Introduction: Solution Methods for State-Dependent and Time-Dependent Models

Lilia Maliar and Serguei Maliar

CEF 2017 Workshop

State-dependent models

- Dynamic stochastic economic models are normally built on the assumption of stationary environment.
- Namely, the economy's fundamentals such as preferences, technologies and laws of motions for exogenous variables do not change over time (or there exists a transformation to stationary environment, such as balanced growth).
- Such models have stationary solutions in which optimal value and decision functions depend on the current state but not on time.
- The state-dependent class of models is convenient for applied work since time-invariant solutions are relatively easy to construct.

Example of state-dependent model

Standard neoclassical growth model:

$$\max_{\{c_t, k_{t+1}\}_{t=0}^{\infty}} E_0 \left[\sum_{t=0}^{\infty} \beta^t u(c_t) \right]$$
$$\text{s.t. } c_t + k_{t+1} = (1-\delta) k_t + f(k_t, z_t),$$
$$z_{t+1} = \varphi(z_t, \varepsilon_{t+1}),$$

– $c_t \geq 0$ and $k_t \geq 0$ are consumption and capital, resp.;
– initial condition (k_0, z_0) is given;
– $u : \mathbb{R}_+ \to \mathbb{R}$ and $f : \mathbb{R}_+^2 \to \mathbb{R}_+$ and $\varphi : \mathbb{R}^2 \to \mathbb{R}$ are **time-invariant** utility function, production function and law of motion for exogenous state variable z_t, resp.;
– ε_{t+1} is i.i.d;
– $\beta \in (0,1)$ = discount factor; $\delta \in [0,1]$ =depreciation rate; $E_t[\cdot]$ = operator of expectation.

Example of state-dependent model (cont.)

Under standard assumptions, a solution to stationary neoclassical growth model is:

- One time-invariant value function $V(k_t, z_t)$.
- One set of time-invariant policy functions, e.g., $k_{t+1} = K(k_t, z_t)$.

Time-dependent models

- At the same time, real-world economies constantly evolve over time, experiencing
  - population growth,
  - technological progress,
  - trends in tastes and habits,
  - policy regime changes,
  - evolution of social and political institutions, etc.
- Also, economic policies change over time, for example, Central Banks can change parameters in the Taylor rule or employ time-dependent unconventional monetary policies such as quantitative easing or forward guidance.
- If the parameters change over time, the resulting models are generally nonstationary, and their optimal value and decision functions are time-dependent.

Example of time-dependent model

Infinitely-lived neoclassical growth model with time-varying fundamentals

$$\max_{\{c_t, k_{t+1}\}_{t=0}^{\infty}} E_0 \left[\sum_{t=0}^{\infty} \beta^t u_t (c_t) \right]$$
$$\text{s.t. } c_t + k_{t+1} = (1-\delta) k_t + f_t (k_t, z_t),$$
$$z_{t+1} = \varphi_t (z_t, \varepsilon_{t+1}),$$

– $c_t \geq 0$ and $k_t \geq 0$ are consumption and capital, resp.;
– initial condition (k_0, z_0) is given;
– $u_t : \mathbb{R}_+ \to \mathbb{R}$ and $f_t : \mathbb{R}_+^2 \to \mathbb{R}_+$ and $\varphi_t : \mathbb{R}^2 \to \mathbb{R}$ are **time-varying** utility function, production function and law of motion for exogenous state variable z_t, resp.;
– sequence of u_t, f_t and φ_t for $t \geq 0$ is known to the agent in period $t = 0$; ε_{t+1} is i.i.d;
– $\beta \in (0, 1)$ = discount factor; $\delta \in [0, 1]$ =depreciation rate; $E_t [\cdot]$ = operator of expectation.

Example of time-dependent model (cont.)

If a model is nonstationary and time-dependency is nontrivial, a solution is:

- An infinite-sequence of time-varying value functions: $V_0(k_0, z_0)$, $V_1(k_1, z_1), \ldots$
- An infinite-sequence of time-varying policy functions: $k_1 = K_0(k_0, z_0)$, $k_1 = K_0(k_0, z_0)$,...

- Conventional numerical methods used for state-dependent models are not directly suitable for analyzing time-dependent models.

