbf{S}_t$ as:

$$\begin{aligned}\widetilde{y}_t &= z_t - \mathbf{H}_t \hat{x}_{t|t-1} \\ \mathbf{S}_t &= \mathbf{H}_t \mathbf{P}_{t|t-1} \mathbf{H}_t^\top.\end{aligned}$$

The Kalman gain matrix is given by:

$$\mathbf{K}_t = \mathbf{P}_{t|t-1} \mathbf{H}_t^\top \mathbf{S}_t^{-1}.$$

With $\widetilde{y}_t$, $\mathbf{S}_t$ and $\mathbf{K}_t$ in hand, the optimal filtered update of the state x_t is

$$\hat{x}_t = \hat{x}_{t|t-1} + \mathbf{K}_t \widetilde{y}_t,$$

and for its associated covariance matrix:

$$\mathbf{P}_t = (I - \mathbf{K}_t \mathbf{H}_t)\, \mathbf{P}_{t|t-1}.$$

The Kalman filter is initialized with x_0 and $\mathbf{P}_0$ determined from their unconditional moments, and is computed until the final time period T of data.

D.2.1 Kalman smoother

With the estimates of the parameters and durations in hand at time period T, the Kalman smoother gives an estimate of $x_{t|T}$, or an estimate of the state vector at each point in time given all available information (see Hamilton, 1994). With $\hat{x}_{t|t-1}$, $\mathbf{P}_{t|t-1}$, $\mathbf{K}_t$ and $\mathbf{S}_t$ in hand from the Kalman filter, the vector $x_{t|T}$ is computed by:

$$x_{t|T} = \hat{x}_{t|t-1} + \mathbf{P}_{t|t-1} r_{t|T},$$

where the vector $r_{T+1|T} = 0$ and is updated with the recursion:

$$r_{t|T} = \mathbf{H}_t^\top \mathbf{S}_t^{-1} \left(z_t - \mathbf{H}_t \hat{x}_{t|t-1}\right) + \left(I - \mathbf{K}_t \mathbf{H}_t\right)^\top \mathbf{P}_{t|t-1}^\top r_{t+1|T}.$$

Finally, to get an estimate of the shocks to each state variable under this model's shock structure, denoted by e_t, we compute:

$$e_t = \mathbf{G}_t \varepsilon_t = \mathbf{G}_t r_{t|T}.$$

From these, we get an estimate of the structural shocks.

D.3 Sampler

This section describes the sampler used to obtain the posterior distribution of interest. Denote by ϑ the vector of parameters to be estimated and by $\mathbf{T}$ the vector of ZLB durations that are observed each period. Denote by $\mathcal{Z} = \{z_\tau\}_{\tau=1}^{T}$ the sequence of vectors of observable variables. The posterior $\mathcal{P}(\vartheta \mid \mathbf{T}, \mathcal{Z})$ satisfies:

$$\mathcal{P}(\vartheta \mid \mathbf{T}, \mathcal{Z}) \propto L(\mathcal{Z}, \mathbf{T} \mid \vartheta) \times \mathcal{P}(\vartheta).$$

With Gaussian errors, the likelihood function $L(\mathcal{Z}, \mathbf{T} \mid \vartheta)$ is computed using the appropriate sequence of structural matrices and the Kalman filter:

$$\log L(\mathcal{Z}, \mathbf{T} \mid \vartheta) = -\left(\frac{N_z T}{2}\right) \log 2\pi - \frac{1}{2}\sum_{t=1}^{T} \log\det \mathbf{H}_t \mathbf{S}_t \mathbf{H}_t^\top - \frac{1}{2}\sum_{t=1}^{T} \widetilde{y}_t^\top \left(\mathbf{H}_t \mathbf{S}_t \mathbf{H}_t^\top\right)^{-1} \widetilde{y}_t.$$

The prior is simply computed using priors over ϑ which are consistent with the literature.

