Spooky Boundaries at a Distance: Technical Appendix

Mahdi Ebrahimi Kahou[1] Jesús Fernández-Villaverde[2]
Sebastián Gómez-Cardona[3] Jesse Perla[4] Jan Rosa[4]

August 12, 2024

Appendix A Robustness

This appendix contains additional robustness results for the neoclassical growth model of Section 4 in the main text.

A.1 Sparse Grids

In our baseline example, we choose $\mathcal{D} \equiv \{0, \dots 29\}$ and minimize equation (13) to find a $k_\theta(t)$ where $|\theta| \approx 40,000$. Alternatively, we use a sparser set of grid points and interpolate when $t \notin \mathcal{D}$. In particular, consider a grid with more points close to the area with high curvature and fewer closer to the steady state, $\mathcal{D}^{\text{Sparse 1}} \equiv \{0, 1, 2, 4, 6, 8, 12, 16, 20, 24, 29\}$, and another grid with fewer points spread evenly over the domain, $\mathcal{D}^{\text{Sparse 2}} \equiv \{0, 1, 4, 8, 12, 16, 20, 24, 29\}$.

Figure A.1 shows the results of these two experiments for an ensemble of 100 initial conditions. The left panel compares the benchmark solution, $k(t)$, relative to the $k_\theta(t)$ for $\mathcal{D}^{\text{Sparse 1}}$ and $\mathcal{D}^{\text{Sparse 2}}$. The right panel compares the benchmark $c(t)$ against the corresponding $c_\theta(t)$. In both cases, the shaded areas show the 10th and the 90th percentiles.

The distribution of the relative error of $k_\theta(t)$ is small, even in the extrapolation region. In the case of $c_\theta(t)$, the error is so small that the 10th and 90th percentile ranges are not visible. This experiment establishes that we can achieve very accurate solutions with sparse grids, even though

[1]Bowdoin College, [2]University of Pennsylvania, [3]Morningstar, and [4]University of British Columbia, Vancouver School of Economics.

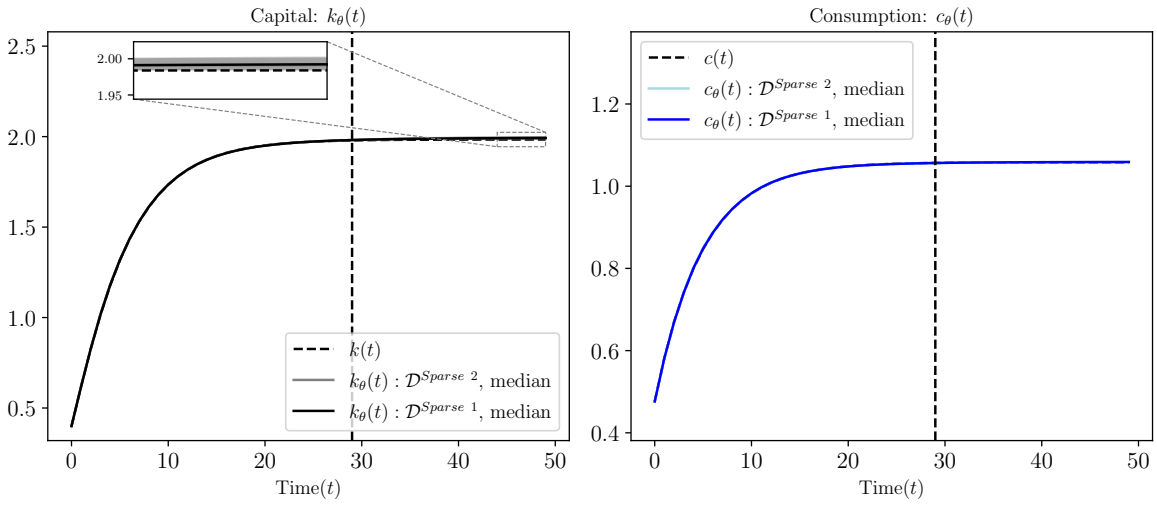

Figure A.1: Solutions to equation (13) with $\mathcal{D}^{\text{Sparse 1}}$ and $\mathcal{D}^{\text{Sparse 2}}$.

the problem remains overparameterized by around four orders of magnitude. ML algorithms do not intrinsically require a large amount of data as long as they have a strong inductive bias.

A.2 Solving on a Short Horizon

A challenge in solving for transition dynamics of models with classic algorithms, such as shooting methods, is the difficulty in choosing the T at which point the solution is close to a steady state. If T is too small, we move toward the steady state too quickly. If T is too large, numerical instabilities can accumulate as the solution iterates forward. Choosing the value of T is an art and requires a good prior on the speed of convergence for a particular model.

To test whether this concern holds with our methods, we solve our model by minimizing equation (13) with the same $\mathcal{H}(\Theta)$, but choose $\mathcal{D} \equiv \{0, 1, \cdots, 9\}$. Not only are there few grid points, but the $t_N = 9$ is far below the point of convergence to the steady state.

Figure A.2 shows the results of this experiment for an ensemble of 100 initial conditions. The left panel shows the median of the approximate capital paths, denoted by $k_\theta(t)$ and the benchmark solution. The right panel shows the median of the approximate consumption paths, denoted by $c_\theta(t)$ and the benchmark solution. The shaded areas represent the 10th and 90th percentiles.

