

*The subscripts on continuous time variables mean that they are functions of time. Derivatives without arguments are evaluated at the optimal values.*

## 1. Ramsey-Cass-Koopmans Model

All we need to add to the Solow model are consumers who make consumption-savings decisions to maximize utility. We assume that all consumers have the same preferences and also eat and save the same amounts, i.e., we assume the existence of a **representative consumer**<sup>1</sup> that maximizes utility  $U(c)$ , where  $c = \{c_0, \dots, c_t, \dots, c_T\}$ ,  $c_t = C_t/L_t$ ,  $T \leq \infty$ , is the stream of per-capita consumption from today onward. Moreover, no longer does capital evolve according to an exogenous law of motion, but according to how much the consumer decides to save.

**Planner's Problem.** Later, we will define markets where consumers have to go out and trade with the (representative) firm(s) and solve for a **competitive equilibrium**. But first we solve for the first-best, which is called the (benevolent) planner's problem (PP): it is as if there is a God who can control everything, and does his best to make everyone happy. In the simple version, the 1st Welfare Theorem (FWT) holds: the competitive equilibrium is a solution to the planner's problem. But once we realize the amount of assumptions we have to make on the planner for the FWT to hold, we will see why communism is not such a great idea (even if the leaders were actually benevolent).

$$\begin{aligned}
 (PP) \quad \max_c U(c) &= \int_0^T e^{-\rho t} u(c_t) dt \\
 \text{s.t.} \quad C_t + \dot{K}_t &= F(K_t, A_t L_t) - \delta K_t, \\
 K_0 &> 0 \text{ given, } K_T &= 0.
 \end{aligned}$$

<sup>1</sup>As a rule of thumb in economics, preferences are usually assumed to be identical across consumers because otherwise there is not much a model can explain. Moreover, there are few known preference structures that exhibit logical consistency, as you will learn in micro, and most well-accepted models seldom deviate from them. The representativeness comes not from the identical preference but from the assumption that consumers make the same decisions; this is a strong assumption that is not always assumed, but we will learn later that depending on the types of insurance opportunities that are available in the market, not entirely undesirable either.

The assumptions we make on the growth rates of  $A_t, L_t$  and the production function  $F(\cdot)$  are the same as in the Solow model. The function  $u(\cdot)$  is the instantaneous utility function, i.e., how much utility the consumer gets by eating *now*. The new parameter  $\rho$  is the discount rate, i.e. how much the consumer values the future as opposed to the present.

Finally, unlike in the Solow model, we assume a terminal condition  $K_T = 0$ , i.e., at the end of time, the consumer dies without leaving anything wasted. A technicality arises when  $T = \infty$  and we have to modify this condition, which we discuss later.

As we did in the Solow model, we can also define per-effective unit consumption  $\hat{c} = C/AL$ .

**Some technicalities.** Sometimes, we will make specific assumptions on the utility function and production function, as we did when we assumed Cobb-Douglas or less restrictively, HD1; but for most results such specific assumptions are not needed. When looking at models and results (and later your own), you want to get into the habit of identifying the necessary conditions and the least restrictive sufficient assumptions. For the growth model, most results go through with the mild assumption that  $u(\cdot)$  and the normalized production function  $f(\cdot)$  are twice-differentiable, strictly increasing and strictly concave. An innocuous assumption is that  $f(0) = 0$ , i.e., zero input results in zero output. Sometimes we need to additionally assume the *Inada Condition*, which means that

$$\begin{aligned} \lim_{c \rightarrow 0} u'(c) = \infty, & \quad \lim_{c \rightarrow \infty} u'(c) = 0 \quad \text{and} \\ \lim_{k \rightarrow 0} f'(k) = \infty, & \quad \lim_{k \rightarrow \infty} f'(k) = 0. \end{aligned}$$

However, for the RCK model and for most of the course, in fact, we will be assuming a constant-relative-risk-aversion (CRRA) utility function:

$$u(c) = \frac{c^{1-\gamma}}{1-\gamma}$$

which introduces another new parameter  $\gamma$ , the RRA coefficient. So henceforth, you might want to keep an eye on when the CRRA assumption was needed vs. was just a convenience.