The Markov Chain Monte Carlo posterior sampler has a single block, corresponding to the parameters ϑ.[35] The sampler at step j is initialized with the last accepted draw of the structural parameters ϑ_{j-1}.

The block is a standard Metropolis-Hastings random walk. First start by selecting which structural parameters to propose a new value for. For those parameters, draw a new proposal ϑ_j from a proposal density centered at ϑ_{j-1} and with thick tails to ensure sufficient coverage of the parameter space and an acceptance rate of roughly 20% to 25%. The proposal ϑ_j is accepted with probability $\frac{\mathcal{P}(\vartheta_j|\mathbf{T},\mathcal{Z})}{\mathcal{P}(\vartheta_{j-1}|\mathbf{T},\mathcal{Z})}$. If ϑ_j is accepted, then set $\vartheta_{j-1} = \vartheta_j$.

E Estimation results

E.1 Data

In the baseline estimation, I use four observable series: real output growth per capita, real consumption growth per capita, GDP deflator inflation, and the Federal Funds rate. For the expected ZLB durations, I use Morgan Stanley's measure of the months until the first rate hike, which was constructed using Federal Funds futures prices. These five series are plotted in Figure E.1. Prior to estimation, I demean output and consumption growth.

The expected ZLB durations follow a hump-shaped pattern, reaching a peak in 2012-13. This accords with the results of Swanson and Williams (2014).

E.2 Calibrated parameters

I calibrate a small set of the structural parameters to values commonly used in the literature. The steady-state value of θ is set at 8, which implies a steady-state markup over marginal costs of $\theta/(\theta-1) = 14\%$. Applying the analysis of Keen and Wang (2005), who compare the Calvo price adjustment parameterization across different values of Rotemberg quadratic cost of price adjustment, I set ϕ_p to 100, which in a Calvo pricing model would imply that with a steady-state markup of 14% over marginal costs, about 25% of firms each quarter reset their prices. With β and φ given by the lifecycle parameterization, and θ and ϕ_p calibrated, the choice of ϕ_p implies that the coefficient on marginal costs in the Phillips curve is 0.07, which is consistent with low estimates of the slope of the

[35] It is worth noting that one can estimate in addition to the structural parameters ϑ, the expected zero lower bound durations can be estimated together with the structural parameters (as in Kulish et al., 2014), in which case an additional block is needed in the posterior sampler (see, for example, Jones, 2015b).