The conclusion is that for the short- to medium-run dynamics, the solutions are very accurate, and the lack of grid points close to the steady state does not feed back to large errors in the short run (as it would with a shooting method). The extrapolation errors are larger than in the baseline

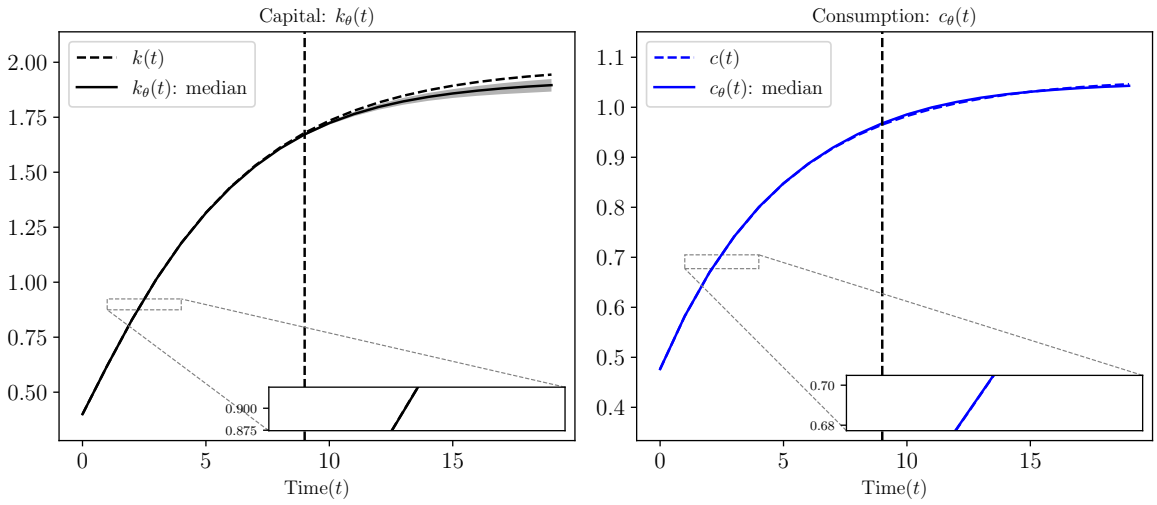

Figure A.2: Solutions to equation (13) with $\mathcal{D} \equiv \{0, 1, \cdots, 9\}$.

case, but getting the long run right was not the goal of the exercise. As discussed, extrapolating and simulating to the steady state is dangerous in general because these solutions are not provably stable. This experiment suggests that the ML methods relying on the inductive bias are not very sensitive to choosing data close to the steady state as long as they are not used to extrapolate too far out of the sample.

A.3 Learning the Scaling Factor

When designing the $\mathcal{H}(\Theta)$ with a BGP, we added in a learnable rescaling: $k_\theta(t) = \exp(\phi t)\text{NN}(t, \theta_{\text{NN}})$, where $\theta \equiv \{\phi, \theta_{\text{NN}}\}$. Given a $\mathcal{D}$ with a large maximum value t_N, the min-norm solution for $\text{NN}(t; \theta_{\text{NN}})$ is achieved by setting $\phi = \log(1 + g)$—at which point $\text{NN}(t; \theta_{\text{NN}})$ could be non-explosive. However, if t_N is relatively small, then we would not expect the approximation to exactly choose the $\phi = \log(1 + g)$ case. A smaller ϕ might yield a lower norm $\text{NN}(t; \theta_{\text{NN}})$ for interpolating a particular $\mathcal{D}$. How well, then, does the algorithm learn g?

Taking the results of Figure 3, which generated solutions using 100 initial conditions, Figure A.3 plots a histogram of the approximated ϕ and compares them to the true growth rate, $g = 0.02$. The results show that the min-norm is biased toward smaller growth rates, as we might expect. However, the solutions in Figure 3 are still extremely accurate. The variations in ϕ within Figure A.3 have compensated changes to $\text{NN}(t; \theta_{\text{NN}})$. A very accurate approximation of the growth rate is not necessary to achieve accurate short- and medium-run dynamics.

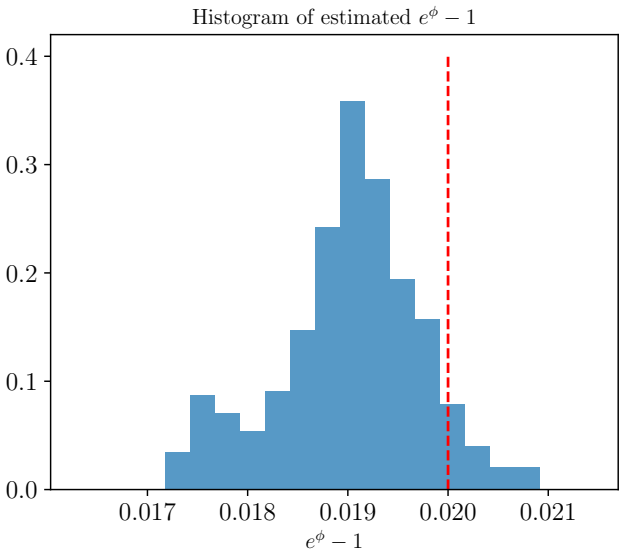

Figure A.3: The distribution of the learned $e^{\phi} - 1$ for the ensemble of 100 seeds used in Figure 3; $g = 0.02$, shown as the dashed line.