Thanks to CRRA, and normalizing  $A_0 = 1$  (otherwise we can just divide everything by  $A_0$ ),

we can rewrite (PP) as

$$(PP)' \quad \max_{c_t, k_t} U = \int_0^T e^{-\hat{\rho}t} u(\hat{c}_t) dt$$

$$\text{s.t. } c_t + \dot{\hat{k}}_t = f(\hat{k}_t) - (n + g + \delta)\hat{k}_t,$$

$$\hat{k}_0 > 0 \text{ given, } \hat{k}_T = 0.$$

where  $\hat{\rho} = \rho - g(1 - \gamma)$ . For future discounting to make sense (and for existence of a BGP), we need to assume  $\hat{\rho} > 0$ , i.e.  $\rho > g(1 - \gamma)$ .

With this, we can now solve the model using a (current value) **Hamiltonian**.<sup>2</sup> I'll just write it down first and explain later. But note that the optimal solution should describe the evolution of  $(\hat{c}_t, \hat{k}_t)$  as *functions of time*, given  $\hat{k}_0 > 0$ .

$$\mathcal{H}(\hat{c}_t, \hat{k}_t, \lambda_t) = u(\hat{c}_t) + \lambda_t [f(\hat{k}_t) - (n + g + \delta)\hat{k}_t - \hat{c}_t]$$

Sufficient conditions for optimality (which are similar to f.o.c.'s for a Lagrangian) are:

$$\hat{c}_t : \frac{\partial \mathcal{H}}{\partial \hat{c}_t} = 0 \quad \Rightarrow \quad u'(\hat{c}_t) = \lambda_t, \quad (1a)$$

$$\hat{k}_t : \frac{\partial \mathcal{H}}{\partial \hat{k}_t} = -\dot{\lambda}_t + \hat{\rho}\lambda_t \quad \Rightarrow \quad -\dot{\lambda}_t = \lambda_t [f'(\hat{k}_t) - (n + g + \delta + \hat{\rho})], \quad (1b)$$

$$\lambda_t : \frac{\partial \mathcal{H}}{\partial \lambda_t} = \hat{k}_t \quad \Rightarrow \quad \dot{\hat{k}}_t = f(\hat{k}_t) - (n + g + \delta)\hat{k}_t - \hat{c}_t, \quad (1c)$$

along with boundary conditions that depend on our assumptions on  $T, \hat{k}_T$ . Here,  $\hat{c}_t$  is the control,  $\hat{k}_t$  is the state, and  $\lambda_t$  the costate. The costate basically plays the role of a Lagrangian multiplier, which is why we denote it by  $\lambda_t$ .

To understand the Hamiltonian, and set  $T = \infty$ , we need to understand a bit of Optimal Control.

---

<sup>2</sup>You can also use a present value Hamiltonian, this is just my personal preference.

## 2. Optimal Control

Let's take a look at the discrete-time version of our problem, which we use extensively later:

$$\begin{aligned} \max_{\{\hat{c}_t, \hat{k}_{t+1}\}_{t=0}^T} U &= \sum_{t=0}^T \hat{\beta}^t u(\hat{c}_t) \\ \text{s.t. } \hat{k}_{t+1} - \hat{k}_t &= f(\hat{k}_t) - \left[ (n + g + ng)\hat{k}_{t+1} + \delta\hat{k}_t \right] - \hat{c}_t, \quad t = 0, \dots, T, \\ \hat{k}_0 &> 0 \text{ given, } \hat{k}_{T+1} &= 0. \end{aligned}$$

where  $\hat{\beta} = \beta(1 + g)^{1-\gamma}$  is the effective **discount factor**, and  $\beta$  the discount factor. Set the current value Lagrangean with multipliers  $\hat{\beta}^t \lambda_t$  for each  $t = 0, \dots, T$ :

$$\mathcal{L} \left( \{\hat{c}_t, \hat{k}_{t+1}, \lambda_t\} \right) = \sum_{t=0}^{\infty} \hat{\beta}^t \left\{ u(\hat{c}_t) + \lambda_t \left[ f(\hat{k}_t) + (1 - \delta)\hat{k}_t - \hat{c}_t - (1 + n)(1 + g)\hat{k}_{t+1} \right] \right\}$$

F.o.n.c.'s:

$$\begin{aligned} \hat{c}_t : \quad & u'(\hat{c}_t) = \lambda_t \\ \hat{k}_{t+1} : \quad & -(\lambda_{t+1} - \lambda_t) = \lambda_{t+1} [f'(\hat{k}_{t+1}) - \delta] - \lambda_t (n + g + ng + 1 - \hat{\beta}) / \hat{\beta} \\ \lambda_t : \quad & \hat{c}_t + (\hat{k}_{t+1} - \hat{k}_t) = f(\hat{k}_t) - \left[ (n + g + ng)\hat{k}_{t+1} + \delta\hat{k}_t \right] \end{aligned}$$

along with boundary conditions we will learn later. Look familiar...?