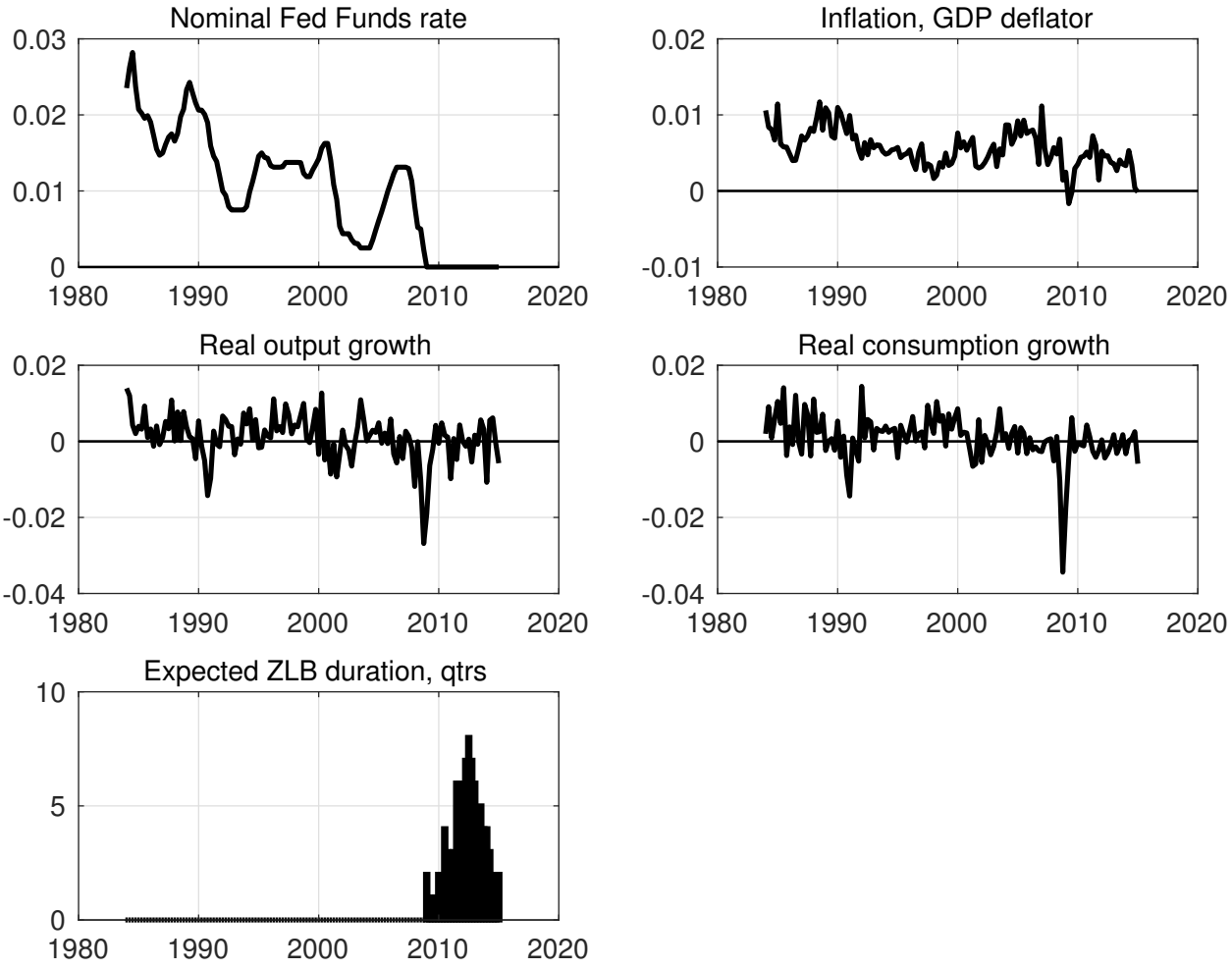

Figure E.1: Data used in estimation. The nominal interest rate is not an observable variable between 2009 and 2015, while the ZLB duration becomes an observable variable during those quarters.

Phillips curve in the DSGE literature, and is close to the value of the slope of the Phillips curve calibrated by Ireland (2004) in a similar model to the aggregate model here.

E.2.1 Full estimation results

I estimate the model over the full sample 1984Q1 to 2015Q1. The prior distributions and posterior estimates for each estimation are given in Table 4. The priors used for the structural parameters are standard in the literature. Prior beliefs suggest preference shocks are relatively large relative to other shocks (including markup shocks, which I scale by $1/\phi_p$, the coefficient on markup shocks in the linearized model). Permanent and transitory technology shocks are given the same prior variance. The policy rule parameters are given standard prior distributions (see, for example, Smets and Wouters, 2007).

Figure E.2 plots R^2 Gelman chain diagnostics for the baseline estimates from 1984Q1 to 2015Q1, and illustrate that the estimated posterior distributions lie comfortably below 1.1, commonly used as a value indicating convergence of the posterior distributions (see, for example, Bianchi, 2013).

The prior and posterior distributions for the estimated parameters are plotted in Figure E.3.