A.4 Learning a Misspecified $\mathcal{H}(\Theta)$

In Figure 3, we used economic insights to choose a $\mathcal{H}(\Theta)$ that included a term for exponential growth. Is it still helpful to suggest problem structure when designing $\mathcal{H}(\Theta)$ if the suggestion is misspecified?

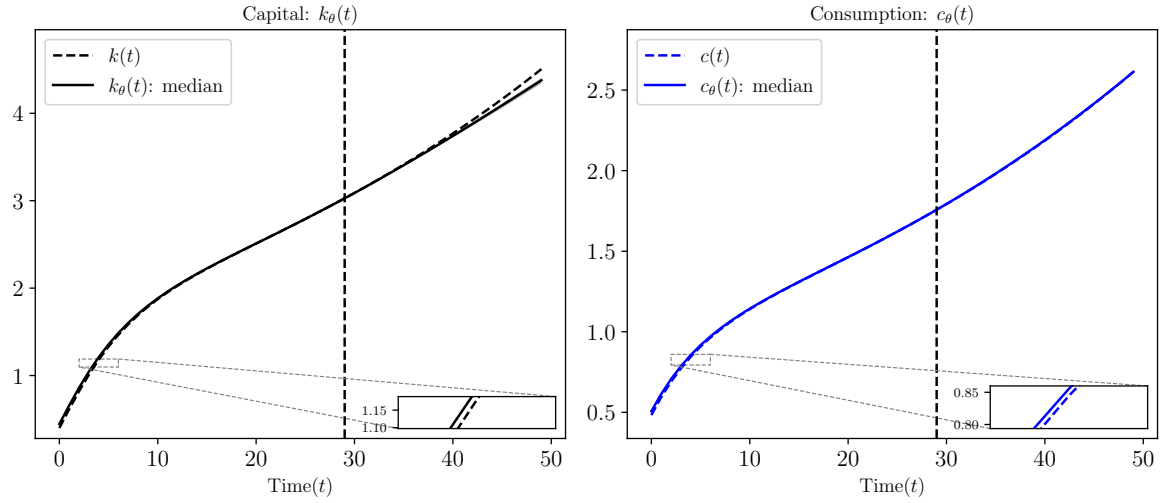

Figure A.4: Solutions to problem (13) with the misspecified $k_\theta(t) = t \cdot \mathrm{NN}(t;\theta) + k_0$ and $g = 0.02$.

To analyze this case, we solve a version where the scaling is assumed to be linear rather than exponential. In particular, $k_\theta(t) = t \cdot \mathrm{NN}(t;\theta) + k_0$. The linear scaling allows some degree of

growth but as $t_N \to \infty$, the $\text{NN}(t;\theta)$ would still need to have an infinite norm in order to capture the true dynamics of the BGP.

Figure A.4 displays the solutions to problem (13) with this specification for 100 initial conditions. The left panel shows the benchmark and the median of the solution for capital, while the right panel does the same for consumption. Although the 10th and 90th percentiles are included, they are so close to each other that they remain indistinguishable even after zooming in.

Compared to the well-specified case of Figure 3, the long-run extrapolation slowly diverges (and would continue to do so for any finite t_N), but this does not cause any issues for the short- and medium-run dynamics.

A.5 Function Norms and the Transversality Condition

Section 4 characterized the set of functions fulfilling the Euler equation and resource constraints as (i) $k_{\max}(t), c_{\max}(t)$, with steady states $k^*_{\max}$ such that $f'(k^*_{\max}) = \delta$ and $c^*_{\max} = 0$; and (ii) $k(t), c(t)$ with interior steady states k^* and c^*. The transversality condition (12) eliminated the first solution to prevent the marginal utility of consumption, $u'(c) = c^{-1}$, from becoming infinite.

When relying on the inductive bias of the function norms in lieu of the transversality condition, we must argue that $||k_{\max}||_\psi > ||k||_\psi$ for a large class of norms, ψ. To see this, Figure A.5 plots the two solutions to the under-determined system. The blue curves show a set of capital, consumption, and marginal utility paths, denoted respectively by $k_{\max}(t)$, $c_{\max}(t)$, and $u'(c_{\max}(t))$, that violate the transversality condition. The black curves show the optimal paths that satisfy the transversality condition and that eventually converge to k^*, c^*. Focusing on the left panel, we see that the path of the $k_{\max}(t)$ function has much steeper changes than that of $k(t)$. Therefore, for a large class of norms and semi-norms, which penalize either the average level or gradients, we have $\|k_{\max}\|_\psi > \|k\|_\psi$.