**EXERCISE 1** Let  $\beta = 1 - \rho\tau$ ,  $\hat{\beta} = 1 - \hat{\rho}\tau$ , and suppose that the growth rates of population and technology are  $(n\tau, g\tau)$ , respectively. Similarly, let the depreciation rate of capital be  $\delta\tau$ . Here,  $\tau$  captures the length of one period in the discrete-time model; so instead of  $\hat{k}_{t+1}$ , write  $\hat{k}_{t+\tau}$ , and similarly for other variables (make sure to multiply flow variables by  $\tau$ !) Show that as  $\tau \rightarrow 0$ , the first order conditions for the Lagrangean become equivalent to the optimality conditions for the Hamiltonian, system (1). Basically, all you need to show is that

1.  $\hat{\rho} \rightarrow \rho - g(1 - \gamma)$
2. the resource constraint becomes equivalent.

## 2.1 Fixed Terminal Time/State

For *any* piecewise differentiable function of time,  $\lambda_t$ , we can write lifetime utility as

$$U(\lambda_t) = \int_0^T e^{-\rho t} \{u(\hat{c}_t) + \lambda_t [(f(\hat{k}_t) - (n + g + \delta)\hat{k}_t - \hat{c}_t - \dot{\hat{k}}_t)]\} dt$$

Integrate  $\int_0^T e^{-\rho t} \lambda_t \dot{\hat{k}}_t dt$  by parts:

$$\int_0^T e^{-\rho t} \lambda_t \dot{\hat{k}}_t dt = e^{-\rho T} \lambda_T \hat{k}_T - \lambda_0 \hat{k}_0 - \int_0^T e^{-\rho t} \{\dot{\lambda}_t \hat{k}_t - \rho \lambda_t \hat{k}_t\} dt$$

which is well defined by the piecewise differentiable assumption. Since  $\hat{k}_T = 0$ , we then have

$$U(\lambda_t) = \int_0^T e^{-\rho t} \{u(\hat{c}_t) + \lambda_t [(f(\hat{k}_t) - (n + g + \delta + \rho)\hat{k}_t - \hat{c}_t) + \dot{\lambda}_t \hat{k}_t]\} dt + \lambda_0 \hat{k}_0$$

Now, we want to characterize the f.o.n.c.'s of this problem. So suppose  $(\hat{c}_t^*, \hat{k}_t^*)$  attain the maximum, and that  $(\hat{c}_t, \hat{k}_t)$  is any other feasible path. Let

$$\Delta_k(t) = \hat{k}_t - \hat{k}_t^*, \quad \Delta_c(t) = \hat{c}_t - \hat{c}_t^*, \quad \Delta_U(\hat{c}_t, \hat{k}_t) = U(\hat{c}_t, \hat{k}_t) - U^*.$$

It follows that

$$\begin{aligned} & \Delta_U(\hat{c}_t, \hat{k}_t) \\ &= \int_0^T e^{-\rho t} \{ [u(\hat{c}_t) - u(\hat{c}_t^*)] + \lambda_t [f(\hat{k}_t) - f(\hat{k}_t^*) - (n + g + \delta + \rho)\Delta_k(t) - \Delta_c(t)] + \dot{\lambda}_t \Delta_k(t) \} dt \end{aligned}$$

Now suppose  $(\hat{c}_t, \hat{k}_t)$  are small perturbations around  $(\hat{c}_t^*, \hat{k}_t^*)$ . A 1st order Taylor expansion gives

$$= \int_0^T e^{-\rho t} [u'(\hat{c}_t^*) - \lambda_t] \Delta_c(t) dt + \int_0^T \{ \lambda_t [f'(\hat{k}_t^*) - (n + g + \delta + \rho)] + \dot{\lambda}_t \} \Delta_k(t) dt.$$

At a maximum, we must have  $\Delta_U(\hat{c}_t^*, \hat{k}_t^*) = 0$ . We can pick  $\lambda_t^*$  (as long as it's piecewise diff.), as well as the control  $\hat{c}_t^*$  to enforce this, i.e.

$$\begin{aligned} -\dot{\lambda}_t^* &= \lambda_t^* [f'(\hat{k}_t^*) - (n + g + \delta + \rho)] \\ u'(\hat{c}_t^*) &= \lambda_t^* \end{aligned}$$

which are the Hamiltonian optimality conditions (1) stated above.

