Supplement to "Solving Models of Economic Dynamics with Ridgeless Kernel Regressions"

MAHDI EBRAHIMI KAHOU
Department of Economics, Bowdoin College

JESSE PERLA
Vancouver School of Economics, University of British Columbia

GEOFF PLEISS
Department of Statistics, University of British Columbia and Vector Institute

This supplement contains the details of the robustness checks on our algorithm, additional applications alongside those presented in the paper, and the derivations for the recursive and stochastic formulations.

1. Uniqueness of DAE Solutions

We relate the existence and uniqueness of the trajectory assumed in Main Paper Assumption 2 to the standard theory of differential-algebraic equations (DAEs) in Brenan et al. (1996), which supplies sufficient conditions on the primitives. The theorems of the main paper use Main Paper Assumptions 1 and 2 directly.

Index one. If $\boldsymbol{H}$ is continuously differentiable with $\nabla_{\boldsymbol{y}}\boldsymbol{H}$ invertible on $\mathbb{V}$, the implicit function theorem determines $\boldsymbol{y}$ locally as a continuously differentiable function of $(\boldsymbol{x},\boldsymbol{\mu})$, and substituting it into Main Paper Equations (1) and (2) reduces the DAE to an ordinary differential equation (ODE) in $(\boldsymbol{x},\boldsymbol{\mu})$ with a unique local solution from each consistent $(\boldsymbol{x}_0,\boldsymbol{\mu}_0,\boldsymbol{y}_0)$. This is the semi-explicit index-one case of Brenan et al. (1996). For a given $\boldsymbol{x}_0$, Main Paper Assumption 2(i) asks that exactly one such trajectory satisfies the transversality condition Main Paper Equation (5).

Singular $\nabla_{\boldsymbol{y}}\boldsymbol{H}$. A singular $\nabla_{\boldsymbol{y}}\boldsymbol{H}$ does not preclude uniqueness. When a jump variable such as the world interest rate of Main Paper Section 4.5 enters only through intertemporal conditions, it is absent from the algebraic equation. Differentiating that equation along solutions until the variable appears, by hand or with the Pantelides algorithm (Pantelides, 1988), gives an index-one system. The index-one system enforces only the differentiated equation, so the original equation must hold at $t=0$ as an additional initial condition. With it, the two systems have the same solutions. The multi-country model of Main Paper Section 4.5 is reduced in this way in Section 4.2. The kernel method can use either formulation. In our examples its solutions satisfy the differentiated constraints approximately without imposing them.

1.1 *Jump-Variable Multiplicity*

Multiple steady states do not by themselves conflict with Main Paper Assumption 2. In the small open economy of Main Paper Section 4.4, the limiting state depends on $\boldsymbol{x}_0$, while the trajectory is unique for each initial condition.

Jump-variable multiplicity is the failure of Main Paper Assumption 2(i) itself. More than one trajectory satisfies Main Paper Equations (1) to (5) for the same $\boldsymbol{x}_0$, so the initial jump variables and co-states are not pinned down. This can arise from coordination failures, rational-expectations indeterminacy, or an unnormalized nominal price scale. In the last case a normalization removes it. The growth model with multiple steady states in Main Paper Section 4.2 can have candidate paths toward different steady states from the same $\boldsymbol{x}_0$. Each candidate is the unique trajectory satisfying Main Paper Equation (5) once $\mathbb{V}$ is restricted to a region containing the path and its steady state, and our theorems apply to each such restricted problem when the remaining assumptions hold there.

2. Robustness

This section provides robustness checks and an exploration of data efficiency for the neoclassical growth model.

2.1 *Sparse training data and data efficiency*

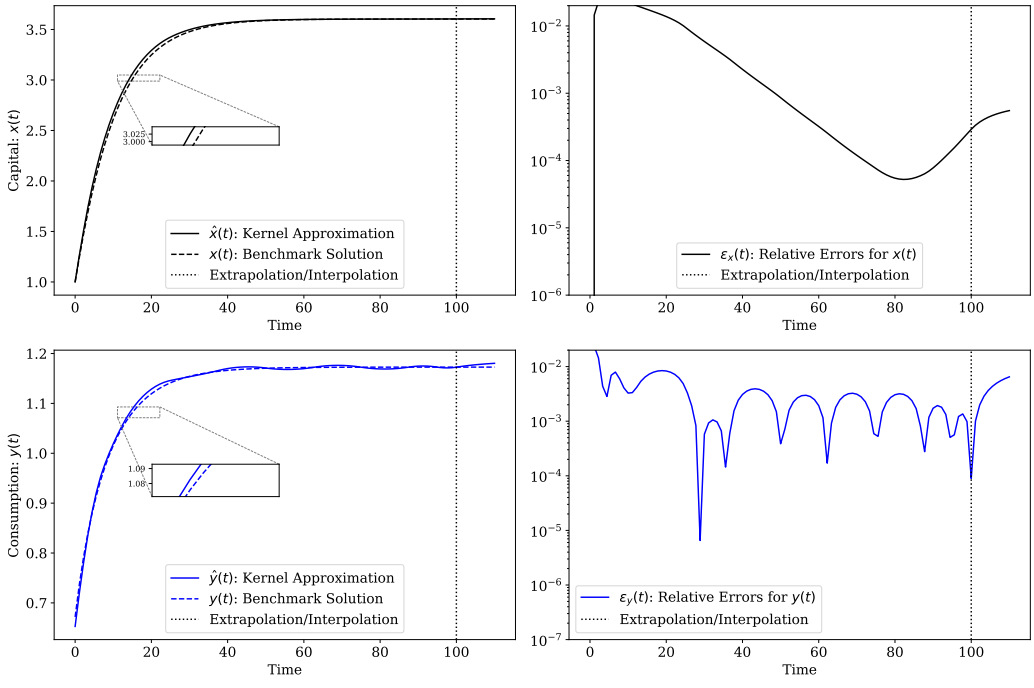


FIGURE 1. Solution of the neoclassical growth model (Main Paper Equations (6) to (8)) without imposing the transversality condition (Main Paper Equation (10)), for sparse training data $\mathcal{D} := \{0, 2.5, 7.5, 12.5, 25, 37.5, 50, 62.5, 75, 87.5, 95, 100\}$. The top panels show the results for capital, $x(t)$, and associated relative errors, and the bottom two show the results for consumption, $y(t)$, and associated relative errors. Accurate solutions can be obtained even with a very sparse $\mathcal{D}$.

Since the number of coefficients grows linearly with the number of training points, the cardinality of $\mathcal{D}$ becomes an impediment in higher dimensions with many state and jump variables, so accuracy with sparse training data matters.

Figure 1 shows the solution of the neoclassical growth model (i.e., Main Paper Equations (6) to (8)) with $\mathcal{D} := \{0, 2.5, 7.5, 12.5, 25, 37.5, 50, 62.5, 75, 87.5, 95, 100\}$, which spans the full horizon with twelve nonuniformly spaced points, denser early in the transition. The approximate paths $\hat{x}(t)$ and $\hat{y}(t)$ and their relative errors $\varepsilon_x(t)$ and $\varepsilon_y(t)$ against the benchmark remain small.

2.2 *Robustness to the choice of the kernel and kernel parameters*

Table 1 shows the approximate solution of the neoclassical growth model, described in Main Paper Equations (6) to (8), for different Matérn kernels and kernel parameters.

The first three rows report the performance of the approximate solutions using three different kernels at the baseline lengthscale. We present the maximum and mean absolute relative errors for both the capital path, $\hat{x}(t)$, and the consumption path, $\hat{y}(t)$. The first row shows the baseline solution using the Matérn kernel with $\nu = \frac{1}{2}$. The second and third rows present results for the Matérn kernels with $\nu = \frac{3}{2}$ and $\nu = \frac{5}{2}$, respectively.

TABLE 1. The robustness of the approximate solutions of the neoclassical growth model (i.e., Main Paper Equations (6) to (8)) is tested using different Matérn kernels, $\nu = \frac{1}{2}, \frac{3}{2}, \frac{5}{2}$, the Gaussian limit $\nu \to \infty$, and lengthscales $\ell = 2, 10, 20$. Entries are the maximum and mean absolute relative errors for capital, $\hat{x}(t)$, and consumption, $\hat{y}(t)$, on the test grid. A "–" entry marks a rejected solution. The Gaussian kernel at large ℓ is rejected because its Gram matrix is ill-conditioned.

ν	ℓ	Max of Rel. Error: $\hat{x}(t)$	Max of Rel. Error: $\hat{y}(t)$	Mean of Rel. Error: $\hat{x}(t)$	Mean of Rel. Error: $\hat{y}(t)$
$1/2$	10	1.5e-03	3.1e-03	1.9e-04	1.1e-04
$3/2$	10	2.3e-04	1.1e-03	2.1e-05	6.7e-05
$5/2$	10	3.6e-05	4.1e-04	3.6e-06	1.2e-05
$1/2$	2	7.0e-03	3.3e-03	1.5e-03	7.7e-04
$1/2$	20	1.8e-03	3.1e-03	2.4e-04	1.3e-04
∞	10	–	–	–	–
∞	2	9.2e-05	2.3e-04	1.5e-05	6.6e-06
∞	20	–	–	–	–

The fourth and fifth rows show the performance of the Matérn-$\frac{1}{2}$ kernel for two different *lengthscales*, $\ell = 2$ and $\ell = 20$, and the last three rows the Gaussian limit $\nu \to \infty$ at the three lengthscales.

Every accepted solution is accurate, and the method is insensitive to the choice of Matérn kernel and lengthscale within the range tested. The Gaussian kernel, whose Gram matrix is ill-conditioned at the two larger lengthscales, fails the acceptance check of Section 6.2 there and is shown as "–".

2.3 *Smaller time horizons: accurate short-run dynamics*

One might suspect that a solution satisfying the transversality condition requires a long horizon in the training data, such as the $\mathcal{D} = \{0, 1, 2, \ldots, 100\}$ of Main Paper Figure 1. This experiment shows that a shorter horizon still gives accurate short-run dynamics.

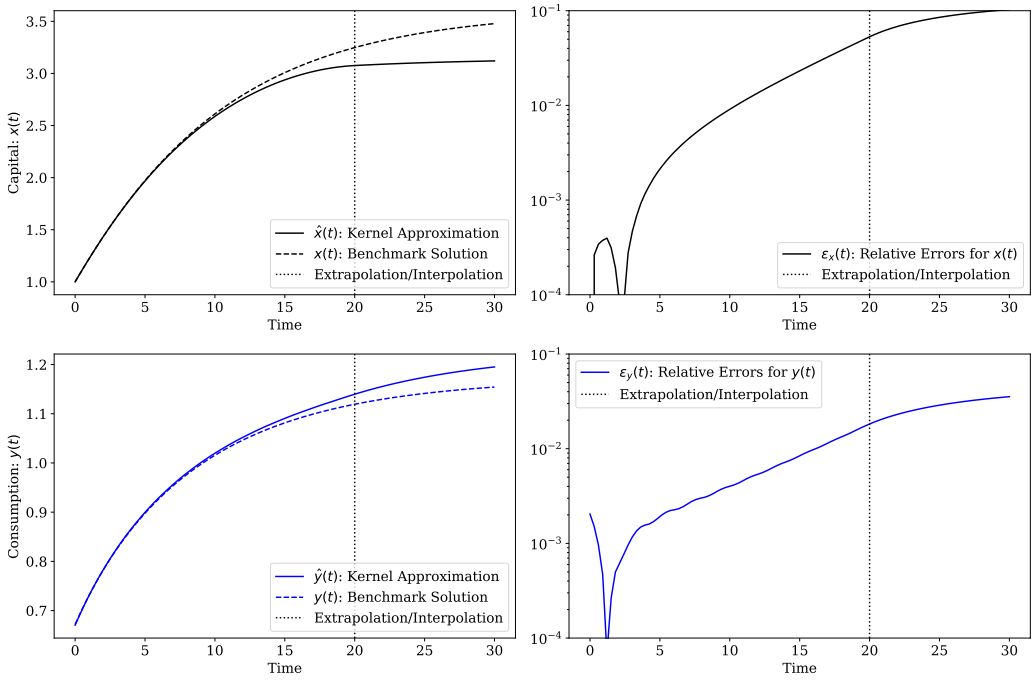


FIGURE 2. Solution of the neoclassical growth model (Main Paper Equations (6) to (8)) without imposing the transversality condition (Main Paper Equation (10)), for training data with a smaller time horizon $\mathcal{D} := \{0, 1, \ldots, 20\}$. The top panels show the results for capital, $x(t)$, and associated relative errors, and the bottom two show the results for consumption, $y(t)$, and associated relative errors. Accurate short-run dynamics are obtained even with a smaller time horizon.