Why cannot we solve a nonstationary model with conventional solution methods?

A stationary growth model (dynamic-programming formulation):

$$V(k,z) = \max_{c,k'} \left\{ u(c) + \beta E\left[V\left(k', z'\right)\right]\right\}$$

$$\text{s.t. } k' = (1-\delta)k + zf(k) - c,$$

$$\ln z' = \rho \ln z + \varepsilon', \qquad \varepsilon' \sim \mathcal{N}\left(0, \sigma^2\right).$$

An interior solution satisfies the Euler equation:

$$u'(c) = \beta E\left[u'\left(c'\right)\left(1 - \delta + z' f'\left(k'\right)\right)\right].$$

- Conventional solution methods: *either iterate on Bellman equation until a fixed-point V is found or iterate on Euler equation until a fixed-point decision function $k' = K(k,z)$ is found.*
- However, if u, f, ρ and σ are time-dependent, then $V_t(\cdot) \neq V_{t+1}(\cdot)$ and $K_t(\cdot) \neq K_{t+1}(\cdot)$, i.e., no fixed-point functions V and K.
- We need to construct a sequence (path) of time-dependent value functions $(V_0(\cdot), V_1(\cdot), ...)$, decision functions $(K_0(\cdot), K_1(\cdot), ...)$.

This workshop

- In Part 1 of this workshop, we review numerical techniques for state-dependent models with an emphasis on problems with a large number of state variables, including:
  - grid techniques *(Smolyak, simulated, cluster, epsilon-distinguishable sets and low-discrepancy sequences)*,
  - integration methods *(quadrature, monomial formulas, Monte Carlo)*,
  - numerically stable approximation techniques *(singular value decomposition (SVD), principal component (PC) approach, linear programming, Tykhonov and other types of regularization, truncated SVD and PC methods)*,
  - alternative iterative procedures *(including endogenous grid and envelope condition methods)*,
  - precomputation techniques *(integrals and intratemporal choice functions)*.
- We illustrate these methods by examples of one and multi-agent neoclassical growth models, as well as a large-scale new Keynesian model.

This workshop (cont.)

- In Part 2, we show a quantitative framework, called extended function path (EFP), for calibrating, solving, simulating and estimating time-dependent models.
- We apply EFP to solve a collection of challenging nonstationary time-dependent and unbalanced-growth applications, including:
  - stochastic growth models with parameters shifts and drifts,
  - capital augmenting technological progress,
  - anticipated regime switches,
  - time-trends in volatility of shocks,
  - seasonal fluctuations,
  - new Keynesian economies with time-varying parameters.
- Also, we show an example of estimation and calibration of parameters in an unbalanced growth model using the data on the U.S. economy.

Papers presented

For general background on global solution methods for large-scale models, we will use:

- Lilia Maliar and Serguei Maliar, (2014). "Numerical methods for large scale dynamic economic models", in: Schmedders, K. and K.L. Judd (Eds.), Handbook of Computational Economics, Volume 3, Chapter 7, 325-477, Amsterdam: Elsevier Science.

Papers presented

Other papers on state-dependent large-scale models that we will cover are:

1. Kenneth L. Judd, Lilia Maliar and Serguei Maliar, (2011). Numerically stable and accurate stochastic simulation approaches for solving dynamic models. Quantitative Economics 2, 173-210.
2. Kenneth L. Judd, Lilia Maliar, Serguei Maliar and Rafael Valero, (2014). "Smolyak method for solving dynamic economic models: Lagrange Interpolation, anisotropic grid and adaptive domain", Journal of Economic Dynamic and Control 44(C), 92-123.
3. Lilia Maliar and Serguei Maliar, (2015). "Merging simulation and projection aproaches to solve high-dimensional problems with an application to a new Keynesian model", Quantitative Economics 6, 1-47.
4. Kenneth L. Judd, Lilia Malia, Serguei Malia and Inna Tsener, (2016). "How to solve dynamic stochastic models computing expectations just once"", Quantitative Economics (forthcoming).

Papers presented (cont.)