The middle and right panels of Figure A.5 also provide intuition on why these methods can be fragile to the right formulation. While $\|k_{\max}\|_\psi > \|k\|_\psi$ for a large norm given the big spread between k^* and $k^*_{\max}$, this is not the case for $c(t)$. If a norm penalized the gradients (e.g., $\int_0^T |c'(t)|dt$), then the norms of $||c_{\max}||_\psi$ and $||c||_\psi$ would be similar. If the level enters the norm, it may even bias the solution toward the wrong answer (i.e., where $c^*_{\max} = 0$). The right panel shows the other extreme, where using the marginal utility makes an even starker difference between the two solutions. The general advice, true for both the sequential formulation and the state-

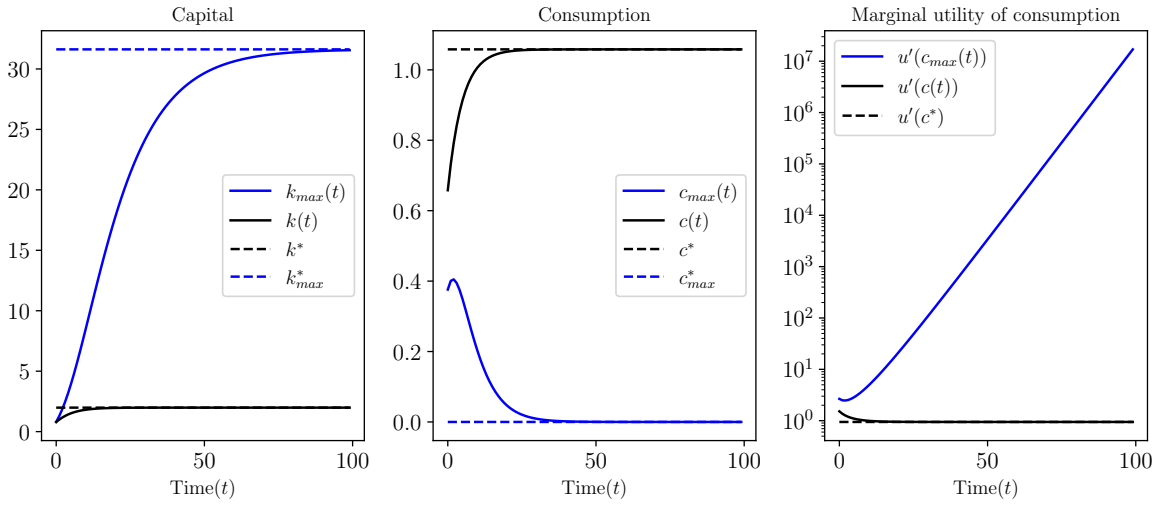

Figure A.5: Comparison between the optimal solution and those violating the transversality condition.

space version, is that it is best to approximate functions that are most explosive if they violate transversality. Co-state variables are the best; state variables often work well, but jump variables are often bounded in a way that makes min-norm solutions harder to disentangle. We see that a similar issue holds in Appendix B.3 for the recursive formulation.

Appendix B State-Space Formulation

This appendix describes the recursive state-space formulation of the neoclassical growth model, in contrast to the sequence-space baseline of Section 4. Inductive bias will serve a similar role in providing a sufficiency condition for transversality, but it will involve norms of the policy functions rather than the trajectories themselves.

B.1 Model

For the state-space $(k, z) \in \mathbb{R}^2_+$, equations (10) and (11) become:

$$u'(c) = u'(c')\beta\left[z'^{1-\alpha} f'(k') + 1 - \delta\right] \tag{B.1}$$

$$k' = z^{1-\alpha} f(k) + (1-\delta)k - c \tag{B.2}$$

where k', c', and z' are the next period capital, consumption, and TFP, respectively, and $u(c) = \log c$. All model primitives and parameters remain the same as in the baseline. The transversality condition (12) must hold for all initial conditions in the state-space formulation:

$$0 = \lim_{T\to\infty} \beta^T u'\left(c_T(k_0, z_0)\right) k_{T+1}(k_0, z_0) \quad \text{for all } (k_0, z_0) \in \mathbb{R}^2_+. \tag{B.3}$$

In this notation, $k_{T+1}(k_0, z_0)$ requires iterating the $k'(\cdot, \cdot)$ policy and $z' = (1+g)z$ law of motion $T+1$ times from (k_0, z_0). Consumption, $c_T(k_0, z_0)$, is found by first iterating to find (k_T, z_T) and then using equation (B.2) to calculate $c_T = z_T^{1-\alpha} f(k_T) + (1-\delta)k_T - k'(k_T, z_T)$.

Transversality with classic methods. The iteration of the policy $k'(\cdot, \cdot)$ in equation (B.3) links stability and transversality. If $k'(\cdot, \cdot)$ was explosive —e.g., $|\nabla_k k'(k, z)| > 1$ for k and z above some threshold— capital would explode until it asymptotically approached the capital maximizing the BGP (or $k^*_{\max}$ if $g = 0$) via equation (B.1). This, in turn, would lead to an infinite marginal utility of consumption in equation (B.3), violating transversality.

In practice, classical methods do not apply the transversality condition as a limit and instead enforce it indirectly in several ways:

- For sequence-space methods, a steady state is found (perhaps after detrending the BGP), which is then used as a terminal boundary condition with shooting methods. Those approaches implicitly use the transversality condition when solving for the correct steady state.

- Linear rational expectations models and LQ control, such as those in Blanchard and Kahn (1980) and Klein (2000), select the non-explosive root via spectral methods.