Figure 2 shows the approximate solutions for the neoclassical growth model (i.e., Main Paper Equations (6) to (8)) for training data with a shorter time horizon, defined as $\mathcal{D} := \{0, 1, \ldots, 20\}$. The panels are as in Figure 1.

2.4 *Trend growth without detrending*

Models with balanced growth can be solved in levels, without detrending. Our theorems do not cover growing paths, since they require the solution to stay in a compact box. The intuition still applies. If the paths are rescaled by the correct trend rate, their derivative seminorms stay finite as T grows. Any other rate leaves an exponential trend and a larger norm.

We solve the neoclassical growth model of Main Paper Section 4.1 with Hicks-neutral productivity $Z(t) = e^{(1-a)gt}$, so that output is $Z(t)f(x(t))$ in Main Paper Equation (6) and the marginal product is $Z(t)f'(x(t))$ in Main Paper Equation (7). Each element of the approximation class is rescaled by $e^{\hat{g}t}$ for x and y and by $e^{-\gamma\hat{g}t}$ for the co-state, where γ is the coefficient of relative risk aversion. The neural network (NN) of Section 6.2 rescales both of its outputs by $e^{\hat{g}t}$, with a learned $\hat{g}$ as one more trainable parameter.

The rate $\hat{g}$ is either fixed at g or learned as a free parameter. With $\hat{g}$ fixed, the accuracy matches the baseline. With $\hat{g}$ learned from zero, the kernel finds $\hat{g} \approx 0.027$ for $g = 0.02$ and the solution stays near the saddle path, with mean relative errors of 8.95×10^{-3} for capital and 4.07×10^{-3} for consumption. The NN is more accurate, but only from a positive initial guess for $\hat{g}$. From zero it fails every time. From the positive guess, it finds the rate to within 7.26×10^{-5}, with mean relative errors of 4.61×10^{-7} for capital and 1.58×10^{-7} for consumption.

3. Derivations for the Small Open Economy

The small open economy of Main Paper Section 4.4 sets $r^\star = r$ and $\chi = \delta^{1-\eta}$. Setting $\dot{x}_1 = \dot{\mu}_1 = 0$ in Main Paper Equations (19) and (21) and using Main Paper Equation (24) pins down steady-state capital, q, and investment,

$$\frac{y_2^\star}{x_1^\star} = \delta, \qquad q^\star = \frac{1}{\eta}, \qquad f'(x_1^\star) = \delta + \frac{r^\star}{\eta} \implies x_1^\star = \left(\frac{a}{\delta + r^\star/\eta}\right)^{\frac{1}{1-a}}, \tag{S1}$$

but setting $\dot{x}_2 = 0$ in Main Paper Equation (20),

$$y_1^\star = f(x_1^\star) - \delta x_1^\star + r^\star x_2^\star, \tag{S2}$$

is one equation in two unknowns. The steady states form a continuum indexed by the long-run asset position $x_2^\star$, which inherits the initial condition through the transition path.

By Main Paper Equations (22) and (23), consumption grows at the constant rate $(r^\star - r)/\gamma$, so with $r^\star = r$ it is flat. Combining Main Paper Equation (20) with the asset component of Main Paper Equation (26) gives the intertemporal budget constraint

$$y_1(0) = \left(r^\star - \frac{r^\star - r}{\gamma}\right)\left[b_0 + \int_0^\infty e^{-r^\star t}\left(f\left(x_1(t)\right) - y_2(t)\right)\mathrm{d}t\right]. \tag{S3}$$

Capital and q evolve by Main Paper Equations (19), (21) and (24), which do not involve b_0, so a change Δb_0 in the initial position leaves capital, q, investment, and the current account unchanged, shifts consumption by $r^\star \Delta b_0$, and shifts the asset path in parallel, $\Delta x_2(t) = \Delta b_0$.

4. Derivations for the Multicountry Growth Model

This section derives the equations of Main Paper Section 4.5 from the household problem, states the reduced system that the replication code solves, and records the structure of the collocation normal equations behind the preconditioned conjugate-gradient solver of Main Paper Figure 6. There are N countries with constant populations $L_i > 0$, and all variables are per worker. A single freely traded good is the numeraire at every date, and a single bond in zero net supply, denominated in the good, pays the world rate $r^\star(t)$. Countries share the installation technology, depreciation, and discount rate r of Main Paper Section 4.4, produce $z_i f\left(x_{1,i}\right)$ with $z_i > 0$, and have constant relative risk

aversion (CRRA) coefficient $\gamma_i > 0$ (log utility at $\gamma_i = 1$). The baseline sets $z_i = 1$, every experiment sets $\gamma_i = \gamma$, and the scaling experiment of Main Paper Figure 6 draws L_i and z_i.

4.1 *Household problem*

Country i takes $r^\star(t)$ as given and solves

$$\max_{y_{1,i},\, y_{2,i}} \int_0^\infty e^{-rt}\, \frac{y_{1,i}(t)^{1-\gamma_i} - 1}{1-\gamma_i}\, \mathrm{d}t \tag{S4}$$

subject to Main Paper Equations (27) and (28), the initial conditions $x_{1,i}(0) = k_{i,0}$ and $x_{2,i}(0) = b_{i,0}$, and the no-Ponzi condition $\liminf_{t\to\infty} e^{-\int_0^t r^\star(s)\,\mathrm{d}s}\, x_{2,i}(t) \geq 0$. The current-value Hamiltonian is

$$\mathcal{H}_i = \frac{y_{1,i}^{1-\gamma_i} - 1}{1-\gamma_i} + \mu_{2,i}\left[r^\star x_{2,i} + z_i f\left(x_{1,i}\right) - y_{1,i} - y_{2,i}\right] + \mu_{1,i}\left[\chi\, y_{2,i}^{\eta}\, x_{1,i}^{1-\eta} - \delta x_{1,i}\right]. \tag{S5}$$

The maximum principle gives $\partial\mathcal{H}_i/\partial y_{1,i} = \partial\mathcal{H}_i/\partial y_{2,i} = 0$, which are Main Paper Equations (31) and (32), and $\dot{\mu}_{j,i} = r\mu_{j,i} - \partial\mathcal{H}_i/\partial x_{j,i}$, which are Main Paper Equations (29) and (30). The transversality conditions $\lim_{t\to\infty} e^{-rt}\mu_{j,i}(t)\, x_{j,i}(t) = 0$ for capital ($j = 1$) and bonds ($j = 2$), are necessary under the conditions of Michel (1982) and stack into Main Paper Equation (35). The bond condition is the no-Ponzi condition binding with equality.

4.2 *Market clearing and the reduced system*

An equilibrium is a collection of paths $\left(x_{1,i}, x_{2,i}, \mu_{1,i}, \mu_{2,i}, y_{1,i}, y_{2,i}\right)_{i=1}^{N}$ and $r^\star$, $6N+1$ in all, such that every household's conditions hold given $r^\star$ and the bond market clears, Main Paper Equation (33). Define the world bond position and the world excess supply of goods,

$$B := \sum_{i=1}^{N} L_i\, x_{2,i}, \qquad S := \sum_{i=1}^{N} L_i\left[z_i f\left(x_{1,i}\right) - y_{1,i} - y_{2,i}\right]. \tag{S6}$$

Summing Main Paper Equation (28) with weights L_i gives

$$\dot{B} = r^\star B + S, \tag{S7}$$

and with $B(0) = 0$ this linear equation makes $B \equiv 0$ equivalent to $S \equiv 0$. Only one clearing condition is an independent restriction. The implementation imposes bond-market clearing and reports the per-capita goods-market gap $S/\sum_i L_i$ as an ex-post diagnostic.

Stacked in the notation of Main Paper Equations (1) to (3), the algebraic block $\boldsymbol{H}$ has $2N+1$ rows in the $2N+1$ jump variables $\left(y_{1,i}, y_{2,i}, r^\star\right)$, but the clearing row depends only on the states $x_{2,i}$, so $\nabla_{\boldsymbol{y}}\boldsymbol{H}$ is singular and no algebraic equation determines $r^\star$ at an instant. Differentiating the clearing condition twice along solutions gives first

$S = 0$ and then $\dot{S} = 0$, in which $r^\star$ appears linearly with the strictly negative coefficient $-\sum_i L_i \left[y_{1,i}/\gamma_i + y_{2,i}/(1-\eta) \right]$, from the growth rates $\dot{y}_{1,i}/y_{1,i} = (r^\star - r)/\gamma_i$ and $\dot{y}_{2,i}/y_{2,i} = \dot{x}_{1,i}/x_{1,i} + \left(r^\star - G_{1,i}\right)/(1-\eta)$, where $G_{1,i}$ is the bracket of Main Paper Equation (29). Replacing the clearing row by $\dot{S} = 0$ solved for $r^\star$ gives an algebraic block that is triangular in the ordering $\left(y_{1,i}, y_{2,i}, r^\star\right)$ with nonzero determinant whenever capital, the co-states, and the jump quantities are positive, so that formulation is index one (Section 1) and is equivalent to the original along solutions given $S(0) = 0$ and $B(0) = 0$. The kernel solver keeps the clearing row.

The reduced system. As in Main Paper Section 4.4, with Tobin's $q_i := \mu_{1,i}/\mu_{2,i}$, the investment first-order condition Main Paper Equation (32) eliminates $y_{2,i}$ through the investment rate $\iota_i := y_{2,i}/x_{1,i} = (\eta \chi q_i)^{\frac{1}{1-\eta}}$, and the co-states enter only through q_i and $\mu_{2,i}$. The paths solved for are $\left(x_{1,i}, x_{2,i}, q_i, \mu_{2,i}, y_{1,i}\right)_{i=1}^{N}$ and $r^\star$, all in levels, and the residual rows at every collocation point are

$$\begin{aligned} \dot{x}_{1,i} &= x_{1,i}\left(\chi\, \iota_i^{\eta} - \delta\right), & \dot{x}_{2,i} &= r^\star x_{2,i} + z_i f\left(x_{1,i}\right) - y_{1,i} - \iota_i x_{1,i}, \\ \dot{q}_i &= \left(r^\star + \delta - (1-\eta)\,\chi\, \iota_i^{\eta}\right) q_i - z_i f'\left(x_{1,i}\right), & \dot{\mu}_{2,i} &= \left(r - r^\star\right)\mu_{2,i}, \\ 0 &= \mu_{2,i}\, y_{1,i}^{\gamma_i} - 1, & 0 &= \textstyle\sum_j L_j\, x_{2,j} \big/ \sum_j L_j, \end{aligned} \tag{S8}$$

the equation for q_i being $\dot{q}_i/q_i = \dot{\mu}_{1,i}/\mu_{1,i} - \dot{\mu}_{2,i}/\mu_{2,i} = r^\star - G_{1,i}$ by Main Paper Equations (29) and (30). Each path is represented as in Main Paper Section 3: $x_{1,i}(0) = k_{i,0}$ and $x_{2,i}(0) = b_{i,0}$ are pinned by Main Paper Equation (34), the initial values of q_i, $\mu_{2,i}$, $y_{1,i}$, and $r^\star$ are free and unpenalized, $3N+1$ in all, and the $5N+1$ coefficient vectors, those of $r^\star$ and $y_{1,i}$ included, carry the ridge penalty of Main Paper Section A. With n collocation points there are $(5N+1)n$ residual rows, scaled by $((5N+1)n)^{-1/2}$, in $(5N+1)n+3N+1$ unknowns. The clearing row at $t = 0 \in \mathcal{D}$ is identically zero. At a node the budget rows and the clearing row give $S(t_j) = \dot{B}(t_j)$, which $B(t_j) = 0$ does not force to vanish. Neither S nor $\dot{S}$ is imposed at the collocation points, and the ridge penalty selects among the remaining freedom.

4.3 *Steady states and aggregation*

At a steady state, Main Paper Equation (30) with $\mu_{2,i} > 0$ requires $r^\star = r$, so Equation (S8) with $\chi = \delta^{1-\eta}$ gives $q_i^\star = 1/\eta$, $\iota_i^\star = \delta$, and $z_i f'\left(x_{1,i}^\star\right) = \delta + r/\eta$, which is Equation (S1) when $z_i = 1$. Setting $\dot{x}_{2,i} = 0$ gives the budgets $y_{1,i}^\star = z_i f\left(x_{1,i}^\star\right) - \delta x_{1,i}^\star + r\, x_{2,i}^\star$, which leave the long-run asset distribution restricted only by $\sum_i L_i\, x_{2,i}^\star = 0$ and $y_{1,i}^\star > 0$, so the steady states form a continuum of dimension $N-1$.