## 2.2 Fixed Terminal Time/Free Terminal State

Without the assumption that  $K_T = 0$ , we need a way to replace the boundary condition. If  $K_T = 0$  is no longer assumed,  $\Delta_U(\hat{c}_t, \hat{k}_t)$  can be written as

$$\begin{aligned} & \Delta_U(\hat{c}_t, \hat{k}_t) \\ &= \int_0^T e^{-\hat{\rho}t} [u'(\hat{c}_t^*) - \lambda_t] \Delta_c(t) dt + \int_0^T \left\{ \lambda_t [f'(\hat{k}_t^*) - (n + g + \delta + \hat{\rho})] + \dot{\lambda}_t \right\} \Delta_k(t) dt - e^{-\hat{\rho}T} \lambda_T \Delta_k(T). \end{aligned}$$

Now, for finite  $T$ , it can never be the case that  $e^{-\hat{\rho}T} \lambda_T = 0$  (why?), so the optimal choice of  $\hat{k}_T = 0$ . Hence, even without assuming that  $K_T = 0$ , we obtain that it is optimal to do so.

## 2.3 Infinite Terminal Time/Free Terminal State

Finally, we arrive at the case when time  $T$  is infinite. Let's start using the Hamiltonian for notational ease: we want to solve

$$\begin{aligned} & \max_{\hat{c}_t, \hat{k}_t, \lambda_t} \int_0^\infty e^{-\hat{\rho}t} \{ \mathcal{H}(\hat{c}_t, \hat{k}_t, \lambda_t) - \lambda_t \dot{\hat{k}}_t \} dt \\ & \text{s.t. } \dot{\hat{k}}_t = f(\hat{k}_t) - (n + g + \delta + \hat{\rho}) \hat{k}_t - \hat{c}_t, \\ & \hat{k}_0 > 0 \text{ given.} \end{aligned}$$

Again using integration by parts, this can be rewritten as (with some heroic mathematical assumptions)

$$\max_{\hat{c}_t, \hat{k}_t, \lambda_t} \int_0^\infty e^{-\hat{\rho}t} \{ \mathcal{H}(\hat{c}_t, \hat{k}_t, \lambda_t) - \rho \lambda_t \hat{k}_t + \dot{\lambda}_t \hat{k}_t \} dt + \lambda_0 k_0 - \lim_{T \rightarrow \infty} e^{-\hat{\rho}T} \lambda_T \hat{k}_T.$$

If the same optimality conditions are to hold, we must have

$$\lim_{T \rightarrow \infty} e^{-\hat{\rho}T} \lambda_T \hat{k}_T = 0,$$

which is called a **transversality condition (TVC)**. A TVC is the boundary condition imposed when time is infinite, and the appropriate expression must be found on a case-to-case basis.

**EXERCISE 2** *What would happen if the TVC in the RCK model does not hold?*

### 3. Ramsey-Cass-Koopmans again

So the optimality conditions are (hope they make sense now)

$$\hat{c}_t : u'(\hat{c}_t) = \lambda_t \quad (2a)$$

$$\hat{k}_t : -\dot{\lambda}_t = \lambda_t [f'(\hat{k}_t) - (n + g + \delta + \hat{\rho})] \quad (2b)$$

$$\lambda_t : \dot{\hat{k}}_t = f(\hat{k}_t) - (n + g + \delta)\hat{k}_t - \hat{c}_t \quad (2c)$$

with the boundary conditions (you always need as many boundary conditions as there are variables!)

$$\hat{k}_0 > 0 \quad \text{given,} \quad \lim_{T \rightarrow \infty} e^{-\hat{\rho}T} u'(\hat{c}_T) \hat{k}_T = 0.$$

The BGP is simply when  $(\dot{\lambda}_t, \dot{\hat{k}}_t) = 0$ :

$$\begin{aligned} f'(\hat{k}^*) &= n + g + \delta + \hat{\rho} \\ \hat{c}^* &= f(\hat{k}^*) - (n + g + \delta)\hat{k}^*. \end{aligned}$$

The  $\hat{k}^*$  solution to the RCK model is called the “modified golden rule,” as opposed to the the “golden rule”  $\hat{k}_S$  in the Solow model. Clearly the MGR  $\hat{k}^*$  is smaller, because people optimally don’t save too much to enjoy more contemporaneous consumption.