- With global solution methods, such as projection and collocation, the transversality is implicitly fulfilled by restricting the domain for the state space. For example, in the growth model, we might approximate with Chebyshev polynomials on a compact hypercube on $[k_{\min}, \bar{k}] \times [z_{\min}, \bar{z}]$. If we chose $\bar{k} < k^*_{\max}$ and $k_{\min} < k^*$, then policy functions violating transversality are rejected since they cannot fulfill the Euler equation before hitting corners. Alternatively, by bounding $c \geq c_{\min} > 0$, algorithms implicitly reject functions that fail transversality by bounding the marginal utility of consumption, $u'(c) \leq u'(c_{\min}) < \infty$.

In low dimensions, where we have a strong prior on the relevant regions of the state space, economists can artfully tinker to ensure that a compact hypercube is placed at the appropriate location and does not contain the solutions violating transversality. Moreover, by plotting the dynamics of the model, we can see when simulations diverge (see Fernández-Villaverde et al., 2016, p.10).[1]

However, this process is not feasible in high dimensions since we cannot constrain ourselves to a compact hypercube and may not have a good prior on the location of a steady state. Even evaluating whether transversality conditions are fulfilled for a given policy is computationally infeasible since it requires iterating the policy function for all initial conditions.

Notice here the connection to the issue of stability in ML methods. Simple forward iterations can accumulate numerical errors and be numerically unstable when the solution is only "approximately" stable. This phenomenon appears even in small models.

B.2 Min-Norm Solution

We approximate the capital policy, $k'_\theta(\cdot,\cdot) \in \mathcal{H}(\Theta)$, using a highly parameterized neural network. Choose $\mathcal{D} \subset \mathbb{R}^2_+$ with N points and minimize the equivalent of equation (13):

$$\min_{\theta\in\Theta} \frac{1}{N} \sum_{(k,z)\in\mathcal{D}} \left[\frac{u'\big(c(k,z;k'_\theta)\big)}{u'\big(c(k'_\theta(k,z),(1+g)z;k'_\theta)\big)} - \beta\big[(1+g)zf'\big(k'_\theta(k,z)\big)+1-\delta\big] \right]^2. \tag{B.4}$$

Consumption is defined through the feasibility constraint for a given policy for capital $k'_\theta(\cdot,\cdot)$:

$$c\big(k,z;k'_\theta\big) \equiv f(k) + (1-\delta)k - k'_\theta(k,z). \tag{B.5}$$

Following the interpretation of ERM as a minimum norm solution, we can think of solutions to equation (B.4) as finding:

$$\min_{k'_\theta\in\mathcal{H}(\Theta)} ||k'_\theta||_\psi \tag{B.6}$$

$$\text{s.t. } \frac{u'\big(c(k,z;k'_\theta)\big)}{u'\big(c(k'_\theta(k,z),(1+g)z;k'_\theta)\big)} = \beta\big[(1+g)zf'\big(k'_\theta(k,z)\big)+1-\delta\big], \quad \text{for all } (k,z)\in\mathcal{D}. \tag{B.7}$$

[1] This is part of the appeal of perturbative solutions, which are provably stable even in high dimensions (if properly pruned).

The norm in problem (B.6) typically depends on gradients due to its bias toward flat solutions. For example, it might have properties similar to those of a Sobolev norm $||k'_\theta||^2_{W^{1,2}} \equiv \int ||\nabla k'_\theta(k,z)||_2^2 dF(k,z)$ for some measure F over the state space, or on a compact subset of the domain.

To informally argue why this bias would choose the non-explosive solution, consider iterating the policy function $k_{t+1} = k'_\theta(k_t, z_t)$. A bias toward solutions with smaller gradients with $|\nabla_k k'_\theta(k,z)| < 1$ for large k will lead to policies that have smaller changes in capital, $k_{t+1} - k_t$. If a steady state exists, it will reach the $k_t \approx k'_\theta(k_t, z_t)$ fixed point. Iterating forward with the policy, the bias leads to trajectories that fulfill the transversality condition (B.3). In Appendix B.3. we demonstrate this by plotting the $k'_\theta(\cdot,\cdot)$ for the trajectories that fulfill the Euler equation with and without transversality.

Results. We solve the minimization problem (B.4) for $\beta = 0.9$, $\alpha = 0.33$, $\delta = 0.1$, $g = 0$, $z_0 = 1$, and $k_0 = 0.4$. In our baseline case, $\mathcal{D}$ is a uniform grid of 16 points between $k_1 = 0.8$ and $k_{N_k} = 2.5$. When $g \neq 0$, we can use a grid $\mathcal{D} \equiv \{k_1, \cdots, k_{N_k}\} \times \{z_1, \cdots, z_{N_z}\}$ of $N = N_z \times N_k$ total points, but the methods could use sampled or simulated points in the state-space. The design of $\mathcal{H}(\Theta)$ is a neural network $\text{NN}(k, z; \theta)$ identical to the sequential version of the model, except that it takes two inputs (k, z) rather than the univariate t. As before, we solve with the L-BFGS optimization algorithm, which is fast and requires little tuning.