Integrating $\dot{\mu}_{2,i} = \left(r - r^\star\right)\mu_{2,i}$ and using the consumption first-order condition,

$$\mu_{2,i}(t) = \mu_{2,i}(0)\, D(t), \qquad y_{1,i}(t) = \mu_{2,i}(0)^{-1/\gamma_i}\, D(t)^{-1/\gamma_i}, \qquad D(t) := \exp\left(\int_0^t \left(r - r^\star(s)\right) \mathrm{d}s\right), \tag{S9}$$

so consumption grows at $(r^\star - r)/\gamma_i$. Capital and q_i evolve by the $\dot{x}_{1,i}$ and $\dot{q}_i$ equations of Equation (S8), which involve neither $x_{2,i}$ nor γ_i, so the bond distribution can move them only through $r^\star$. Goods-market clearing with Equation (S9) substituted reads

$$\sum_{i=1}^{N} L_i\, \mu_{2,i}(0)^{-1/\gamma_i}\, D(t)^{-1/\gamma_i} = \sum_{i=1}^{N} L_i \left[z_i f\left(x_{1,i}(t)\right) - \iota_i(t)\, x_{1,i}(t) \right]. \tag{S10}$$

Under a common γ the left side is $\Lambda\, D(t)^{-1/\gamma}$ with $\Lambda := \sum_i L_i\, \mu_{2,i}(0)^{-1/\gamma}$, and at $t = 0$, where $D(0) = 1$, Equation (S10) pins Λ from the $k_{i,0}$ and the $q_i(0)$ alone. The block $\left(x_{1,i}, q_i, r^\star, \Lambda\right)$ therefore solves a system in which the $b_{i,0}$ do not appear. Given uniqueness of its bounded solution, the paths of $r^\star$, every $x_{1,i}$, q_i, and ι_i, and aggregate consumption are invariant to redistributions of the $b_{i,0}$ with $\sum_i L_i\, b_{i,0} = 0$ that keep every country's consumption positive, which only split Λ across countries and move consumption levels and asset paths.

4.4 *A structure-exploiting preconditioner for large N*

Every kernel solution in the paper minimizes the ridge objective of Main Paper Section A with the ridge Levenberg–Marquardt solver of `nlls-gram` (Perla, 2026a), which works in coordinates whitened by the Gram metric, can anneal the ridge toward a floor, and solves the linear system of each step by a dense factorization or by matrix-free conjugate gradient (CG) with a user-supplied preconditioner. The multicountry model supplies the residual rows Equation (S8), the Gram metric, and the preconditioner below. The dense factorization costs $O(N^3 n^3)$ per step, while the preconditioner costs $O(Nn^3 + nN^3)$ to build and $O(Nn^2)$ to apply.

The whitened normal equations. Collect the unknowns as $\theta = \left(\boldsymbol{\alpha}^{(1)}, \ldots, \boldsymbol{\alpha}^{(5N+1)}, \theta_0\right)$: one coefficient block $\boldsymbol{\alpha}^{(b)} \in \mathbb{R}^n$ per path and the $3N + 1$ free initial values θ_0. Let $\boldsymbol{r}(\theta)$ stack the collocation residuals and let $\boldsymbol{K}_{\mathcal{D}} = \boldsymbol{R}^\top \boldsymbol{R}$ be the Cholesky factorization of the Gram matrix on $\mathcal{D}$, so that the ridge penalty is Euclidean in the whitened coefficients $\boldsymbol{R}\boldsymbol{\alpha}^{(b)}$. With $\tilde{\theta}$ the unknowns in whitened coordinates and $\tilde{\boldsymbol{J}}$ the Jacobian of $\boldsymbol{r}$ with respect to them, each step solves

$$\left(\tilde{\boldsymbol{J}}^\top \tilde{\boldsymbol{J}} + \lambda\, \boldsymbol{E} + \varepsilon\, \boldsymbol{I}\right) \boldsymbol{\delta} = -\left(\tilde{\boldsymbol{J}}^\top \boldsymbol{r} + \lambda\, \boldsymbol{E}\, \tilde{\theta}\right), \tag{S11}$$

with λ the ridge weight, ε the damping, and $\boldsymbol{E}$ the projector onto the coefficient coordinates.

Row structure. Every residual row at node t_j depends on the paths only through the levels $\left(\tilde{\boldsymbol{K}}\boldsymbol{\alpha}^{(b)}\right)_j$, the derivatives $\left(\boldsymbol{K}_{\mathcal{D}}\boldsymbol{\alpha}^{(b)}\right)_j$, and θ_0, where $\tilde{K}_{jl} := \int_0^{t_j} k(\tau, t_l)\, \mathrm{d}\tau$ is the integrated kernel matrix. Differentiating the rows of a node with respect to these local values gives scalar weights $c_j^{\rho b}$ and $d_j^{\rho b}$ such that row ρ of node t_j of $\tilde{\boldsymbol{J}}$, restricted to block b, is

$$c_j^{\rho b}\, \check{\boldsymbol{K}}_{j\cdot} + d_j^{\rho b} \left(\boldsymbol{R}^\top\right)_{j\cdot}, \qquad \check{\boldsymbol{K}} := \tilde{\boldsymbol{K}}\, \boldsymbol{R}^{-1}, \qquad \boldsymbol{K}_{\mathcal{D}} \boldsymbol{R}^{-1} = \boldsymbol{R}^\top, \tag{S12}$$

so the coefficient blocks of $\tilde{\boldsymbol{J}}^\top \tilde{\boldsymbol{J}}$ are sums over nodes of outer products of rows of the fixed matrices $\check{\boldsymbol{K}}$ and $\boldsymbol{R}^\top$, and the Jacobian is never materialized.

Arrowhead plus a clearing correction. Country i's five rows in Equation (S8) involve only its own five paths, its three free initial values, and the rate, so their contribution to the operator of Equation (S11) is an arrowhead,

$$\bar{\boldsymbol{A}} = \begin{pmatrix} \boldsymbol{A}_1 & & & \boldsymbol{C}_1^\top \\ & \ddots & & \vdots \\ & & \boldsymbol{A}_N & \boldsymbol{C}_N^\top \\ \boldsymbol{C}_1 & \cdots & \boldsymbol{C}_N & \boldsymbol{A}_0 \end{pmatrix}, \tag{S13}$$

with country blocks $\boldsymbol{A}_i$ of size $5n+3$, a border $\boldsymbol{A}_0$ of size $n+1$ for the rate, and couplings $\boldsymbol{C}_i$. The clearing rows are linear in the bond coefficients, so their whitened Jacobian is $\boldsymbol{U}^\top$ for a fixed $\boldsymbol{U}$ with n columns, nonzero only in the bond blocks, and

$$\tilde{\boldsymbol{J}}^\top \tilde{\boldsymbol{J}} + \lambda \boldsymbol{E} + \varepsilon \boldsymbol{I} = \bar{\boldsymbol{A}} + \boldsymbol{U}\boldsymbol{U}^\top. \tag{S14}$$

The preconditioner. The arrowhead inverts by block elimination, one batched Cholesky of the N country blocks and one of the border Schur complement

$$\boldsymbol{\Sigma} := \boldsymbol{A}_0 - \sum_{i=1}^{N} \boldsymbol{C}_i \, \boldsymbol{A}_i^{-1} \boldsymbol{C}_i^\top, \tag{S15}$$

and the clearing term is restored by the Woodbury identity,

$$\left(\bar{\boldsymbol{A}} + \boldsymbol{U}\boldsymbol{U}^\top\right)^{-1} = \bar{\boldsymbol{A}}^{-1} - \bar{\boldsymbol{A}}^{-1}\boldsymbol{U}\left(\boldsymbol{I}_n + \boldsymbol{U}^\top \bar{\boldsymbol{A}}^{-1}\boldsymbol{U}\right)^{-1} \boldsymbol{U}^\top \bar{\boldsymbol{A}}^{-1}, \tag{S16}$$

with one further Cholesky factorization of the $n \times n$ matrix $\boldsymbol{I}_n + \boldsymbol{U}^\top \bar{\boldsymbol{A}}^{-1}\boldsymbol{U}$. The weights $c_j^{\rho b}, d_j^{\rho b}$ move with the iterate, so the preconditioner is rebuilt after every step.

Settings of the scaling experiment. Every solution in Main Paper Figure 6 uses the kernel and 192-point grid of the baseline. The ridge starts at 10^{-7} and is annealed by factors of 10^{-3} toward a floor of 3×10^{-14} (10^{-14} at $N = 100$), damping starts at 10^{-4}, and CG runs at most 40 iterations per Levenberg–Marquardt step. The optimization converges when the whitened gradient norm and the scaled residual norm are both below 5×10^{-8} within 100 steps, and a solution is accepted when it converges with finite paths, positive where the model requires, and a largest training residual below 10^{-5}. Across sizes, the largest mean validation residual is 3.68×10^{-5} and the largest validation residual 1.73×10^{-3}.

5. Recursive and Stochastic Models

This appendix derives the state-space system Main Paper Equations (36) to (40) for each example of Main Paper Section 5 from its Hamilton–Jacobi–Bellman (HJB) equation, then records the three reductions applied before solving. Co-states are eliminated

through the algebraic conditions, the remaining rows are rescaled to a form in log consumption, and the jump variables are reparameterized through investment. Section 5.3 describes the solvers, Section 5.4 the results, and Section 5.5 the calibration. The theorems of Main Paper Section 3 do not apply here and nothing terminal is imposed in any solution. The diagnostics of Section 5.3 check transversality after the fact at a long finite horizon. The unknown functions are taken to be twice continuously differentiable with interior controls.

We carry a co-state only for each endogenous state, since exogenous states carry no decision, and write $\mathcal{L}$ for the generator Main Paper Equation (41) and $\Gamma(\varphi,\psi) := \nabla\varphi^\top \boldsymbol{B}\boldsymbol{B}^\top \nabla\psi$ for its covariance term. When $\boldsymbol{B}=\mathbf{0}$, $\mathcal{L}\varphi$ is the derivative along the drift and $\Gamma \equiv 0$, so every expression below reduces to its deterministic counterpart.

5.1 *The Stochastic Growth Model*

The state-space system. A planner maximizes $\mathbb{E}\big[\int_0^\infty e^{-rt}u\big(c(t)\big)\mathrm{d}t\big]$ with $u(c)=\frac{c^{1-\gamma}-1}{1-\gamma}$ subject to Main Paper Equations (52) and (53), where $f(k)=k^a$ and productivity is Hicks-neutral with z a zero-mean log productivity process. The HJB equation is

$$rV(z,k)=\max_c\Big\{u(c)+V_k(z,k)\big[e^z f(k)-\delta k-c\big]-\kappa z\,V_z(z,k)+\tfrac{\sigma^2}{2}V_{zz}(z,k)\Big\}, \tag{S17}$$

with first-order condition $u'(c)=V_k$. Differentiating Equation (S17) in k and applying the envelope theorem, so that the terms in $\partial c/\partial k$ drop out,

$$rV_k=V_k\big[e^z f'(k)-\delta\big]+V_{kk}\big[e^z f(k)-\delta k-c\big]-\kappa z\,V_{kz}+\tfrac{\sigma^2}{2}V_{kzz}. \tag{S18}$$

Define the co-state $\mu(z,k):=V_k(z,k)$. The last three terms of Equation (S18) are $\mathcal{L}\mu$ for the generator Main Paper Equation (54), since $\mu_k=V_{kk}$, $\mu_z=V_{kz}$, and $\mu_{zz}=V_{kzz}$, so Equation (S18) is

$$r\mu=\mu\big[e^z f'(k)-\delta\big]+\mathcal{L}\mu, \tag{S19}$$

which is Main Paper Equation (42) with $\boldsymbol{x}=(z,k)$, jump variable $\boldsymbol{y}=c$, co-state $\boldsymbol{\mu}=\mu$, and

$$\boldsymbol{F}=\begin{bmatrix}-\kappa z\\ e^z f(k)-\delta k-c\end{bmatrix},\qquad \boldsymbol{B}=\begin{bmatrix}\sigma & 0\\ 0 & 0\end{bmatrix},\qquad G=e^z f'(k)-\delta,\qquad H=\mu\,c^\gamma-1, \tag{S20}$$

where the algebraic condition $H=0$ is the first-order condition $c^{-\gamma}=\mu$. The value function has disappeared: Equation (S19) and $H=0$ restrict (μ,c) alone. Stacking gives the residual Main Paper Equation (55),

$$\boldsymbol{R}(z,k;\mu,c)=\begin{bmatrix}r\mu-\mu\big[e^z f'(k)-\delta\big]-\mathcal{L}\mu\\ \mu\,c^\gamma-1\end{bmatrix}, \tag{S21}$$

and the transversality condition Main Paper Equation (40) for the endogenous state is

$$0=\lim_{t\to\infty}\mathbb{E}\big[e^{-rt}\,k(t)\,\mu\big(z(t),k(t)\big)\,\big|\,z_0,k_0\big]. \tag{S22}$$

Reductions. The implementations solve Equation (S21) after three reductions.