#### 3.1 Phase Diagram

We can then characterize the dynamics of the system by a phase diagram on the  $(\hat{c}, \hat{k})$  plane. The phase diagram is characterized by the two loci on which we know that either  $\hat{c}$  or  $\hat{k}$  do not change.

**c-loci** From (2b), and the strict concavity of the utility function, we know that when

$$f'(\hat{k}) = (n + g + \delta + \hat{\rho})$$

$\hat{c}$  will not change. Note that

1. This is just a straight line parallel to the  $c$ -axis, such that  $\hat{k} = \hat{k}^*$ .
2. To the left of the loci  $\lambda$  must decrease, meaning  $\hat{c}$  must increase.

3. Conversely, to the right,  $\hat{c}$  must decrease.

**k-loci** From (2c), we know that when

$$\hat{c}(\hat{k}) = f(\hat{k}) - (n + g + \delta)\hat{k}$$

$\hat{k}$  will not change. There are several things to notice:

1. The loci begins at the point  $(0, 0)$ .
2. There exist  $\hat{k}^* < \hat{k}_S < \hat{k}_u < \infty$  s.t.  $\hat{c}$  peaks at  $k_S$ , and  $\hat{c}(\hat{k}_u) = 0$ . ( $\hat{k}_S$  is the golden rule level of capital from the Solow model.)
3. Below the loci,  $\hat{k}$  is increasing, and above, decreasing.

Now, we want to show that

**PROPOSITION 1** 1. Any optimal path must converge to the steady state.

2. The optimal path is unique and asymptotes toward the steady state.

Any path converging toward the steady state will satisfy the system. We must now rule out paths that asymptote toward  $\hat{c} = 0$  or  $\hat{k} = 0$  (there are no other possibilities). Hitting those values in finite time is obviously suboptimal due to the Inada conditions; we only need to rule out the asymptotic cases.

**Candidate solution converging to  $\hat{k} = 0$ ?** This can only happen if  $\hat{k}$  slows down (strictly) as it approaches 0. But taking the time derivative of (2c),

$$\lim_{\hat{k} \rightarrow 0} \dot{\hat{k}} = \lim_{\hat{k} \rightarrow 0} \left\{ [f'(\hat{k}) - (n + g + \delta)]\hat{k} - \hat{c} \right\} < 0,$$

since we are before the peak of the  $k$ -loci,  $\hat{k}_S, \hat{k} < 0$ , and  $\hat{c} > 0$ . Since  $\hat{k} < 0$ , this means that  $\hat{k}$  decelerates at an increasing pace, so asymptoting toward  $\hat{k} = 0$  can't happen.

**Candidate solution converging to  $\hat{c} = 0$ ?** In this case, any other paths are either infeasible or dominated by a path s.t.  $\hat{k} \rightarrow \hat{k}_u$ . To not violate the TVC, it must be that the marginal utility ( $\lambda = u'(\hat{c}_t)$ ) grows at a slower rate than  $\hat{\rho}$ . But from (2b), the growth rate of MU as it approaches  $\hat{k}_u$  is

$$\lim_{\hat{k} \rightarrow \hat{k}_u} \frac{\dot{\lambda}_t}{\lambda_t} = \lim_{\hat{k} \rightarrow \hat{k}_u} [-f'(\hat{k}) + n + g + \delta + \hat{\rho}] > \hat{\rho}$$



so the TVC is violated.

**EXERCISE 3** Argue we do not need to consider  $\hat{c}_t \rightarrow \infty$  as  $\hat{k} \rightarrow 0$ .

### 3.2 Local Dynamics

Now that we know the only possible trajectory converges toward the steady state, we need to show two more things to show uniqueness: i) the choice of  $\hat{c}$  is unique for any  $\hat{k}$ , and ii) the path decelerates (asymptotes) as it approaches the steady state. While there are other ways to show this, here we will show this locally.<sup>3</sup> Take logs on both sides of (2a) and differentiate w.r.t time:

$$\frac{u''(\hat{c}_t)}{u'(\hat{c}_t)} \cdot \dot{\hat{c}}_t = \frac{\dot{\lambda}_t}{\lambda_t} = -[f'(\hat{k}_t) - (n + g + \delta + \hat{\rho})],$$