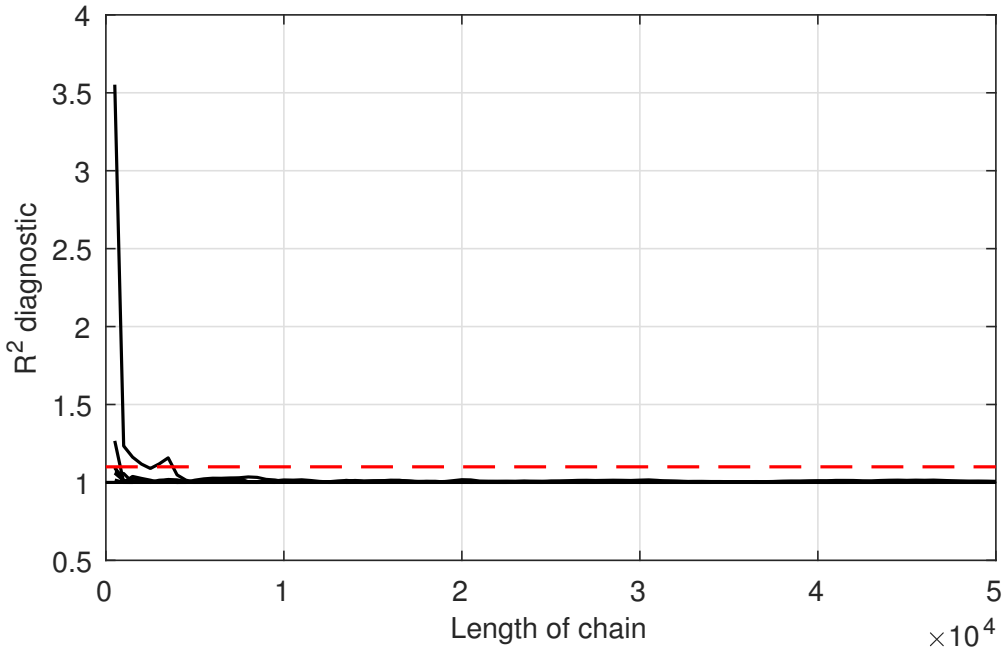

Figure E.2: Gelman chain diagnostic. The black lines are the calculated R^2 statistic for each structural parameter.