5. Lilia Maliar and Serguei Maliar, (2013). "Envelope Condition Method versus Endogenous Grid Method for Solving Dynamic Programming Problems", Economic Letters 120, 262-266.
6. Cristina Arellano, Lilia Maliar, Serguei Maliar and Viktor Tsyrennikov, (2016). "Envelope condition method with an application to default risk models", Journal of Economic Dynamics and Control 69, 436-459.
7. Kenneth L. Judd, Lilia Maliar and Serguei Maliar, (2016). "Lower bounds on approximation errors to numerical solutions of dynamic economic models", Econometrica (forthcoming).
8. Vadym Lepetuyk, Lilia Maliar and Serguei Maliar (2017). "Should central banks worry about nonlinearities of their large-scale macroeconomic models?", Bank of Canada working paper 2017-21.

Papers presented (cont.)

Time-dependent models are analyzed by using the EFP framework developed in:

1. Lilia Maliar, Serguei Maliar, John B. Taylor and Inna Tsener (2015). "A tractable framework for analyzing a class of nonstationary Markov models", NBER 21155.
2. Lilia Maliar (2016). "Forward guidence puzzle and turnpike theorem", manuscript.
3. Lilia Maliar, Serguei Maliar, John B. Taylor and Inna Tsener (2017). "Extended function path method", manuscript.

Computer codes

Please, download the code from https://stanford.edu/~maliarl/Codes.html

- "*GSSA_Two_Models.zip*" - *Generalized Stochastic Simulation Algorithm (GSSA)*,
- "*ECM_and_EGM_MM_2013.zip*" - *Envelope condition and endogeneous grid for growth model with valued leisure*,
- "*7_methods_for_growth_model_AMMT_2016.zip*" - *Comparison of 7 iterative methods for a growth model (including value iteration, policy iteration, Euler equation, envelope condition and endogenous grid)*,
- "*Smolyak_Anisotropic_JMMV_2014.zip*" - *Smolyak method*,
- "*EDSCGA_Maliars_QE6_2015.zip*" - *Epsilon-distingushable set and cluster-grid methods*,
- "*Precomputation_JMMT_QE_2016.zip*" - Precomputation of integrals (= get rid off expectations before solving the model),
- "*EFP_MMTT_2015.zip*" - *Extended Function Path (EFP) method for time-dependent models.*

Model with elastic labor supply: a divisible-labor version

We consider a standard growth *model with elastic labor supply*. The agent solves:

$$\max_{\{k_{t+1},c_t,\ell_t\}_{t=0,\ldots,\infty}} E_0\left\{\sum_{t=0}^{\infty}\beta^t u\left(c_t,\ell_t\right)\right\}$$
$$\text{s.t. } c_t + k_{t+1} = (1-\delta)\,k_t + \theta_t f\left(k_t,\ell_t\right),$$
$$\ln\theta_{t+1} = \rho\ln\theta_t + \sigma\epsilon_{t+1}, \qquad \epsilon_{t+1}\sim\mathcal{N}(0,1),$$

where initial condition (k_0,θ_0) is given;
$f(\cdot)$ = production function;
c_t = consumption; k_{t+1} = capital; θ_t = productivity level;
β = discount factor; δ = depreciation rate of capital;
ρ = autocorrelation coefficient of the productivity level;
σ = standard deviation of the productivity shock ϵ_{t+1}.

Model with elastic labor supply: a divisible-labor version (cont.)

- Assume that the agents value leisure.
  $1 =$ total time endowment,
  $l_t =$ leisure,
  $\ell_t =$ working hours.
- The agent can choose any number of working hours between 0 and 1.

$$\ell_t + l_t = 1.$$

- $u(c_t, l_t) =$ the momentary utility (strictly increasing, and concave).
- A common assumption is the CRRA utility function:

$$u(c_t, l_t) = \frac{\left(c_t^v l_t^{1-v}\right)^{1-\sigma} - 1}{1-\sigma},$$

  $v =$ share of consumption; $\sigma =$ coefficient of relative risk aversion.
- If $\sigma = 1$, then $u(c_t, l_t) = \ln c_t + A \ln l_t$.