So for CRRA utility, we can use the CRRA coefficient to write the dynamics of system as<sup>4</sup>

$$\dot{\hat{c}}_t = \frac{1}{\gamma} [f'(\hat{k}_t) - (n + g + \delta + \hat{\rho})] \hat{c}_t \quad (3a)$$

$$\dot{\hat{k}}_t = f(\hat{k}_t) - (n + g + \delta) \hat{k}_t - \hat{c}_t. \quad (3b)$$

Taylor approximate the system (3) around the BGP:

$$\begin{pmatrix} \dot{\hat{c}}_t \\ \dot{\hat{k}}_t \end{pmatrix} \approx \begin{pmatrix} 0 & -\sigma \\ -1 & \hat{\rho} \end{pmatrix} \begin{pmatrix} \hat{c}_t - \hat{c}^* \\ \hat{k}_t - \hat{k}^* \end{pmatrix}$$

where  $\sigma = -\frac{1}{\gamma} f''(k^*) c^* > 0$ . Differentiate the second equation w.r.t. time and substitute back in  $(\dot{\hat{c}}_t, \dot{\hat{k}}_t)$  to get:

$$\ddot{\hat{k}}_t \approx -\dot{\hat{c}}_t + \hat{\rho} \dot{\hat{k}}_t = \sigma (\hat{k}_t - \hat{k}^*) + \hat{\rho} \dot{\hat{k}}_t$$

or

$$\ddot{\hat{k}}_t - \hat{\rho} \dot{\hat{k}}_t - \sigma \hat{k}_t \approx -\sigma \hat{k}^*.$$

<sup>3</sup>Since we assume concavity, Inada conditions, and the choice set is convex, the solution is unique.

<sup>4</sup>Note that we didn't need to use the CRRA assumption up to now, except when deriving normalized consumption.

Now solving this second order ODE is easy: get the general and particular solution s.t.

$$\hat{k}_t \approx \hat{k}_t^g + \hat{k}_t^p.$$

Clearly  $\hat{k}_t^p = k^*$ . The general solution is given by  $\hat{k}_t^g \approx C_1 e^{\alpha_1 t} + C_2 e^{\alpha_2 t}$  where  $(\alpha_1, \alpha_2)$  are the solutions to  $\alpha^2 - \hat{\rho}\alpha - \beta = 0$ , i.e.

$$\alpha_{1,2} = \frac{\hat{\rho} \pm \sqrt{\hat{\rho}^2 + 4\sigma}}{2} = \frac{\hat{\rho}}{2} \pm \sqrt{\sigma + \frac{\hat{\rho}^2}{4}}.$$

We see that the two roots are real and distinct (one larger and one smaller than 0). Let  $\alpha_1 < 0 < \alpha_2$ . This implies that locally around the BGP the dynamic system is saddle-path stable, i.e., there is a unique stable manifold leading to the steady state because for any value other than  $C_2 = 0$ ,  $\hat{k}_t$  will explode, violating the TVC. Hence

$$\hat{k}_t \approx \hat{k}^* + C_1 e^{\alpha_1 t},$$

and  $C_1$  is pinned down by the initial condition  $\hat{k}_0$ , so finally

$$\hat{k}_t \approx \hat{k}^* + (\hat{k}_0 - \hat{k}^*) e^{\alpha_1 t},$$

and the corresponding solution for  $\hat{c}_t$  can be found from

$$\begin{aligned} \dot{\hat{k}}_t &\approx -(\hat{c}_t - \hat{c}^*) + \hat{\rho}(\hat{k}_t - \hat{k}^*) \\ \Rightarrow \hat{c}_t &\approx \hat{c}^* + \hat{\rho}(\hat{k}_t - \hat{k}^*) - \dot{\hat{k}}_t \\ &\approx \hat{c}^* + (\hat{\rho} - \alpha_1)(\hat{k}_0 - \hat{k}^*) e^{\alpha_1 t}. \end{aligned}$$

Note that the speed of convergence is determined by  $|\alpha_1|$ , which is increasing in the elasticity of intertemporal substitution  $1/\gamma$  and decreasing in the effective discount rate  $\hat{\rho}$ . So the more the people are willing to substitute toward the future they accumulate faster, and the more they discount they accumulate slower.

**EXERCISE 4** Note that we did not need to assume any parametric form for the production function. Show that for a BGP to exist,  $u(c)$  must at least converge to CRRA in the limit. To show this, re-solve the RCK model by dividing variables by  $L$  **but not by**  $A$ . Is this assumption required even if  $g = 0$ ?