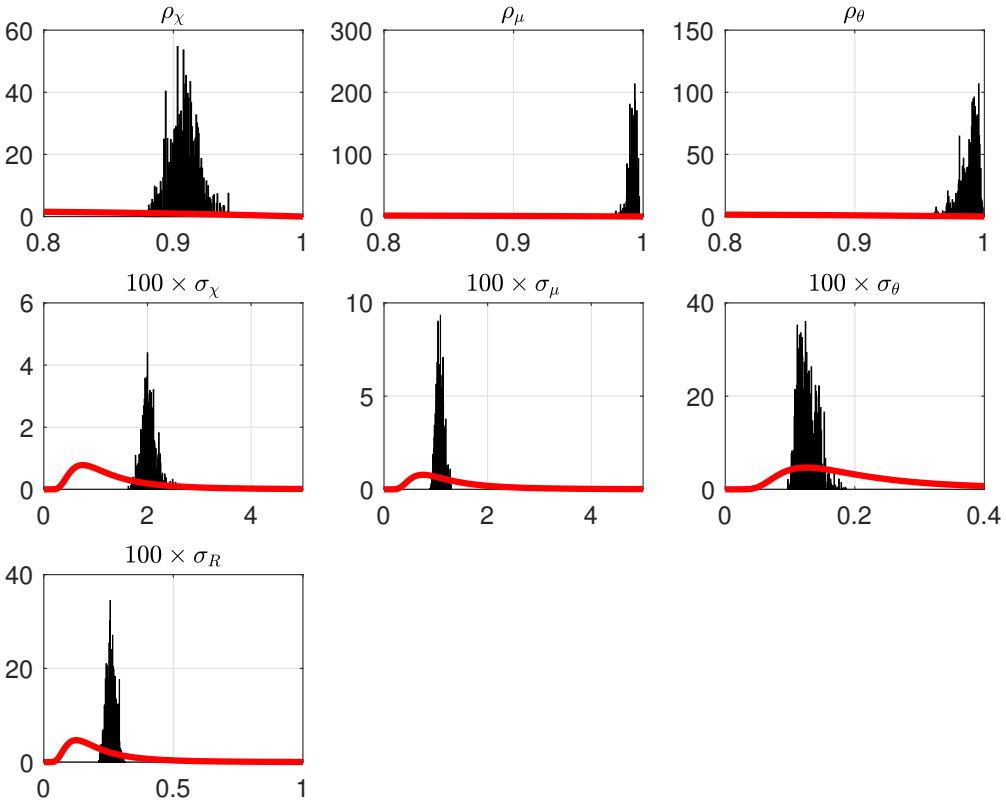

Figure E.3: Estimated posterior distributions. This figure shows the priors (red) and posteriors (black) of the estimated parameters.

Table 4: Estimated parameters

	Prior			Posterior		
Parameter	Dist	Mode	St Dev	Mode	Mean	St Dev
ρ_χ	B	0.600	0.20	0.903	0.908	0.011
ρ_μ	B	0.600	0.20	0.995	0.992	0.003
ρ_ξ	B	0.600	0.20	0.996	0.989	0.007
$100 \times \sigma_\chi$	IG	0.750	1.50	2.004	2.035	0.150
$100 \times \sigma_\mu$	IG	0.750	1.50	1.140	1.078	0.070
$100 \times \sigma_\xi$	IG	0.125	0.25	0.122	0.126	0.015
$100 \times \sigma_r$	IG	0.125	0.25	0.256	0.259	0.017

F Data sources

F.1 Data sources for OLG model

This section details the data series used for calibration of the lifecycle model.

Current Population Survey I use the Current Population Survey to get estimates of labor force participation rates by age in the calibration of the disutility of providing labor.

Census/American Community Survey I use Census and American Community Survey extracts from IPUMS-USA to compute the experience-productivity profiles following Elsby and Shapiro (2012).

Social Security Administration I use Social Security Administration estimates and forecasts for mortality rates γ_t^s. These numbers are directly fed into the model as anticipated paths.

BLS-Multifactor Productivity Program I use the BLS-MFP data to construct a measure of the real interest rate from observed capital-output ratios.

From the BLS website: "Capital input data–service-flows of equipment, structures, intellectual property products, inventories, and land. BLS measures of capital service inputs are prepared using NIPA data on real gross investment in depreciable assets and inventories. Labor input data–hours worked by all persons engaged in a sector–is based on information on employment and average weekly hours collected in the monthly BLS survey of establishments and the hours at work survey. Labor composition data are based on March supplements to the Current Population Survey."

F.2 Data sources for estimating the New Keynesian model

I use data on output, consumption, inflation, and interest rates.[36] Construction of the data series follows Smets and Wouters (2007). The codes for each raw data series are as follows:

- Real Gross Domestic Product, 3 Decimal (GDPC96). Current, $.
- Gross Domestic Product: Implicit Price Deflator (GDPDEF). Index, 2009=100.
- Personal Consumption Expenditures (PCEC). Current, $.

[36] A public version of the data list corresponding to the Smets and Wouters (2007) series can be obtained at https://research.stlouisfed.org/pdl/803.

- Total population (CNP16OV), Thousands of Persons.

To map these data series to the model variables, I do the following transformations.

1. Construct the series LNSindex, which is an index of CNP16OV where 1992Q3=1. I adjust the CNP16OV series to account for breaks in the series each January, due to revisions from updated Census reports, which can be substantial. To do this, I impute an estimate of each January's monthly change in population and construct an estimate of the revised change in population from the actual change to the constructed imputed change. I then distribute that revised change in population across the preceding 12 months.

2. Construct the series CE16OVIndex, which is an index of CE16OV where 1992Q3=1.

3. Output = Y_t = ln(GDPC96 / LNSindex) * 100. Then compute the percentage change in output as an observable, $\ln Y_t - \ln Y_{t-1}$.