Eliminate the co-state. The second row of Equation (S21) is imposed exactly by substituting $\mu = c^{-\gamma}$, which leaves the first row as a single equation in the policy $c(z,k)$.

Log form. By Itô's lemma applied to $\mu = e^{-\gamma \log c}$,

$$\frac{\mathcal{L}\mu}{\mu} = -\gamma\,\mathcal{L}\log c + \frac{\gamma^2}{2}\Gamma(\log c, \log c), \tag{S23}$$

and dividing the first row of Equation (S21) by $\gamma\mu$ gives the residual that is set to zero,

$$\mathcal{R}^c(z,k) := \mathcal{L}\log c - \frac{\gamma}{2}\Gamma(\log c, \log c) - \frac{e^z f'(k) - \delta - r}{\gamma} \;=\; \frac{R_1(z,k;\, c^{-\gamma}, c)}{\gamma\, c^{-\gamma}}. \tag{S24}$$

With $\boldsymbol{B} = \mathrm{diag}(\sigma, 0)$, $\Gamma(\log c, \log c) = \sigma^2\big(\partial_z \log c\big)^2$ and $\mathcal{L}\log c = \nabla \log c \cdot \boldsymbol{F} + \frac{\sigma^2}{2}\partial_{zz}\log c$, and Equation (S24) reads

$$\mathcal{R}^c(z,k) = \underbrace{\nabla \log c \cdot \boldsymbol{F}}_{\text{drift}} + \underbrace{\frac{\sigma^2}{2}\Big[\partial_{zz}\log c - \gamma\big(\partial_z \log c\big)^2\Big]}_{\text{Itô and precautionary}} - \frac{e^z f'(k) - \delta - r}{\gamma}. \tag{S25}$$

Equation (S25) and the first row of Equation (S21) have the same zero set, since they differ by the positive factor $\gamma c^{-\gamma}$, but they are not the same objective when a residual is only approximately zero.

Parameterize the jump variable through investment. Rather than c itself, the unknown function is investment $\iota := e^z f(k) - c$, parameterized through a rescaled softplus U_ς as

$$\iota(\boldsymbol{x}) = \delta k_0\, U_\varsigma\big(h(\boldsymbol{x})\big), \qquad U_\varsigma(h) := \frac{1}{\log 2}\log\Big(1 + \exp\big(2\log(2)\,\varsigma h\big)\Big) = \log_2\big(1 + 4^{\varsigma h}\big), \tag{S26}$$

where h is the kernel approximation or the NN of Section 5.3, k_0 is the initial capital stock, and $\varsigma = 1$ here. Rescaling the standard softplus $\log(1+\exp(u))$ in this way gives $U_\varsigma(0) = 1$ and $U_\varsigma'(0) = \varsigma$, so $h \equiv 0$ is the constant policy $\iota = \delta k_0$, replacement investment at the initial capital stock, and $\iota > 0$ for every h. Consumption $c = e^z f(k) - \iota$ is then defined by the resource constraint, and its positivity is not imposed. Finally h takes the state in the coordinates $(z, \log k)$, and the simulations integrate $\log k$, whose drift is $\iota/k - \delta$ with no diffusion, in place of Main Paper Equation (53).

The deterministic case. Setting $\kappa = \sigma = 0$ and $z_0 = 0$ removes the exogenous state. The primitives are then Equation (S20) with $\boldsymbol{F} = \big(0,\ f(k) - \delta k - c\big)$ and $\boldsymbol{B} = \boldsymbol{0}$, so $\mathcal{L}$ is the derivative along the drift, $\Gamma \equiv 0$, and Equation (S25) is the deterministic Euler equation $\nabla \log c \cdot \boldsymbol{F} = \big(f'(k) - \delta - r\big)/\gamma$ evaluated along the flow. The implementation keeps the same policy class over (z,k) and the same change of variables Equation (S26), so the deterministic and stochastic residuals differ only in the drift and volatility of z. The model is that of Main Paper Section 4.1: along the trajectory from k_0, $x(t) = k(t)$, $y(t) = c\big(k(t)\big)$, and $\mu(t) = c\big(k(t)\big)^{-\gamma}$ turn Equation (S19) into Main Paper Equation (7), the second row of Equation (S21) into the CRRA form of Main Paper Equation (8) used in Main Paper Section 4.1, and Equation (S22) into Main Paper Equation (10).

5.2 *The Stochastic Small Open Economy*

The state-space system. We return to the model of Main Paper Section 4.4 and make two of its primitives stochastic: productivity, and the world interest rate that was fixed at $r^\star$ in Main Paper Equations (20) and (22). The household maximizes $\mathbb{E}\big[\int_0^\infty e^{-rt}u\big(c(t)\big)\,\mathrm{d}t\big]$ subject to

$$\mathrm{d}k(t) = \big[\chi\,\iota(t)^{\eta}k(t)^{1-\eta} - \delta k(t)\big]\mathrm{d}t, \tag{S27}$$

$$\mathrm{d}b(t) = \big[r^\star(t)\,b(t) + e^{z(t)}f\big(k(t)\big) - c(t) - \iota(t)\big]\mathrm{d}t, \tag{S28}$$

$$\mathrm{d}z(t) = -\kappa_z\,z(t)\,\mathrm{d}t + \sigma_z\,\mathrm{d}W_1(t), \tag{S29}$$

$$\mathrm{d}r^\star(t) = -\kappa_r\big[r^\star(t) - \bar r^\star\big]\mathrm{d}t + \sigma_r\Big[\varrho\,\mathrm{d}W_1(t) + \sqrt{1-\varrho^2}\,\mathrm{d}W_4(t)\Big], \tag{S30}$$

where k is installed capital, b is the net foreign asset position, ι is investment expenditure, $f(k)=k^a$, and ϱ is the correlation between the productivity and world-rate shocks. Equation (S30) is the continuous-time counterpart of the world interest rate process estimated by Uribe and Yue (2006). In the notation of Main Paper Section 5.1 the state is $\boldsymbol{x}=(z,k,b,r^\star)$, the jump variables are $\boldsymbol{y}=(c,\iota)$, $\boldsymbol{W}$ is a standard four-dimensional Brownian motion of which only two components are used, and

$$\boldsymbol{F} = \begin{bmatrix} -\kappa_z z \\ \chi\,\iota^{\eta}k^{1-\eta} - \delta k \\ r^\star b + e^z f(k) - c - \iota \\ -\kappa_r\big(r^\star - \bar r^\star\big) \end{bmatrix}, \qquad \boldsymbol{B} = \begin{bmatrix} \sigma_z & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ \varrho\,\sigma_r & 0 & 0 & \sqrt{1-\varrho^2}\,\sigma_r \end{bmatrix}, \tag{S31}$$

so that $\boldsymbol{B}\boldsymbol{B}^{\top}$ has diagonal entries σ_z^2 and σ_r^2 in the z and $r^\star$ positions and off-diagonal entry $\varrho\,\sigma_z\sigma_r$.

The HJB equation is

$$\begin{aligned} rV(\boldsymbol{x}) = \max_{c,\,\iota}\Big\{ & u(c) + V_k\big[\chi\,\iota^{\eta}k^{1-\eta} - \delta k\big] + V_b\big[r^\star b + e^z f(k) - c - \iota\big] \\ & - \kappa_z z\,V_z - \kappa_r\big(r^\star - \bar r^\star\big)V_{r^\star} + \tfrac{1}{2}\big[\sigma_z^2 V_{zz} + 2\varrho\,\sigma_z\sigma_r V_{zr^\star} + \sigma_r^2 V_{r^\star r^\star}\big]\Big\}, \end{aligned} \tag{S32}$$

with first-order conditions

$$u'(c) = V_b, \qquad \eta\,\chi\,V_k\left(\frac{\iota}{k}\right)^{\eta-1} = V_b. \tag{S33}$$

Define the co-states $\mu_1 := V_k$ for installed capital and $\mu_2 := V_b$ for wealth. Differentiating Equation (S32) in b, the only drift term that depends on b is $V_b r^\star b$, and the envelope theorem gives $r\mu_2 = r^\star\mu_2 + \mathcal{L}\mu_2$. Differentiating in k, using $\partial_k\big[\chi\iota^{\eta}k^{1-\eta} - \delta k\big] = (1-\eta)\chi(\iota/k)^{\eta} - \delta$, gives $r\mu_1 = \mu_1\big[(1-\eta)\chi(\iota/k)^{\eta} - \delta\big] + \mu_2\,e^z f'(k) + \mathcal{L}\mu_1$. Both are Main Paper Equation (42) with

$$\boldsymbol{G} = \begin{bmatrix} (1-\eta)\,\chi\left(\dfrac{\iota}{k}\right)^{\eta} - \delta + \dfrac{\mu_2}{\mu_1}\,e^z f'(k) \\ r^\star \end{bmatrix}, \qquad \boldsymbol{H} = \begin{bmatrix} \mu_2\,c^{\gamma} - 1 \\ \eta\,\chi\,\mu_1\left(\dfrac{\iota}{k}\right)^{\eta-1} - \mu_2 \end{bmatrix}, \tag{S34}$$

which are Main Paper Equations (21) to (24) with $f'(k)$ replaced by $e^z f'(k)$ and the world rate a state rather than a parameter. The residual Main Paper Equation (44) has four rows,

$$\boldsymbol{R}(\boldsymbol{x};\boldsymbol{\mu},\boldsymbol{y}) = \begin{bmatrix} r\mu_1 - \mu_1 G_1(\boldsymbol{x},\boldsymbol{\mu},\boldsymbol{y}) - \mathcal{L}\mu_1 \\ r\mu_2 - \mu_2\, r^\star - \mathcal{L}\mu_2 \\ \mu_2\, c^\gamma - 1 \\ \eta\,\chi\,\mu_1\,(\iota/k)^{\eta-1} - \mu_2 \end{bmatrix}, \tag{S35}$$

and the transversality conditions Main Paper Equation (40) for the two endogenous states are

$$0 = \lim_{t\to\infty} \mathbb{E}\big[e^{-rt}\, k(t)\, \mu_1\big(\boldsymbol{x}(t)\big)\big], \qquad 0 = \lim_{t\to\infty} \mathbb{E}\big[e^{-rt}\, b(t)\, \mu_2\big(\boldsymbol{x}(t)\big)\big]. \tag{S36}$$

Reductions. Eliminate the co-states. The third row of Equation (S35) gives $\mu_2 = c^{-\gamma}$, and the fourth gives the ratio

$$q := \frac{\mu_1}{\mu_2} = \frac{1}{\chi\eta}\left(\frac{\iota}{k}\right)^{1-\eta}, \tag{S37}$$

Tobin's q, the inversion of Main Paper Equation (24) used in Main Paper Section 4.4. Both rows are imposed exactly by substituting $\mu_2 = c^{-\gamma}$ and $\mu_1 = q\, c^{-\gamma}$, so the unknown functions are the two jump variables.

Parameterize the jump variables through investment and the current account. In place of (c,ι) the unknown functions are investment and the current account, $CA := r^\star b + e^z f(k) - c - \iota$, the drift of b in Equation (S28). Consumption is then defined by the budget constraint, $c = r^\star b + e^z f(k) - \iota - CA$. Investment follows the change of variables Equation (S26), $\iota(\boldsymbol{x}) = \delta k_0\, U_\varsigma\big(h_1(\boldsymbol{x})\big)$ with $\varsigma = e^{z_0} f(k_0)/(\delta k_0)$, and the current account is a signed linear output, $CA(\boldsymbol{x}) = e^{z_0} f(k_0)\, h_2(\boldsymbol{x})$, so $h \equiv \mathbf{0}$ is replacement investment at the initial capital stock with a zero current account.