Figure B.1 shows the median of solutions for capital (top row) and consumption (bottom row) for an ensemble of 100 initial conditions. The consumption path $\tilde{c}(t)$ is calculated with equation (B.5) given the trajectory of the state space. The benchmark solutions, $k(t)$ and $c(t)$, are obtained using value function iteration. The left panels show the median of the approximate capital, $\hat{k}(t)$, and consumption, $\hat{c}(t)$, paths, along with the benchmark solutions (i.e., $\hat{k}(t)$ and $\hat{c}(t)$ are the results of iterating the solution from a particular initial condition). The right panels show the median of the relative errors for capital, $\varepsilon_k(t) \equiv (\hat{k}(t) - k(t))/k(t)$, and consumption, $\varepsilon_c(t) \equiv (\hat{c}(t) - c(t))/c(t)$. The shaded regions show the 10th and 90th percentiles. The gray region in the top-left panel shows the interpolation region, defined as the convex hull of $\mathcal{D}$. The dashed parts of the curves show the median of the relative errors in the extrapolation region. The shaded regions show the 10th and 90th percentiles of the solutions for the 100 random seeds for optimization of θ.

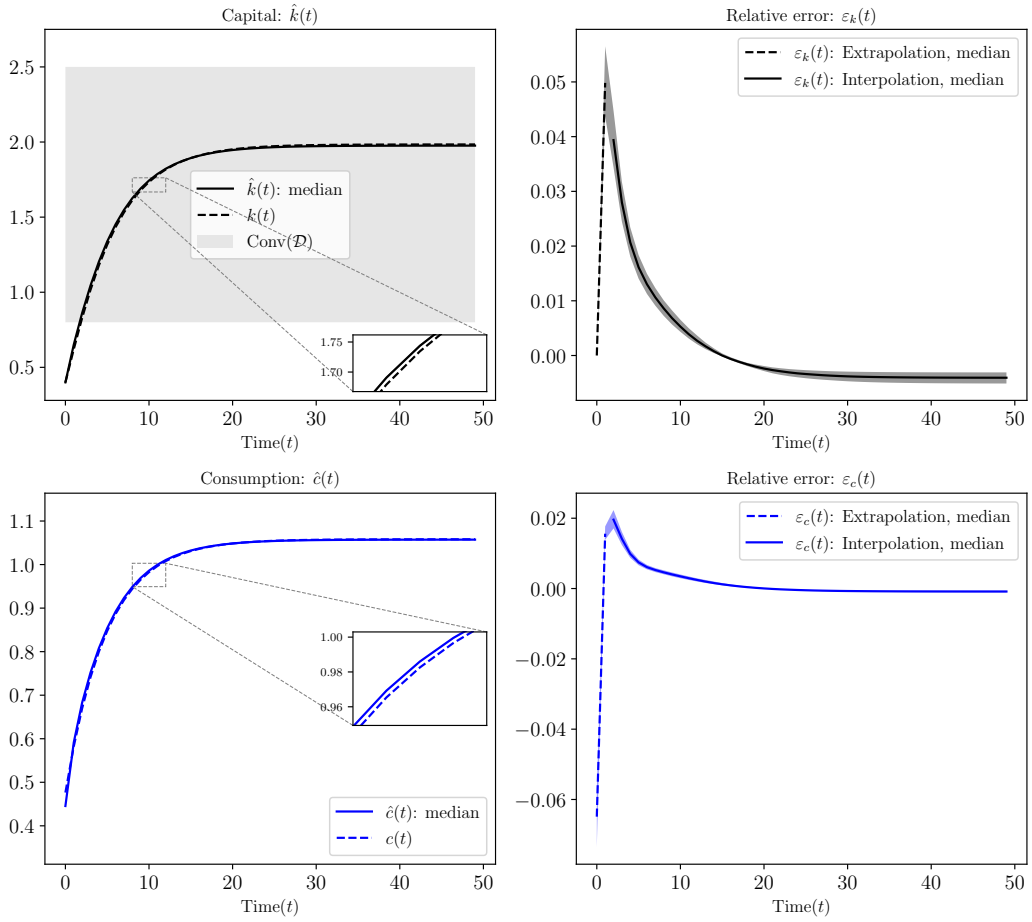

Figure B.1: Solutions obtained by solving problem (B.4) for $g = 0$.

The results show that the inductive bias rules out solutions that violate the transversality conditions in all cases and achieves a good approximation despite only using 16 data points. Even when k_0 is outside the minimum value of $\mathcal{D}$, the errors are small. An inductive bias leads to good generalization behavior even outside of the convex hull of Conv($\mathcal{D}$).

BGP. Since we know that the solution will be homothetic when $g = 0.02$, we now design $\mathcal{H}(\Theta)$ as $k'_\theta(k, z) \equiv z \cdot \text{NN}(k/z, z; \theta)$. We set $\mathcal{D}$ as the cartesian product of 16 points in $[0.8, 3.5]$ for capital with 8 points in $[0.8, 1.8]$. As before, using a small $\mathcal{D}$ highlights the strength of the inductive bias. This implementation minimizes the problem (B.4) with different $\mathcal{H}(\Theta)$ and $\mathcal{D}$ for 100 seeds on the initial condition for the optimizer.[2]

Figure B.2 shows the results for a simulated trajectory from $k_0 = 0.4$ and $z_0 = 1$ and compares

[2]In the exactly homothetic case, we could further simplify this to a univariate NN($k/z; \theta$), but we leave in the z parameter as a check for cases that are almost homothetic and as a further check that the inductive bias avoids overfitting.