4. Inflation = Π_t = ln(GDPDEF / GDPDEF(-1)) * 100.

5. Consumption = C_t = ln((PCEC / GDPDEF) / LNSindex) * 100. Then compute the percentage change in consumption as an observable, $\ln C_t - \ln C_{t-1}$.

The interest rate is the Federal Funds Rate, taken from the Federal Reserve Economic Database.

References

Stephanie Aaronson, Thomaz Cajner, Bruce Fallick, Felix Galbis-Reig, Christopher Smith, and William Wascher. Labor Force Participation: Recent Developments and Future Prospects. *Brookings Papers on Economic Activity*, Fall:197–255, 2014.

Yunus Aksoy, Henrique S. Basso, Tobias Grasl, and Ron P. Smith. Demographic Structure and Macroeconomic Trends. 2015.

Sungbae An and Frank Schorfheide. Bayesian Analysis of DSGE Models. *Econometric Reviews*, 26(2-4): 113–172, 2007.

Juan Antolin-Diaz, Thomas Drechsel, and Ivan Petrella. Following the Trend: Tracking GDP when Long-Run Growth is Uncertain. 2014.

S Boragan Aruoba, Pablo Cuba-Borda, and Frank Schorfheide. Macroeconomic Dynamics Near the ZLB: A Tale of Two Countries. 2016.

Orazio Attanasio, Sagiri Kitao, and Giovanni L. Violante. Global Demographic Trends and Social Security Reform. *Journal of Monetary Economics*, 54:144–198, 2007.

A. J. Auerbach and L. J. Kotlikoff. *Dynamic Fiscal Policy*. Cambridge University Press, 1987.

David Backus, Thomas Cooley, and Espen Henriksen. Demography and Low-Frequency Capital Flows. *Journal of International Economics*, 92:S94–S102, 2014.

Ben S. Bernanke. Why Are Interest Rates So Low? 2015.

Francesco Bianchi. Regime Switches, Agents' Beliefs, and Post-World War II U.S. Macroeconomic Dynamics. *Review of Economic Studies*, 80(2):463–490, 2013.

M. Binder and H.M. Pesaran. Multivariate Rational Expectations Models and Macroeconomic Modelling: A Review and Some New Results. Cambridge Working Papers in Economics 9415, Faculty of Economics, University of Cambridge, 1995.

Olivier Blanchard, Eugenio Cerutti, and Lawrence Summers. Inflation and Activity: Two Explorations, and their Monetary Policy Implications. 2015.

Olivier J Blanchard. Debt, Deficits, and Finite Horizons. *Journal of Political Economy*, 93(2):223–47, April 1985.

David E. Bloom, David Canning, and Jaypee Sevilla. Economic Growth and the Demographic Transition. Working Paper 8685, National Bureau of Economic Research, December 2001.

David E. Bloom, David Canning, and Gunther Fink. Implications of Population Aging for Economic Growth. 2011.

Fabio Canova, Filippo Ferroni, and Christian Matthes. Approximating Time Varying Structural Models with Time Invariant Structures. 2015.

Carlos Carvalho, Andrea Ferrero, and Fernanda Nechio. Demographics and Real Interest Rates: Inspecting the Mechanism. 2015.

Carlos Carvalho, Eric Hsu, and Fernanda Nechio. Measuring the Effect of the Zero Lower Bound on Monetary Policy. 2016.

Maria Casanova. Revisiting the Hump-Shaped Wage Profile. 2013.

Lawrence Christiano, Martin Eichenbaum, and Sergio Rebelo. When Is the Government Spending Multiplier Large? *Journal of Political Economy*, 119(1):78 – 121, 2011.

Lawrence J. Christiano, Martin S. Eichenbaum, and Mathias Trabandt. Understanding the Great Recession. *American Economic Journal: Macroeconomics*, 7(1):110–167, 2015.