Log forms. By Itô's lemma applied to $\mu_2 = e^{-\gamma \log c}$ and $\mu_1 = e^{\log q - \gamma \log c}$,

$$\frac{\mathcal{L}\mu_2}{\mu_2} = -\gamma\, \mathcal{L}\log c + \frac{\gamma^2}{2}\Gamma(\log c, \log c), \tag{S38}$$

$$\frac{\mathcal{L}\mu_1}{\mu_1} = \mathcal{L}\log q + \tfrac{1}{2}\Gamma(\log q, \log q) - \gamma\,\Gamma(\log q, \log c) + \frac{\mathcal{L}\mu_2}{\mu_2}. \tag{S39}$$

Dividing the second row of Equation (S35) by $\gamma\mu_2$, and subtracting the first row divided by μ_1 from the second row divided by μ_2, gives the two residuals the implementations set to zero,

$$\mathcal{R}^c(\boldsymbol{x}) := \mathcal{L}\log c - \frac{\gamma}{2}\Gamma(\log c, \log c) - \frac{r^\star - r}{\gamma} = \frac{R_2}{\gamma\,\mu_2}, \tag{S40}$$

$$\mathcal{R}^q(\boldsymbol{x}) := \mathcal{L}\log q + \tfrac{1}{2}\Gamma(\log q, \log q) - \gamma\,\Gamma(\log q, \log c) - \left[r^\star + \delta - (1-\eta)\,\chi\left(\frac{\iota}{k}\right)^\eta - \frac{e^z f'(k)}{q}\right]$$
$$= -\frac{R_1}{\mu_1} + \frac{R_2}{\mu_2}, \tag{S41}$$

with μ_1 and μ_2 evaluated at $\mu_2 = c^{-\gamma}$ and $\mu_1 = q\,c^{-\gamma}$. The bracket in Equation (S41) is $r^\star - G_1$. The map $(R_1, R_2) \mapsto (\mathcal{R}^c, \mathcal{R}^q)$ is inverted by $R_2 = \gamma\mu_2\mathcal{R}^c$ and $R_1 = \mu_1\big(\gamma\,\mathcal{R}^c - \mathcal{R}^q\big)$, so the two pairs have the same zero set. Here $\Gamma(\varphi,\psi) = \sigma_z^2\,\partial_z\varphi\,\partial_z\psi + \varrho\sigma_z\sigma_r\big(\partial_z\varphi\,\partial_{r^\star}\psi + \partial_{r^\star}\varphi\,\partial_z\psi\big) + \sigma_r^2\,\partial_{r^\star}\varphi\,\partial_{r^\star}\psi$.

Equation (S40) is Equation (S24) with $r^\star$ in place of $e^z f'(k) - \delta$. Equation (S41) is the no-arbitrage condition for installed capital under the stochastic discount factor $e^{-rt}c(t)^{-\gamma}$: $\mathcal{L}\log q + \frac{1}{2}\Gamma(\log q, \log q)$ is the expected capital gain, and $\gamma\,\Gamma(\log q, \log c)$ is the risk premium, the expected excess return on installed capital, absent from Main Paper Equation (21) and zero when $\boldsymbol{B} = \mathbf{0}$.

The deterministic case. Setting $\sigma_z = \sigma_r = 0$ with $z_0 = 0$ and $r^\star(0) = \bar{r}^\star$ makes productivity and the world rate constant, Γ vanishes, and Equations (S27) and (S28) and the rows of Equation (S35) are Main Paper Equations (19) to (24) along the trajectory $\boldsymbol{x}(t) = \big(k(t), b(t)\big)$, $\boldsymbol{y}(t) = \big(c(\boldsymbol{x}(t)), \iota(\boldsymbol{x}(t))\big)$. Neither formulation imposes transversality.

5.3 *Approximation and Solution*

Kernel policies. The kernel implementations solve the fixed-distribution problem Main Paper Equation (48) with $\mathcal{D}$ a fixed set of states and a small fixed ridge, in the form of Main Paper Section A. For the growth models, with the residual Equation (S25),

$$\min_{\boldsymbol{\alpha},\boldsymbol{\beta},h_0} \frac{1}{|\mathcal{D}|}\sum_{\boldsymbol{x}_i\in\mathcal{D}} \mathcal{R}^c(\boldsymbol{x}_i)^2 + \lambda\,\boldsymbol{\alpha}^\top\big(\boldsymbol{K} + \epsilon\boldsymbol{I}\big)\boldsymbol{\alpha}, \qquad h(\boldsymbol{x}) = \sum_{j=1}^{|\mathcal{D}|}\alpha_j K(\boldsymbol{x},\boldsymbol{x}_j) + \boldsymbol{\beta}^\top\begin{bmatrix} z \\ \log k\end{bmatrix} + h_0, \tag{S42}$$

where h is the unknown function of Equation (S26), $\boldsymbol{K}_{ij} = K(\boldsymbol{x}_i,\boldsymbol{x}_j)$ is the Gram matrix of the state kernel K, λ is a small fixed ridge, $\epsilon > 0$ keeps the metric factorizable when states in $\mathcal{D}$ coincide, and the linear and constant block is unpenalized. Levenberg–Marquardt starts from $\boldsymbol{\alpha} = \mathbf{0}$, $\boldsymbol{\beta} = \mathbf{0}$, $h_0 = 0$, the replacement-investment policy of Equation (S26). For the open economy there are two expansions h_1 and h_2 on the same centers with coefficient vectors $\boldsymbol{\alpha}^\iota$ and $\boldsymbol{\alpha}^{CA}$, penalized as $\lambda\big[\boldsymbol{\alpha}^{\iota\top}(\boldsymbol{K} + \epsilon\boldsymbol{I})\boldsymbol{\alpha}^\iota + \boldsymbol{\alpha}^{CA\top}(\boldsymbol{K} + \epsilon\boldsymbol{I})\boldsymbol{\alpha}^{CA}\big]$, each with an unpenalized linear term in the standardized state and no constant, and the objective averages Equations (S40) and (S41) over the $2|\mathcal{D}|$ rows. Since the stochastic residuals differentiate the policy twice, K must be at least twice differentiable, which rules out the Matérn-$1/2$ kernel of the sequential examples.

For the growth models the state kernel is the sum of a squared exponential along a ray coordinate and an anisotropic squared exponential in the two states,

$$K(\boldsymbol{x},\boldsymbol{x}') = A_\zeta^2\exp\left(-\frac{(\zeta-\zeta')^2}{2\ell_\zeta^2}\right) + A^2\exp\left(-\frac{(z-z')^2}{2\ell_z^2} - \frac{(\log k - \log k')^2}{2\ell_k^2}\right), \tag{S43}$$

with the ray coordinate $\zeta := \log k - s_\zeta\,z$. The first component lets the policy vary along a single direction in the state, capital relative to a productivity-scaled level, with the two-dimensional component carrying the remainder. The slope s_ζ is a tuned hyperparameter like the amplitudes and lengthscales, $s_\zeta = 2$ for the stochastic model, and

the balanced-growth relation between z and k is not imposed. In the deterministic model $z \equiv 0$, so $\zeta = \log k$ and the kernel is a sum of two squared exponentials in $\log k$. For the open economy it is an anisotropic squared exponential $K(\tilde{\boldsymbol{x}}, \tilde{\boldsymbol{x}}') = A^2 \exp\big(-\frac{1}{2}\sum_{m=1}^{4}(\tilde{x}_m - \tilde{x}'_m)^2/\ell_m^2\big)$ in the standardized coordinates $\tilde{\boldsymbol{x}} := \big(z/\mathrm{sd}_z,\ \log k - \log k_0,\ (b - b_0)/(20\, e^{z_0} f(k_0)),\ (r^\star - \bar{r}^\star)/\mathrm{sd}_r\big)$, with $\mathrm{sd}_z = \sigma_z/\sqrt{2\kappa_z}$ and $\mathrm{sd}_r = \sigma_r/\sqrt{2\kappa_r}$ the stationary standard deviations of the exogenous states. These are also the reference NN's input coordinates.

$\mathcal{D}$ is fixed before solving and its states are the kernel centers. No policy values from the source of $\mathcal{D}$ enter the objective, and no integration is performed while solving. In the deterministic growth model $\mathcal{D}$ is the benchmark path sampled at $t = 0, 1, \ldots, 60$. In the stochastic models it is one simulated batch of trajectories of the reference NN policy below, integrated with the stochastic Runge–Kutta scheme for additive noise of strong order 1.5 (SRA1) of Rößler (2010) from the initial distribution of Section 5.5. The trajectory and state counts are in Table 3.

NN policies. The NN implementations solve the sample average approximation Main Paper Equation (51). Write $\mathrm{NN}(\cdot;\theta) : \mathbb{R}^{d_x} \to \mathbb{R}^M$ for an NN with d hidden layers of width w, alternating affine maps with a pointwise nonlinearity ϕ,

$$\mathrm{NN}(\boldsymbol{x};\theta) = W_{d+1}\,\phi\big(W_d\,\phi(\cdots\phi(W_1\boldsymbol{x} + b_1)\cdots) + b_d\big) + b_{d+1}, \tag{S44}$$

with $\theta := \{W_j, b_j\}_{j=1}^{d+1}$, $W_1 \in \mathbb{R}^{w\times d_x}$, $W_j \in \mathbb{R}^{w\times w}$ for $1 < j \le d$, $W_{d+1} \in \mathbb{R}^{M\times w}$, and conformable biases. The unknown function h of Equation (S26) is

$$h(\boldsymbol{x}) = \mathrm{NN}\big(\tilde{\boldsymbol{x}};\theta\big), \tag{S45}$$

with $\tilde{\boldsymbol{x}} = (z - z_0, \log k - \log k_0)$ in the growth model ($d_x = 2$) and the standardized state of the kernel above in the open economy ($d_x = 4$, with $M = 2$ outputs (h_1, h_2)). Here $d = 2$, $w = 16$, and ϕ is the SiLU.[1] The change of variables Equation (S26) rescales the output, and the output layer W_{d+1} is initialized near zero, so training starts near the replacement-investment policy. The approximants of Main Paper Equation (51) are the functions of h given by the reductions of Sections 5.1 and 5.2: $\hat{\boldsymbol{y}} = \hat{c}$ and $\hat{\boldsymbol{\mu}} = \hat{c}^{-\gamma}$ in the growth model, $\hat{\boldsymbol{y}} = (\hat{c}, \hat{\imath})$ and $\hat{\boldsymbol{\mu}} = \big(\hat{q}\,\hat{c}^{-\gamma}, \hat{c}^{-\gamma}\big)$ in the open economy. The objective is Main Paper Equation (51) with the reduced residuals in place of $\boldsymbol{R}$,

$$\min_{\theta}\ \frac{1}{n_r}\sum_{s=1}^{S}\sum_{i=1}^{N}\big\|\mathcal{R}\big(\hat{\boldsymbol{x}}_i^{(s)}\big)\big\|_2^2, \tag{S46}$$

with $\mathcal{R} = \mathcal{R}^c$ of Equation (S25) and $n_r = SN$ in the growth model, and $\mathcal{R} = \big(\mathcal{R}^c, \mathcal{R}^q\big)$ of Equations (S40) and (S41) and $n_r = 2SN$ in the open economy, where the states $\hat{\boldsymbol{x}}_i^{(s)}$ are simulated under the current policy. The simulation runs in the coordinates of the implementation, $\hat{\boldsymbol{x}} = (z, \log k)$ or $(z, \log k, b, r^\star - \bar{r}^\star)$, with drift $\hat{\boldsymbol{F}}(\hat{\boldsymbol{x}}) := \boldsymbol{F}\big(\hat{\boldsymbol{x}}, \hat{\boldsymbol{\mu}}(\hat{\boldsymbol{x}}), \hat{\boldsymbol{y}}(\hat{\boldsymbol{x}})\big)$ in those coordinates, whose log-capital component is $\hat{\imath}/k - \delta$ in the growth model and $\chi(\hat{\imath}/k)^\eta - \delta$ in the open economy, and with $\boldsymbol{B}$ unchanged since capital carries no diffusion. In the deterministic models the increments of Main Paper Equation (49) drop

[1] The sigmoid linear unit, $\phi(u) = u/(1 + e^{-u})$.