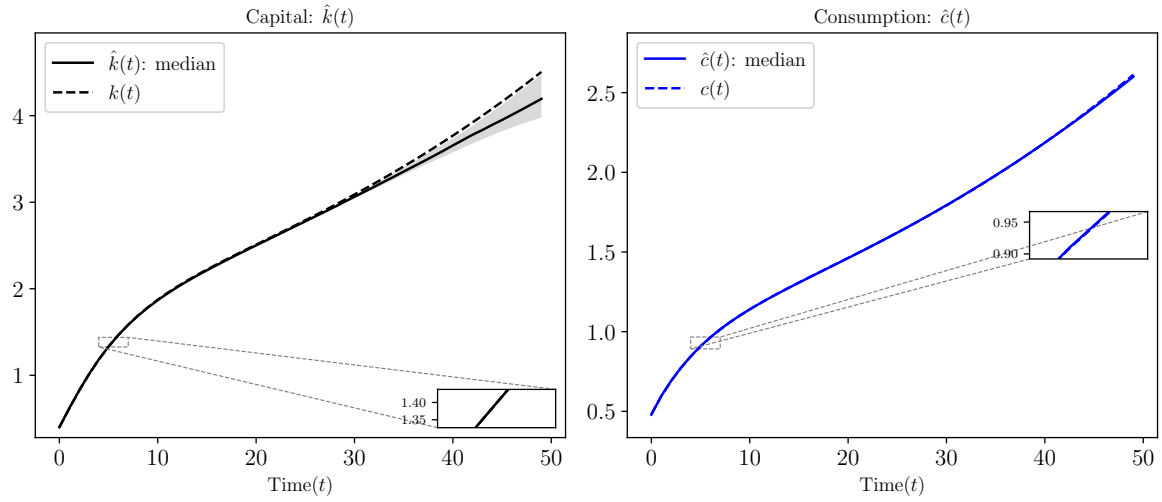

Figure B.2: Solutions to problem B.4 for $g = 0.02$.

the dynamics given the benchmark solution. The left panel shows the median of the approximate capital path, denoted by $\hat{k}(t)$. The right panel shows the median of the approximate consumption path, denoted by $\hat{c}(t)$. The shaded regions show the 10th and 90th percentiles.

The results indicate that, even in the case of growing TFP, the solution is very accurate in the short run, and the differences relative to the benchmark are difficult to see even after zooming in on the graph. The long-run extrapolation is less accurate than in the benchmark (where we could manually rescale due to homotheticity). In other words, we can obtain very accurate short- and medium-run solutions, even though the initial condition for capital lies outside the interpolation region.

B.3 Failures of Euler Residuals Minimization

Appendix A.5 discussed the importance of choosing the right formulation of the problem to ensure that the inductive bias toward min-norm solutions would select the solution that fulfills transversality. This issue is often even more stark in state-space formulations of the problem. Understanding this phenomenon is especially important before we move toward high-dimensional problems in macroeconomics, where failures of transversality are less obvious.

We demonstrate this problem by comparing an equivalent formulation of the neoclassical growth model where we approximate $c_\theta(k, z)$ to our previous results in Figures B.1 and B.2. The inductive bias toward min-norm solutions will consistently choose the wrong solution that violates

transversality.

Let $z = 1$ and $g = 0$ for simplicity, approximate $c_\theta(k) \in \mathcal{H}(\Theta)$ with a neural network, and implicitly define the investment choice as $k'\big(k; c_\theta\big) \equiv f(k) + (1-\delta)k - c_\theta(k)$. The equivalent to the ERM objective function (B.4) becomes:

$$\min_{\theta \in \Theta} \frac{1}{N} \sum_{k \in \mathcal{D}} \underbrace{\left[\frac{u'\big(c_\theta(k)\big)}{u'\Big(c_\theta\big(k'(k; c_\theta)\big)\Big)} - \beta \left[f'\big(k'(k; c_\theta)\big) + 1 - \delta \right] \right]}_{\equiv \varepsilon_E^c(k)}^2 . \tag{B.8}$$