George M. Constantinides. Intertemporal asset pricing with heterogeneous consumers and without demand aggregation. *The Journal of Business*, 55(2):253–267, 1982.

Gauti B. Eggertsson and Neil R. Mehrotra. A Model of Secular Stagnation. Brown University, 2014.

Gauti B Eggertsson and Michael Woodford. The Zero Bound on Interest Rates and Optimal Monetary Policy. *Brookings Papers on Economic Activity*, 1:139–233, 2003.

Gauti B. Eggertsson, Neil R. Mehrotra, and Jacob A. Robbins. A Quantitative Model of Secular Stagnation. Brown University, 2016.

Michael W. L. Elsby and Matthew D. Shapiro. Why Does Trend Growth Affect Equilibrium Employment: A New Explanation of an Old Puzzle. *American Economic Review*, 102(4):1378–1413, 2012.

John Fernald. Productivity and Potential Output Before, During and After the Great Recession. Technical Report 18, Federal Reserve Bank of San Francisco, 2012.

John Fernald. Productivity and Potential Output before, during, and after the Great Recession. *NBER Macroeconomics Annual*, 29(1):1–51, 2015.

Jesús Fernández-Villaverde, Grey Gordon, Pablo A. Guerrón-Quintana, and Juan Rubio-Ramírez. How Structural Are Structural Parameters? *NBER Macroeconomics Annual*, 22:83–132, June 2007.

James Feyrer. Demographics and Productivity. *The Review of Economics and Statistics*, 89(1):100–109, February 2007.

S Fujita and I Fujiwara. Aging and Deflation: Japanese Experience. 2014.

Etienne Gagnon, Benjamin K. Johannsen, and David Lopez-Salido. Understanding the New Normal: The Role of Demographics. Working paper, Board of Governors of the Federal Reserve System, 2016.

Robert Gordon. *The Rise and Fall of Economic Growth: The U.S. Standard of Living Since the Civil War*. Princeton University Press, 2016.

Luca Guerrieri and Matteo Iacoviello. Occbin: A Toolkit to Solve Models with Occasionally Binding Constraints Easily. *Journal of Monetary Economics*, 70:22–38, March 2015.

Fatih Guvenen, Fatih Karahan, Serdar Ozkan, and Jae Song. What Do Data on Millions of U.S. Workers Reveal about Life-Cycle Earnings Risk? Working Paper 20913, National Bureau of Economic Research, January 2015.

Robert E. Hall and Charles I. Jones. Why do Some Countries Produce So Much More Output Per Worker than Others? *The Quarterly Journal of Economics*, 114(1):83–116, 1999.

James D. Hamilton. *Time Series Analysis*. Princeton University Press, 1994.

James D. Hamilton, Ethan S. Harris, Jan Hatzius, and Kenneth D. West. The Equilibrium Real Funds Rate: Past, Present and Future. Working Paper 21476, National Bureau of Economic Research, August 2015.

Gary D. Hansen and Selahattin Imrohoroglu. Consumption Over the Life Cycle: The Role of Annuities. *Review of Economic Dynamics*, 11(3):566–583, 2008. ISSN 1094-2025. doi: http://dx.doi.org/10.1016/j.red.2007.12.004.

Peter Ireland. Technology Shocks in the New Keynesian Model. *Review of Economics and Statistics*, 86(4):923–936, November 2004.

Peter Ireland. Changes in the Federal Reserve's Inflation Target: Causes and Consequences. *Journal of Money, Credit and Banking*, 39(8):1851–1882, 2007.

Nir Jaimovich and Henry Siu. The Trend is the Cycle: Job Polarization and Jobless Recoveries. 2012.

Callum Jones. Unanticipated Shocks and Forward Guidance at the Zero Lower Bound. New York University, 2015a.

Callum Jones. Trend Growth, Forward Guidance and the Zero Lower Bound. 2015b.