out and the Euler–Maruyama step Main Paper Equation (50) is replaced by a fixed-step fourth-order Runge–Kutta step in the same form,

$$\begin{aligned} \boldsymbol{k}_1 &= \hat{\boldsymbol{F}}\big(\hat{\boldsymbol{x}}_i^{(s)}\big), \qquad \boldsymbol{k}_2 = \hat{\boldsymbol{F}}\big(\hat{\boldsymbol{x}}_i^{(s)} + \tfrac{\Delta t}{2}\boldsymbol{k}_1\big), \\ \boldsymbol{k}_3 &= \hat{\boldsymbol{F}}\big(\hat{\boldsymbol{x}}_i^{(s)} + \tfrac{\Delta t}{2}\boldsymbol{k}_2\big), \qquad \boldsymbol{k}_4 = \hat{\boldsymbol{F}}\big(\hat{\boldsymbol{x}}_i^{(s)} + \Delta t\,\boldsymbol{k}_3\big), \\ \hat{\boldsymbol{x}}_{i+1}^{(s)} &= \hat{\boldsymbol{x}}_i^{(s)} + \tfrac{\Delta t}{6}\big(\boldsymbol{k}_1 + 2\boldsymbol{k}_2 + 2\boldsymbol{k}_3 + \boldsymbol{k}_4\big), \end{aligned} \tag{S47}$$

for $i = 1, \ldots, N-1$. In the stochastic models the step is Main Paper Equation (50),

$$\hat{\boldsymbol{x}}_{i+1}^{(s)} = \hat{\boldsymbol{x}}_i^{(s)} + \Delta t\,\hat{\boldsymbol{F}}\big(\hat{\boldsymbol{x}}_i^{(s)}\big) + \boldsymbol{B}\,\Delta\boldsymbol{W}_i^{(s)}, \qquad \Delta\boldsymbol{W}_i^{(s)} = \sqrt{\Delta t}\,\boldsymbol{Z}_i^{(s)}. \tag{S48}$$

Both schemes are run by the differentiable JAX solvers of Perla (2026b). Every state after the initial condition depends on θ through the preceding steps, so the objective is a function of θ through the whole discretized solution of the ODE or stochastic differential equation (SDE), and its derivatives are taken through the scheme by automatic differentiation: the discretization comes first and the differentiation second, rather than through the continuous adjoint equations of Chen et al. (2018), Li et al. (2020). Within an epoch the draws are fixed, so Equation (S46) is a deterministic nonlinear least-squares problem in θ whose gradient passes through the fixed draws (Murphy, 2023, section 6.3), minimized by Levenberg–Marquardt with geodesic acceleration and the identity metric on θ (Perla, 2026a). The Jacobian of the residual in θ is assembled through the simulation by forward- or reverse-mode differentiation, whichever side is smaller, and factorized densely. Inside $\mathcal{R}$ the generator and Γ are evaluated by nested forward-mode differentiation along $\boldsymbol{F}$ and the columns of $\boldsymbol{B}$, so the objective differentiates the policy twice in the state and once more in θ.

The draws are held fixed for an epoch of solver steps. At each interior epoch boundary the Brownian increments are renewed, in the stochastic growth model together with the initial capital stocks, the damping is reset, and the solver stops early if the loss on the fresh batch is below the threshold. The solver restarts from a fresh initialization if its loss becomes nonfinite or, in the single-epoch deterministic case, if the threshold is not reached within the step budget. In the stochastic models the NN with the smallest mean test residual is the reference policy whose simulated states form the kernel's $\mathcal{D}$.

TABLE 2. Settings of the recursive NN solver. Simulated $\mathcal{D}$ is trajectories $\times$ states per trajectory with the step Δt; epochs $\times$ Levenberg–Marquardt steps per epoch, both budgets; damping (initial, factor after an accepted step, factor after a rejected step); restart attempts in all; training-loss threshold.

Model	Simulated $\mathcal{D}$	Epochs $\times$ steps	Damping (initial, accept, reject)	Attempts	Loss threshold
Deterministic growth	1 traj $\times$ 61, $\Delta t = 1$	1×125	4.74×10^{-4}, 0.511, 2.82	2	2.15×10^{-14}
Stochastic growth	16 traj $\times$ 41, $\Delta t = 1.5$	4×125	3.26×10^{-5}, 0.482, 2.71	4	5×10^{-13}
Stochastic open economy	4 traj $\times$ 33, $\Delta t = 1.25$	5×100	5.1×10^{-4}, 0.195, 3.22	2	10^{-10}

Diagnostics. Test residuals evaluate the reduced residual of the model along trajectories resimulated under the fitted policy, from fresh initial-condition draws where the

initial distribution is dispersed and with fresh shocks in the stochastic models: 1×41 on $[0, 40]$ for the deterministic growth model, 50×41 on $[0, 40]$ for the stochastic one, and 8×51 on $[0, 50]$ for the open economy, all with $\Delta t = 1$. Path errors compare the simulated paths against a reference: the benchmark of Main Paper Section 4.1 for the deterministic growth model, and for the stochastic kernels the reference NN simulated from the same initial conditions with the same Brownian increments.[2] The stochastic NNs have no independent reference. Transversality diagnostics evaluate Equations (S22) and (S36) at a long but finite horizon. For the growth models the diagnostic is

$$\log\left[e^{-rT}c(T)^{-\gamma}k(T)\right], \qquad T = 200, \tag{S49}$$

along one simulated path per training initial condition, reported as its maximum over those initial conditions. For the open economy the diagnostics are $\log \mathbb{E}\left[e^{-rT}c(T)^{-\gamma}q(T)k(T)\right]$ and $\log \mathbb{E}\left[e^{-rT}c(T)^{-\gamma}|b(T)|\right]$ at $T = 160$, with the expectations estimated by averages over 128 simulated paths. The absolute value is strictly stronger than the signed limit in Equation (S36).

5.4 *Results*

Table 3 collects the results for every model and solver, with the calibration of Section 5.5. NN entries are medians over five seeds. Section 6 places the deterministic growth rows beside the sequential solvers on the same calibration.

Growth models. In the deterministic model both solvers recover the benchmark path with the accuracy of the sequential solvers, and the diagnostic Equation (S49) is the same for both because their simulated paths coincide to the reported accuracy. In the stochastic model the NNs attain a median test residual of 4.86×10^{-8}. The kernel policy, fit on its fixed $\mathcal{D}$, has a mean absolute test residual of 7.74×10^{-5}. Under common shocks it matches the reference paths with mean relative errors of 2.65×10^{-4} in capital and 1.59×10^{-4} in consumption. Figure 3 shows the simulated capital paths under the kernel policy for the 50 test trajectories and, at each date, the median and 90th percentile of the relative errors against the reference NN.

Open economy. The training residuals Equations (S40) and (S41) are 9.83×10^{-8} for the NNs (median over seeds) and 1.05×10^{-5} for the kernel. Test values are in Table 3. Under common shocks the kernel matches the reference paths with a mean relative capital error of 3.95×10^{-4} and a mean asset error of 0.0217. The asset error is measured relative to initial output, since b changes sign. Figure 4 shows the simulated paths under the kernel policy for the eight test trajectories and, at each date, the median and 90th percentile of the residuals across them.

5.5 *Calibration*

Table 4 collects the annual calibrations. The growth block maps the benchmark of Aruoba et al. (2006), their quarterly discount factor, risk aversion, and productivity process converted to annual rates, with their high-volatility column for σ and the annual

[2] The reference NN's own test residual is measured on its Euler–Maruyama test paths, while the kernel's is measured on SRA1 paths, so those two residuals are not on common shocks.

TABLE 3. Recursive and stochastic models: kernel and NN policies. Size is the number of penalized kernel coefficients (one per center for the growth models, two for the open economy) or NN parameters. Simulated $\mathcal{D}$ is the number of trajectories and states per trajectory on which the residual is enforced with the step Δt, on $[0, 60]$ for the growth models and $[0, 40]$ for the open economy: the benchmark path for the deterministic growth kernel, the reference NN's simulated paths for the stochastic kernels, and the NN's own fourth-order Runge–Kutta or Euler–Maruyama trajectories otherwise. Test residuals are means of $|\mathcal{R}|$ on the test trajectories, and path errors are mean absolute relative errors in capital against the benchmark (deterministic) or the reference NN under common shocks (stochastic kernels). For the growth models the diagnostic is Equation (S49). For the open economy the two diagnostics are the logs of the capital and absolute-asset expectations at $T = 160$. Times are seconds for one solution.

Model	Solver	Size	Simulated $\mathcal{D}$	Test $\lvert\mathcal{R}\rvert$	Path error	Transversality	Time
Growth model							
Deterministic	Kernel	61 coeffs	1 traj $\times$ 61, $\Delta t = 1$	5.8×10^{-6}	3.22×10^{-5}	-7.4	0.272
Deterministic	NN	337 params	1 traj $\times$ 61, $\Delta t = 1$	1.39×10^{-7}	1.34×10^{-6}	-7.4	0.138
Stochastic	NN	337 params	16 traj $\times$ 41, $\Delta t = 1.5$	4.86×10^{-8}	—	-7.295	8.75
Stochastic	Kernel	336 coeffs	16 traj $\times$ 21, $\Delta t = 3$	7.74×10^{-5}	2.65×10^{-4}	-7.296	0.56
Small open economy (diagnostics: capital, assets)							
Stochastic	NN	386 params	4 traj $\times$ 33, $\Delta t = 1.25$	2.75×10^{-6}	—	$-4.93, -4.66$	3.89
Stochastic	Kernel	672 coeffs	16 traj $\times$ 21, $\Delta t = 2$	5.04×10^{-5}	3.95×10^{-4}	$-4.81, -4.44$	1.5

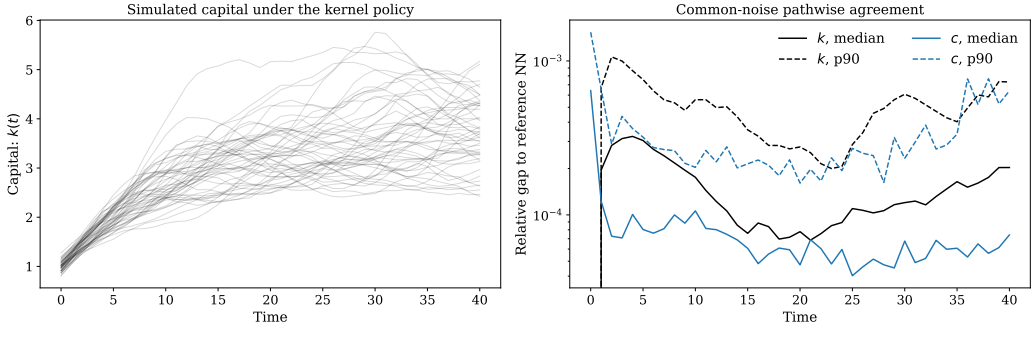


FIGURE 3. The stochastic growth model. Left: the 50 simulated capital paths under the kernel policy, one per test shock realization. Right: the absolute relative error of the kernel policy's capital and consumption paths against the reference NN under common shocks (median and 90th percentile across the 50 test trajectories at each date).

depreciation of King and Rebelo (1999), and coincides with the calibration of Main Paper Section 4.1. The open economy keeps the calibration of Main Paper Section 4.4 and adds the same productivity process together with the world-rate process of Uribe and Yue (2006) converted to an annual Ornstein–Uhlenbeck process, whose stationary dispersion of about 1.3 percentage points matches the international risk-free real rate estimated by Fernández-Villaverde et al. (2011). The initial debtor position $b_0 = -0.8$, roughly half of steady-state output, follows Mendoza and Villalvazo (2020). Initial states

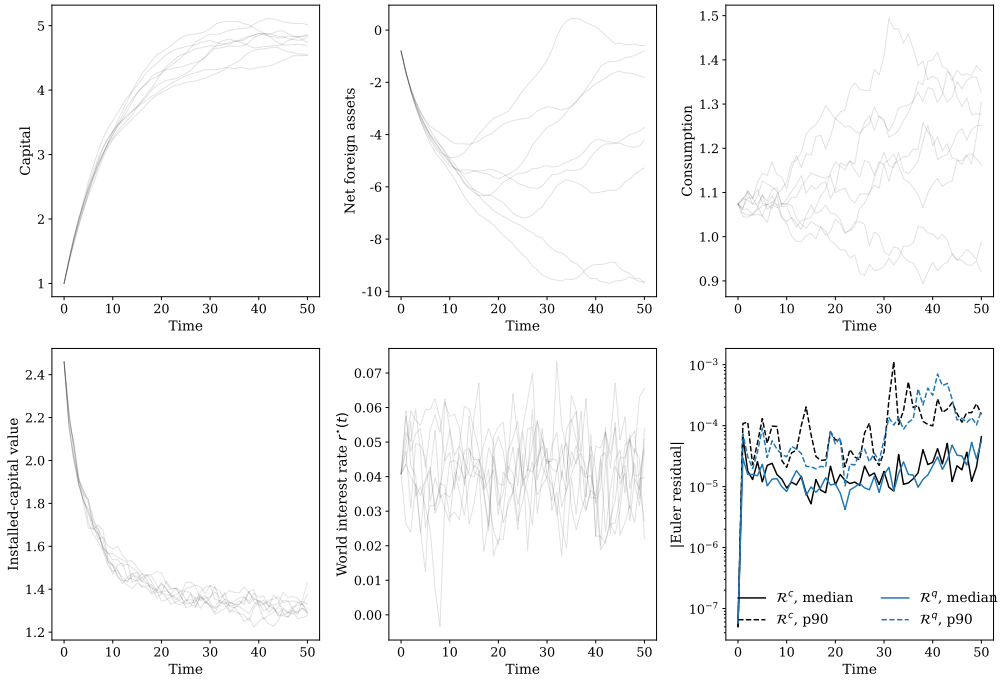


FIGURE 4. The stochastic small open economy. Simulated paths under the kernel policy for capital, net foreign assets, consumption, the installed-capital value, and the world interest rate across the eight test shock realizations, together with the absolute residuals $|\mathcal{R}^c|$ and $|\mathcal{R}^q|$ of Equations (S40) and (S41) (median and 90th percentile across trajectories at each date).

are point masses except that the stochastic growth training and test trajectories draw $\log k_0$ with variance 0.01 around $\log k = 0$.