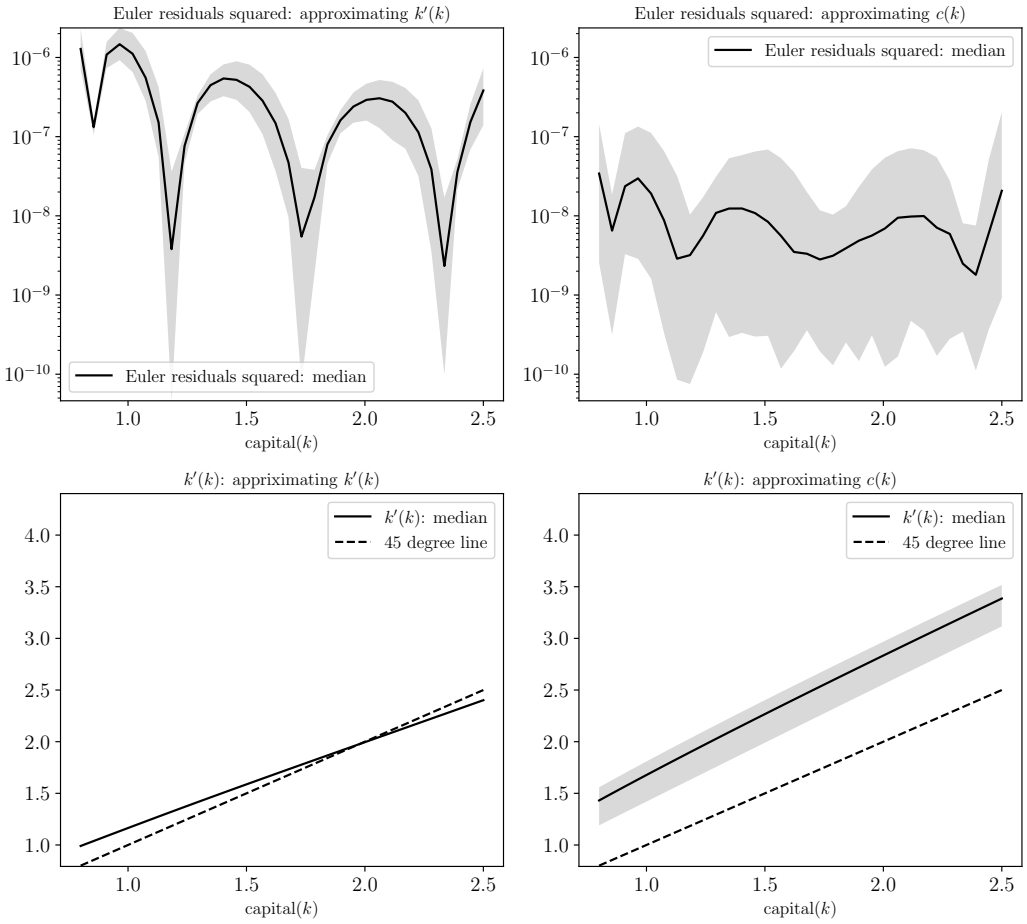

Figure B.3: Comparison between approximating the policy function for capital $k'(k)$ vs. the consumption function $c(k)$ with a deep neural network.

Figure B.3 shows the comparison between approximating the policy function for capital $k'_\theta(\cdot)$ vs. approximating the consumption function $c(\cdot)$ with a deep neural network.[3] The left panels

[3]Primitives and parameters are identical to our baseline case. Given the parameters, the steady-state solution fulfilling transversality is $k^* \approx 2.0$.

show the results using the baseline k'_θ approximation, as in Figure B.1, but plots the Euler error in the top panel and the policy function $k'_\theta(k)$ in the bottom panel. The $k'(k)$ in the bottom panel crosses the 45-degree line around $k^* \approx 2.0$, which is the closed-form steady state. The top right panel instead plots the square Euler error when approximation the c_θ function, while the bottom right panel plots the implied $k'(k; c_\theta)$ policy from $k'\big(k; c_\theta\big) \equiv f(k) + (1-\delta)k - c_\theta(k)$. The solid curves show the medians, and the shaded regions show the 10th and 90th percentiles over 100 different seeds.

Approximating the consumption functions with a neural network leads to solutions that violate the transversality condition. Given the c_θ approximation, the squared Euler residual error, $\varepsilon^c_E(k)$, is defined in equation (B.8) and when approximating with k'_θ, an equivalent definition of $\varepsilon^k_E(k)$ exists from equation (B.4). The Euler errors in both cases are very small and close to numerical precision, so the optimizer has a solution that interpolates the Euler equation and implicitly fulfills the resource constraint on $\mathcal{D}$. If anything, the Euler errors are smaller for the c_θ approximation. However, the bottom right panel does not have the $k'(k)$ intersecting the 45-degree line. It has chosen a c_θ such that $\nabla_k k'(k; c_\theta) > 1$ for all k. This leads to explosive $\tilde{k}(t)$ trajectories and fails the transversality condition in all cases.

The reason why the inductive bias works in the wrong direction in this formulation can be seen if we return to the middle panel of Figure A.5. The consumption trajectory that violates transversality converges to 0 and would have a smaller norm for many ψ that penalizes the level of the function. Even without penalizing the level, the slope of the solution fulfilling transversality is not systematically smaller in absolute value.

To conclude, low Euler (or value-function) errors are insufficient to ensure that an ML algorithm has successfully solved the problem, and inductive bias with the wrong problem formulation might systematically choose the policy that violates transversality. The broad advice is to ensure that the problem is formulated in a way that violations of transversality lead to explosive behavior (e.g., diverging states or formulating in terms of the marginal utility or co-state variables).

References

Blanchard, O. J. and C. M. Kahn (1980): "The solution of linear difference models under rational expectations," *Econometrica*, 48, 1305–1311.

Fernández-Villaverde, J., J. Rubio-Ramírez, and F. Schorfheide (2016): "Solution and estimation methods for DSGE models," in *Handbook of Macroeconomics*, ed. by J. B. Taylor and H. Uhlig, Elsevier, vol. 2, 527–724.

Klein, P. (2000): "Using the generalized Schur form to solve a multivariate linear rational expectations model," *Journal of Economic Dynamics and Control*, 24, 1405–1423.