Fatih Karahan, Benjamin Pugsley, and Aysehul Sahin. Understanding the 30-year Decline in the Startup Rate: A General Equilibrium Approach. 2015.

Benjamin D. Keen and Yongsheng Wang. What is a Realistic Value for Price Adjustment Costs in New Keynesian Models? 2005.

Narayana Kocherlakota. Why Americans Feel Poor, in One Chart. 2016.

Y.C. Kong, B. Ravikumar, and G. Vandenbroucke. Explaining Cross-Cohort Differences in Life-Cycle Earnings. 2016.

Mariano Kulish and Adrian Pagan. Estimation and Solution of Models with Expectations and Structural Changes. *Journal of Applied Econometrics*, 2016.

Mariano Kulish, Christopher Kent, and Kathryn Smith. Aging, Retirement, and Savings: A General Equilibrium Analysis. *The BE Journal of Macroeconomics*, 10(1), 2010.

Mariano Kulish, James Morley, and Tim Robinson. Estimating the Expected Duration of the Zero Lower Bound in DSGE Models with Forward Guidance. Research Paper 2014 ECON 32, Australian School of Business, 2014.

Nicole Maestas, Kathleen J. Mullen, and David Powell. The Effect of Population Aging on Economic Growth, the Labor Force and Productivity. Working Paper 22452, National Bureau of Economic Research, July 2016.

Lilia Maliar and Serguei Maliar. The Representative Consumer in the Neoclassical Growth Model with Idiosyncratic Shocks. *Review of Economic Dynamics*, 6:362–380, 2003.

Alisdair McKay, Emi Nakamura, and Jon Steinsson. The Power of Forward Guidance Revisited. *American Economic Review*, 106(10):3133–3158, 2016.

Virgiliu Midrigan and Thomas Philippon. Household Leverage and the Recession. 2016.

Lukasz Rachel and Thomas Smith. Secular Drivers of the Global Real Interest Rate. 2015.

Felix Reichling and Charles Whalen. Review of Estimates of the Frisch Elasticity of Labor Supply. Working Paper 13, Congressional Budget Office, 2012.

Jose-Victor Rios-Rull, Frank Schorfheide, Cristina Fuentes-Albero, Maxym Kryshko, and Raul Santaeulalia-Llopis. Methods Versus Substance: Measuring the Effects of Technology Shocks. 59(8):826–846, 2012.

Kenneth Rogoff. Debt Supercycle, Not Secular Stagnation. 2015.

Christopher Sims. Solving Linear Rational Expectations Models. *Computational Economics*, 20:1–20, 2002.

Frank Smets and Rafael Wouters. Shocks and Frictions in US Business Cycles: A Bayesian DSGE Approach. *American Economic Review*, 96(3):586–606, 2007.

Lawrence H. Summers. U.S. Economic Prospects: Secular Stagnation, Hysteresis, and the Zero Lower Bound. *Business Economics*, 49(2):65–73, 2014.

Eric Swanson and John Williams. Measuring the Effect of the Zero Lower Bound on Medium- and Long-term Interest Rates. *American Economic Review*, 104(10):3154–3185, 2014.

Iván Werning. Managing a Liquidity Trap: Monetary and Fiscal Policy. 2012.

E. Whitehouse. The Value of Pension Entitlements: A Model of Nine OECD Countries. Technical Report 9, OECD, 2003.

Johannes Wieland. Are Negative Supply Shocks Expansionary at the Zero Lower Bound? Manuscript, January 2014.

Arlene Wong. Population Aging and the Transmission of Monetary Policy to Consumption. 2015.

Jing Cynthia Wu and Fan Dora Xia. Measuring the Macroeconomic Impact of Monetary Policy at the Zero Lower Bound. manuscript, 2016.

Menahem Yaari. Uncertain Lifetime, Life Insurance, and the Theory of the Consumer. *Review of Economic Studies*, 32:137–150, 1965.