TABLE 4. Annual calibration of the recursive and stochastic models.

	Growth (z, k)	Open economy $(z, k, b, r^\star)$
Preferences γ, r	2, $-4\log 0.9896 \approx 0.042$	2, $-\log 0.96 \approx 0.041$
Technology a, δ	$\frac{1}{3}$, 0.1	0.33, $-\log 0.94 \approx 0.062$
Installation η, χ	—	0.76, $\delta^{1-\eta}$
Productivity κ, σ	$-4\log 0.95 \approx 0.205$, 0.072	$\kappa_z = -4\log 0.95 \approx 0.205$, $\sigma_z = 0.072$
World rate $\bar{r}^\star$, κ_r, σ_r, ϱ	—	r, $-4\log 0.83 \approx 0.745$, 0.016, 0
Initial states	$z_0 = 0$, $k_0 = 1$	$z_0 = 0$, $k_0 = 1$, $b_0 = -0.8$, $r^\star(0) = \bar{r}^\star$

6. SOLVER COMPARISON

The constraints of Main Paper Equation (13) can also be interpolated by an NN in place of the kernel expansion, with selection supplied by the optimizer's inductive bias rather than an explicit minimum norm (Ebrahimi Kahou et al., 2024). This section first records

the classical shooting approach and its sensitivity to the horizon, then states the NN solver for the sequential problems and compares it with the kernel solver. Section 5.3 states the NN solver for the state-space problems.

6.1 *Shooting and finite-horizon boundary-value methods*

Write $(\boldsymbol{x}, \boldsymbol{\mu}, \boldsymbol{y})(t; \tilde{\boldsymbol{\mu}}_0)$ for the solution of Main Paper Equations (1) to (3) from $(\boldsymbol{x}_0, \tilde{\boldsymbol{\mu}}_0)$, and let $\boldsymbol{x}(\infty)$ be the steady state to which the solution converges. Shooting solves the M equations

$$\Psi(\tilde{\boldsymbol{\mu}}_0; T) := \boldsymbol{x}(T; \tilde{\boldsymbol{\mu}}_0) - \boldsymbol{x}(\infty) = \mathbf{0} \tag{S50}$$

in the M unknowns $\tilde{\boldsymbol{\mu}}_0$ by Newton's method, or bisection when $M = 1$, integrating the system with a standard DAE or ODE initial value solver at each evaluation. Any trajectory whose $(\boldsymbol{x}, \boldsymbol{\mu})$ converges to a finite steady state satisfies Main Paper Equation (5), which is why the terminal condition stands in for the asymptotic one.

Sensitivity to the horizon. Suppose the steady state is a hyperbolic saddle with exactly M unstable directions, that perturbations of the initial co-state map invertibly onto those directions, and that the directions map invertibly onto the terminal state. Evaluate the Jacobian $\nabla\Psi(\cdot; T)$ along the solution. Each unstable mode then grows at the rate of its own eigenvalue, so with $\lambda_{\max}$ and $\lambda_{\min}$ the largest and smallest real parts among the unstable eigenvalues of the linearization at the steady state, $\|\nabla\Psi(\cdot; T)\|$ grows like $e^{\lambda_{\max} T}$, up to polynomial factors when the unstable block is not diagonalizable. For $M = 1$ the Jacobian is a scalar, so the issue is magnitude rather than conditioning: locally around the finite-horizon root a fixed sufficiently small terminal tolerance corresponds to an interval of guesses of width shrinking like $e^{-\lambda_{\max} T}$, and integration errors with unstable components can be amplified at that rate. For $M \geq 2$ with distinct unstable real parts the condition number of $\nabla\Psi(\cdot; T)$ grows like $e^{(\lambda_{\max} - \lambda_{\min})T}$ as well. Against this, a small T forces the path toward the steady state before the transition is complete and biases the approximation, so the horizon trades bias against sensitivity, and the tuning becomes harder with the dimension.

Finite-horizon boundary-value methods. Related approaches discretize time and solve the nonlinear system that enforces the dynamics, the initial condition Main Paper Equation (4), and the artificial terminal condition $\boldsymbol{x}(T) = \boldsymbol{x}(\infty)$ jointly. These avoid the unstable forward propagation of single shooting, since the terminal error is not carried through one long integration, but retain the truncation error of the artificial terminal condition and its sensitivity to the choice of T. Both formulations as described require identifying the relevant steady state, which hysteresis or multiple steady states complicate because the limit may depend on $\boldsymbol{x}_0$.

6.2 *The NN solver for sequential problems*

Parameterization. Here $\mathrm{NN}(\cdot; \theta) : \mathbb{R} \to \mathbb{R}^M$ is the NN of Equation (S44) with scalar input t and $\phi = \tanh$. The values of d (depth), w (width), and the size of θ are in Table 5. Let $\hat{\psi}_1, \ldots, \hat{\psi}_M$ be the unknown functions of time that a model's formulation solves for,

listed below: components of the paths $\hat{\boldsymbol{x}}$, $\hat{\boldsymbol{\mu}}$, $\hat{\boldsymbol{y}}$ themselves, or transformations of them such as Tobin's q in the open economy. Function m applies the rescaled softplus U_1 of Equation (S26) to the m-th output,

$$\hat{\psi}_m(t) = s_m\, U_1\big(\mathrm{NN}_m(\omega t;\theta)\big), \qquad U_1(v) := \frac{1}{\log 2}\log\Big(1+\exp\big(2\log(2)\,v\big)\Big) = \log_2\big(1+4^v\big), \tag{S51}$$

with ω a fixed time scale and s_m a fixed scale in the units of $\hat{\psi}_m$, built from the initial conditions and parameters. Signed functions, those that can take either sign (net foreign assets in the open-economy and multicountry models, and the price in the asset-pricing model), omit U_1. Because this rescaled softplus satisfies $U_1(0)=1$ and $U_1'(0)=1$, an output near zero gives $\hat{\psi}_m$ near s_m, with slope s_m in the output. In the trend-growth variant of Section 2.4 each function also carries a factor $e^{\hat{g}t}$ with $\hat{g}$ fixed or learned. At initialization the output layer W_{d+1} is near zero, so every positive function starts near the constant s_m and every signed one near zero. The functions are in levels, $\dot{\hat{\psi}}_m(t)$ is the exact derivative of Equation (S51) by automatic differentiation in t, and initial conditions enter as residual rows.

Objective. Let $\boldsymbol{r}(\theta)$ stack the rows listed for each model below: the differential and algebraic rows over $t_i \in \mathcal{D}$, and the initial-condition rows $\hat{\boldsymbol{x}}(0) - \boldsymbol{x}_0$ once. The NN solves

$$\min_{\theta}\ \frac{1}{n_r}\big\|\boldsymbol{r}(\theta)\big\|_2^2, \tag{S52}$$

with n_r the number of scalar rows: the least-squares problem of Main Paper Section A with $\lambda = 0$, the initial condition Main Paper Equation (4) as rows rather than pinned, and the rows of each model's DAE after the reductions stated below. There is no norm penalty and no terminal, transversality, or steady-state row. In the growth models the co-state is eliminated through the algebraic condition. In the open-economy models investment is eliminated through Main Paper Equation (24), μ_1 enters through $q = \mu_1/\mu_2$ as in Main Paper Section 4.4, and the differential rows of the positive paths are divided by the path, so that the residual is the error in the growth rate. Each change preserves the zero set but changes the least-squares weighting. The minimization is Levenberg–Marquardt on θ with geodesic acceleration and the identity metric on the parameters, in place of the Gram metric of the kernel solver, using the same implementation (Perla, 2026a): the Jacobian of $\boldsymbol{r}$ in θ is assembled by automatic differentiation and factorized densely.

Rows by model. Linear asset pricing (Main Paper Section 4.3). The one path is the price $\hat{\mu}$, signed with $s = 1$. The dividend path $x(t) = (x_0 + c/g)e^{gt} - c/g$ solves Main Paper Equation (14) in closed form and enters as data, as in the kernel implementation. The one row is $\dot{\hat{\mu}}(t_i) - r\hat{\mu}(t_i) + x(t_i)$. The price has no initial condition.

Neoclassical growth (Main Paper Section 4.1) and its concave–convex variant. The co-state is eliminated through Main Paper Equation (8) in its CRRA form, $\mu = y^{-\gamma}$, which turns Main Paper Equation (7) into the Euler equation in consumption, the level form of the reduction in Section 5.1. The paths $(\hat{x}, \hat{y})$ are positive with scales x_0 and $f(x_0) - \delta x_0$,

and the rows are

$$\dot{\hat{x}} - \big[f(\hat{x}) - \delta\hat{x} - \hat{y}\big], \qquad \dot{\hat{y}} - \hat{y}\,\frac{f'(\hat{x}) - \delta - r}{\gamma}, \qquad \hat{x}(0) - x_0, \tag{S53}$$

with each model's production function. Section 2.4 states the trend-growth variant.

Small open economy (Main Paper Section 4.4). The paths are the reduced system $(\hat{x}_1, \hat{x}_2, \hat{q}, \hat{\mu}_2, \hat{y}_1)$ of that section, with $\hat{x}_2$ signed and scales $\big(k_0,\ f(k_0),\ 1/\eta,\ c_s^{-\gamma},\ c_s\big)$ for $c_s := f(k_0) - \delta k_0$. With the investment rate $\hat{y}_2/\hat{x}_1 = \big(\eta\chi\hat{q}\big)^{\frac{1}{1-\eta}}$ from Main Paper Equation (24), the rows are

$$\begin{aligned}
&\frac{\dot{\hat{x}}_1}{\hat{x}_1} - \Big[\chi\Big(\frac{\hat{y}_2}{\hat{x}_1}\Big)^{\eta} - \delta\Big], && \dot{\hat{x}}_2 - \big[r^\star\hat{x}_2 + f(\hat{x}_1) - \hat{y}_1 - \hat{y}_2\big],\\
&\frac{\dot{\hat{q}}}{\hat{q}} - \Big[r^\star + \delta - (1-\eta)\,\chi\Big(\frac{\hat{y}_2}{\hat{x}_1}\Big)^{\eta} - \frac{f'(\hat{x}_1)}{\hat{q}}\Big], && \frac{\dot{\hat{\mu}}_2}{\hat{\mu}_2} - \big(r - r^\star\big),\\
&\hat{\mu}_2\,\hat{y}_1^{\gamma} - 1, && \hat{x}_1(0) - k_0, \qquad \hat{x}_2(0) - b_0,
\end{aligned} \tag{S54}$$

the five country rows of Equation (S8) for one country with $z_i = 1$ and a fixed $r^\star$, with the $\dot{\hat{x}}_1$, $\dot{\hat{q}}$, and $\dot{\hat{\mu}}_2$ rows divided by $\hat{x}_1$, $\hat{q}$, and $\hat{\mu}_2$.

Multicountry (Main Paper Section 4.5). The same five paths per country with $z_i f$ and γ_i in place of f and γ, and the world rate $r^\star(t)$ as a further positive path with scale r: $5N + 1$ outputs. The capital scale is $k_{i,0}$, the asset and consumption scales are the population-weighted means of $z_i f(k_{i,0})$ and of $c_{s,i} := z_i f(k_{i,0}) - \delta k_{i,0}$, and $\hat{\mu}_{2,i}$ is scaled by $\bar{c}_s^{-\gamma_i}$. The rows are the five differential and algebraic rows of Equation (S54) for every country with the rate path in place of $r^\star$, the clearing row $\sum_i L_i\hat{x}_{2,i}(t_j)\big/\sum_i L_i$ at every $t_j \in \mathcal{D}$, and the $2N$ initial-condition rows with the $b_{i,0}$ population-demeaned.

Hyperparameters. Table 5 lists the settings. The time scale is $\omega = 0.01$ (0.025 for asset pricing). Damping starts at 10^{-5}, is halved after an accepted step, and is doubled after a rejected one. For asset pricing it starts at 10^{-3} and quadruples after a rejection. The optimization converges when the residual norm of Equation (S52) falls below $10^{-7.5}$ (10^{-7} for the open-economy and trend-growth models). One that does not converge within 400 steps (250 for the growth models), or whose loss becomes nonfinite, restarts from a fresh initialization, up to five attempts in all. Acceptance is checked afterwards: the largest absolute DAE-row residual on $\mathcal{D}$ (initial-condition rows excluded) and the mean absolute residual on a uniform validation grid of 250 points (450 for the open-economy models) must be below 10^{-5} (10^{-6} for asset pricing) and 10^{-6} respectively, and in the open-economy models every path must be finite, and every path but the asset position positive, on the training, validation, and test grids. The multicountry runs omit the validation threshold. Restarts respond only to the optimizer, so a converged solution that fails the thresholds is reported as rejected rather than re-run, and the acceptance counts of Table 6 record how often the returned solution passes its acceptance check. The NN is run with seven seeds, with a single seed at each of the 70 initial conditions of the concave–convex model and for the open economy's initial-debt and initial-asset runs. The open-economy grids follow the kernel's sampling, concentrated at early dates, with 81 points on $[0, 40]$ and 15 on $(40, 160]$.

TABLE 5. Settings of the sequential NN solver: width $\times$ depth of the hidden layers, parameter count, and collocation grid.

Model	$w \times d$	Params	$\mathcal{D}$
Asset pricing	16×2	321	41 on $[0, 40]$
Neoclassical growth	32×3	2242	101 on $[0, 100]$
Concave–convex	32×3	2242	41 on $[0, 40]$
Open economy	32×3	2341	96 on $[0, 160]$
Multicountry, $N = 2$	32×3	2539	96 on $[0, 160]$

6.3 *Kernel and NN solvers on the sequential models*

Table 6 compares the two solvers on the sequential models of Main Paper Section 4: accuracy against each model's benchmark, fit time, and acceptance counts (the NN is run with seven seeds, while the kernel solutions are deterministic). Table 7 compares all four solvers for the neoclassical growth model, sequential and recursive, kernel and NN, which share the calibration of Section 5.5 exactly. Errors are against the common benchmark over each solver's own test window, $[0, 110]$ for the sequential solvers and $[0, 40]$ for the recursive ones, and fit times are for each solver's own design.

TABLE 6. Kernel and NN solvers on the sequential models. Errors are absolute relative errors in capital against each model's benchmark (in the price for asset pricing). NN entries are medians over accepted seeds. The concave–convex rows aggregate over the 70 initial conditions, every one of which reaches the same basin as the benchmark with both solvers, the mean averaging the per-initial-condition means and the maximum taken over all of them. Times are medians over accepted runs.

Model	Solver	Mean rel. error	Max rel. error	Time (s)	Accepted
Asset pricing	Kernel	5.7e-04	1.7e-03	0.06	1/1
	NN	4.3e-05	2.0e-04	0.018	7/7
Neoclassical growth	Kernel	1.9e-04	1.5e-03	0.14	1/1
	NN	2.6e-07	3.2e-06	0.33	7/7
Concave–convex	Kernel	1.5e-04	4.0e-03	0.0021	70/70
	NN	2.6e-06	8.4e-05	0.37	70/70
Open economy	Kernel	8.0e-05	2.4e-04	0.16	1/1
	NN	3.3e-07	1.8e-06	0.94	7/7

7. SENSITIVITY ANALYSIS VIA IMPLICIT DIFFERENTIATION

The structural parameters and initial conditions enter the collocation residuals as explicit arguments. The solver's stationarity condition at the returned solution therefore defines the solution as an implicit function of them. By the implicit function theorem, the derivative of any functional of the solved paths with respect to every parameter and initial condition follows at once. Each scalar observable costs one reverse-mode

TABLE 7. The neoclassical growth model across the four solvers, sharing one calibration: absolute relative errors against the boundary-value benchmark over each solver's own test window, and fit times in seconds (NN entries are medians over accepted seeds).

Solver	Mean rel. error: k	Max rel. error: k	Mean rel. error: c	Time (s)	Accepted
Sequential kernel	1.9e-04	1.5e-03	1.1e-04	0.14	1/1
Sequential NN	2.6e-07	3.2e-06	1.0e-07	0.33	7/7
Recursive kernel	3.2e-05	7.4e-05	1.6e-05	0.27	1/1
Recursive NN	1.3e-06	2.2e-06	4.7e-07	0.14	5/5

pass. There is no re-solving at perturbed parameters and no finite differencing. The tangent system uses the Gauss–Newton operator of the ridge objective, which drops the residual-weighted second derivatives and is exact when the residual vanishes. For the NNs, whose parameters outnumber the residual rows, the undamped tangent system at the trained parameters is rank deficient and is solved by a truncated singular value decomposition pseudoinverse, which selects the minimum-norm parameter tangent. Because no steady-state, transversality, or other terminal information is supplied to the solvers, comparative statics that theory pins down are out-of-sample checks on the differentiated solution.

To check the differentiated solution against something known, we use the baseline calibration and solvers of Main Paper Section 4.1, extend the training grid to $\mathcal{D} = \{0, 1, \ldots, 120\}$, and differentiate $x(T)$ and $y(T)$ at $T = 120$ with respect to $\theta \in \{a, \delta, r, \gamma, x_0\}$. At that horizon the solution has reached the steady state, whose comparative statics are available in closed form. The steady state solves $f'(x^\star) = \delta + r$, so

$$x^\star = \left(\frac{a}{\delta + r}\right)^{\frac{1}{1-a}}, \qquad y^\star = (x^\star)^a - \delta x^\star, \tag{S55}$$

and differentiating, using $a\,(x^\star)^{a-1} - \delta = r$,

$$\frac{\partial x^\star}{\partial \delta} = \frac{\partial x^\star}{\partial r} = -\frac{x^\star}{(1-a)\,(\delta + r)}, \qquad \frac{\partial x^\star}{\partial a} = x^\star \left[\frac{1}{a(1-a)} + \frac{\log\left(a/(\delta + r)\right)}{(1-a)^2}\right], \tag{S56}$$

$$\frac{\partial y^\star}{\partial \delta} = r\,\frac{\partial x^\star}{\partial \delta} - x^\star, \qquad \frac{\partial y^\star}{\partial r} = r\,\frac{\partial x^\star}{\partial r}, \qquad \frac{\partial y^\star}{\partial a} = (x^\star)^a \log x^\star + r\,\frac{\partial x^\star}{\partial a}. \tag{S57}$$

Neither γ nor x_0 appears, so their steady-state derivatives are zero, and the finite-T sensitivities converge to Equations (S56) and (S57) as the transition decays. At $T = 120$, $x(120) = 3.60344$ against $x^\star = 3.60348$.

Table 8 reports the comparison. The nonzero rows agree with the closed forms to an absolute gap of at most 1.49×10^{-3}, and the two solvers agree with each other to 4.08×10^{-4}. The γ and x_0 rows are numerical zeros, at most 3.93×10^{-5} in absolute value.

TABLE 8. Sensitivities of the neoclassical growth model solution at $T = 120$, differentiated through the kernel and NN solutions (NN entries are medians over seven seeds), against the closed-form steady-state derivatives Equations (S56) and (S57). The γ and x_0 rows are exactly zero in the closed form.

	$\partial x(T)/\partial\theta$			$\partial y(T)/\partial\theta$		
θ	Kernel	NN	Closed Form	Kernel	NN	Closed Form
a	23.14	23.14	23.14	2.933	2.933	2.933
δ	-38.11	-38.11	-38.11	-5.197	-5.197	-5.197
r	-38.11	-38.11	-38.11	-1.594	-1.594	-1.594
γ	-3.064×10^{-5}	-3.926×10^{-5}	0	-6.692×10^{-6}	-4.982×10^{-6}	0
x_0	3.899×10^{-5}	4.345×10^{-6}	0	-2.948×10^{-6}	7.029×10^{-7}	0

REFERENCES

Aruoba, S. Borağan, Jesús Fernández-Villaverde, and Juan F. Rubio-Ramírez (2006), "Comparing solution methods for dynamic equilibrium economies." *Journal of Economic Dynamics and Control*, 30 (12), 2477–2508. []

Brenan, Kathryn E, Stephen L Campbell, and Linda R Petzold (1996), *Numerical Solution of Initial-Value Problems in Differential-Algebraic Equations*, volume 14 of *Classics in Applied Mathematics*. SIAM, Philadelphia. []

Chen, Ricky TQ, Yulia Rubanova, Jesse Bettencourt, and David K Duvenaud (2018), "Neural ordinary differential equations." *Advances in neural information processing systems*, 31, 6571–6583. []

Ebrahimi Kahou, Mahdi, Jesús Fernández-Villaverde, Sebastián Gómez-Cardona, Jesse Perla, and Jan Rosa (2024), "Spooky boundaries at a distance: Inductive bias, dynamic models, and behavioral macro." Working Paper 32850, National Bureau of Economic Research. []

Fernández-Villaverde, Jesús, Pablo Guerrón-Quintana, Juan F. Rubio-Ramírez, and Martín Uribe (2011), "Risk matters: The real effects of volatility shocks." *American Economic Review*, 101 (6), 2530–2561. []

King, Robert G. and Sergio T. Rebelo (1999), "Resuscitating real business cycles." In *Handbook of Macroeconomics*, volume 1, 927–1007, Elsevier. []

Li, Xuechen, Ting-Kam Leonard Wong, Ricky T. Q. Chen, and David Duvenaud (2020), "Scalable gradients for stochastic differential equations." In *Proceedings of the 23rd International Conference on Artificial Intelligence and Statistics*, volume 108 of *Proceedings of Machine Learning Research*, 3870–3882. []

Mendoza, Enrique G. and Sergio Villalvazo (2020), "FiPIt: A simple, fast global method for solving models with two endogenous states & occasionally binding constraints." *Review of Economic Dynamics*, 37, 81–102. []

Michel, Phillippe (1982), "On the transversality condition in infinite horizon optimal problems." *Econometrica: Journal of the Econometric Society*, 50 (4), 975–985. []

Murphy, Kevin P. (2023), *Probabilistic Machine Learning: Advanced Topics*. Adaptive Computation and Machine Learning, MIT Press, URL https://github.com/probml/pml-book. []

Pantelides, Constantinos C. (1988), "The consistent initialization of differential-algebraic systems." *SIAM Journal on Scientific and Statistical Computing*, 9 (2), 213–231, URL https://doi.org/10.1137/0909014. []

Perla, Jesse (2026a), "nlls_gram: Metric-aware Levenberg–Marquardt nonlinear least squares for JAX." https://github.com/HighDimensionalEconLab/nlls_gram. Version 2.10.0. []

Perla, Jesse (2026b), "tinydiffeq: Differentiable ODE/SDE/DAE solvers for JAX." https://github.com/HighDimensionalEconLab/tinydiffeq. Version 2.6.1. []

Rößler, Andreas (2010), "Runge–kutta methods for the strong approximation of solutions of stochastic differential equations." *SIAM Journal on Numerical Analysis*, 48 (3), 922–952. []

Uribe, Martín and Vivian Z. Yue (2006), "Country spreads and emerging countries: Who drives whom?" *Journal of International Economics*, 69 (1), 6–36. []