Time invariant decision functions

- Our goal is to solve for a recursive Markov equilibrium in which the decisions on next-period capital, consumption and labor are made according to some time invariant state contingent functions

$$k' = K(k, \theta), \quad c = C(k, \theta), \quad \ell = \mathcal{L}(k, \theta).$$

- A version of model in which the agent does not value leisure and supplies to the market all her time endowment is referred to as *a model with inelastic labor supply*.
- Such model is obtained by replacing $u(c_t, \ell_t)$ and $f(k_t, \ell_t)$ with $u(c_t)$ and $f(k_t)$, respectively.

First-order conditions

We assume that a solution to the model is interior and satisfies budget constraint

$$c_t + k_{t+1} = (1 - \delta) k_t + \theta_t f(k_t, \ell_t)$$

and the first-order conditions (FOCs)

$$u_1(c_t, \ell_t) = \beta E_t \{ u_1(c_{t+1}, \ell_{t+1}) [1 - \delta + \theta_{t+1} f_1(k_{t+1}, \ell_{t+1})] \}, \quad (1)$$

$$u_2(c_t, \ell_t) = u_1(c_t, \ell_t) \theta_t f_2(k_t, \ell_t). \quad (2)$$

- FOC (1) is the *Euler equation* or inter-temporal FOC (relates variables of different periods).
- FOC (2) is intra-temporal FOC (relates variables within the same period).

Three broad classes of numerical methods

1. Projection methods, Judd (1992), Christiano and Fisher (2000), etc.
   - solution domain = prespecified grid of points;
   - accurate and fast with few state variables but cost grows exponentially with the number of state variables (curse of dimensionality!).
2. Perturbation methods, Judd and Guu (1993), Gaspar and Judd (1997), Juillard (2003), etc.
   - solution domain = one point (steady state);
   - practical in large-scale models but the accuracy can deteriorate dramatically away from the steady state.
3. Stochastic simulation methods, Marcet (1988), Smith (2001), Judd et al. (2011), etc.
   - solution domain = simulated series;
   - simple to program but often numerically unstable, and the accuracy is lower than that of the projection methods.

An example of a global projection-style Euler equation method

- We approximate functions K, C, and $\mathcal{L}$ numerically.
- Let us consider a projection-style method in line with Judd (1992) that approximates these functions to satisfy the FOCs on a grid of points.

An outline of a global projection-style Euler equation method

(EEM): A global projection-style Euler equation method.

Step 1. Choose functional form $\widehat{K}(\cdot, b)$ for representing K, where b is the coefficients vector.
Choose a grid $\{k_m, \theta_m\}_{m=1,...,M}$ on which $\widehat{K}$ is constructed.
Step 2. Choose nodes, ϵ_j, and weights, ω_j, $j = 1, ..., J$, for approximating integrals. Compute next-period productivity $\theta'_{m,j} = \theta_m^{\rho} \exp(\epsilon_j)$ for all j, m.
Step 3. Solve for b that approximately satisfies the model's equations:

$$u_1(c_m, \ell_m) = \beta \sum_{j=1}^{J} \omega_j \cdot \left[u_1\left(c'_{m,j}, \ell'_{m,j}\right) \left(1 - \delta + \theta'_{m,j} f_1\left(k'_m, \ell'_{m,j}\right)\right)\right],$$

$$u_2(c_m, \ell_m) = u_1(c_m, \ell_m)\, \theta_m f_2(k_m, \ell_m),$$

$$c_m = (1-\delta) k_m + \theta_m f(k_m, \ell_m) - k'_m$$

$$u_2\left(c'_{m,j}, \ell'_{m,j}\right) = u_1\left(c'_{m,j}, \ell'_{m,j}\right) \theta'_{m,j} f_2\left(k'_m, \ell'_{m,j}\right),$$

$$c'_{m,j} = (1-\delta) k'_m + \theta'_{m,j} f\left(k'_m, \ell'_{m,j}\right) - k''_{m,j}$$

Discussion of Step 3

In Step 3, we use the assumed decision function $\widehat{K}(\cdot, b)$:

- the choices in the current period $k'_m = \widehat{K}(k_m, \theta_m; b) \equiv k'_m(b)$;
- future shocks $\theta'_{m,j} = \theta_m^{\rho} \exp(\epsilon_j)$
- the choices in J possible future states $k''_{m,j} = \widehat{K}\left(k'_m, \theta'_{m,j}; b\right)$

$$u_1(c_m, \ell_m) = \beta \sum_{j=1}^{J} \omega_j \left[u_1\left(c'_{m,j}, \ell'_{m,j}\right) \left(1 - \delta + \theta'_{m,j} f_1\left(k'_m(b), \ell'_{m,j}\right)\right)\right.$$

$$u_2(c_m, \ell_m) = u_1(c_m, \ell_m)\, \theta_m f_2(k_m, \ell_m),$$

$$c_m = (1-\delta) k_m + \theta_m f(k_m, \ell_m) - k'_m(b)$$

$$u_2\left(c'_{m,j}, \ell'_{m,j}\right) = u_1\left(c'_{m,j}, \ell'_{m,j}\right) \theta'_{m,j} f_2\left(k'_m(b), \ell'_{m,j}\right),$$

$$c'_{m,j} = (1-\delta) k'_m(b) + \theta'_{m,j} f\left(k'_m(b), \ell'_{m,j}\right) - \widehat{K}\left(k'_m(b), \theta'_{m,j}; b\right)$$

- We must use these $2J+3$ equations to identify $2J+2$ unknowns $c_m, \ell_m, \left\{c'_{m,j}, \ell'_{m,j}\right\}_{j=1}^{J}$ and the coefficients b.

Unidimensional grid points and basis functions

- To solve the model, we discretize the state space in Step 1 into a finite set of grid points $\{k_m, \theta_m\}_{m=1,...,M}$.
- Our construction of a multidimensional grid begins with unidimensional grid points and basis functions.
- The simplest possible choice is a family of ordinary polynomials and a grid of uniformly spaced points.
- However, many other choices are possible.
- In particular, a useful alternative is a family of Chebyshev polynomials and a grid composed of extrema of Chebyshev polynomials.
- Such polynomials are defined in the interval $[-1, 1]$, and thus, the model's variables such as k and θ must be rescaled to be inside this interval prior to any computation.

Unidimensional grid of uniformly spaced points and ordinary polynomials

Table: Unidimensional grid of uniformly spaced points and ordinary polynomials

n	Ordinary polyn. of degree $n-1$	Uniform grid of n points on $[-1,1]$
1	1	0
2	x	$-1, \quad 1$
3	x^2	$-1 \quad 0 \quad 1$
4	x^3	$-1, \quad -\frac{2}{3}, \quad \frac{2}{3}, \quad 1$
5	x^4	$-1 \quad -\frac{1}{2} \quad 0 \quad -\frac{1}{2} \quad 1$

Notes: Ordinary polynomial of degree $n-1$ is given by $P_{n-1}(x) = x^{n-1}$.

Unidimensional Chebyshev polynomials and a grid of their extrema

Table: Unidimensional Chebyshev polynomials and a grid of their extrema

n	Chebyshev polyn. of degree $n-1$	n extrema of Chebyshev polyn. of degree $n-1$
1	1	0
2	x	$-1, \quad 1$
3	$2x^2-1$	$-1 \quad 0 \quad 1$
4	$4x^3-3x$	$-1, \quad -\frac{1}{2}, \quad \frac{1}{2}, \quad 1$
5	$8x^4-8x^2+1$	$-1 \quad -\frac{1}{\sqrt{2}} \quad 0 \quad -\frac{1}{\sqrt{2}} \quad 1$

Notes: Chebyshev polynomial of degree $n-1$ is given by $T_{n-1}(x) = cos((n-1)cos^{-1}(x))$; and finally, n extrema of Chebyshev polynomials of degree $n-1$ are given by $\zeta_j^n = -cos(\pi(j-1)/(n-1)), j = 1, ..., n$.

Ordinary versus Chebyshev polynomials

- As we see, Chebyshev polynomials are just linear combinations of ordinary polynomials.
- If we had an infinite arithmetic precision on a computer, it would not matter which family of polynomials we use.
- But with a finite number of floating points, Chebyshev polynomials have an advantage over ordinary polynomials.

Ordinary versus Chebyshev polynomials

Figure 1a. Ordinary polynomials.

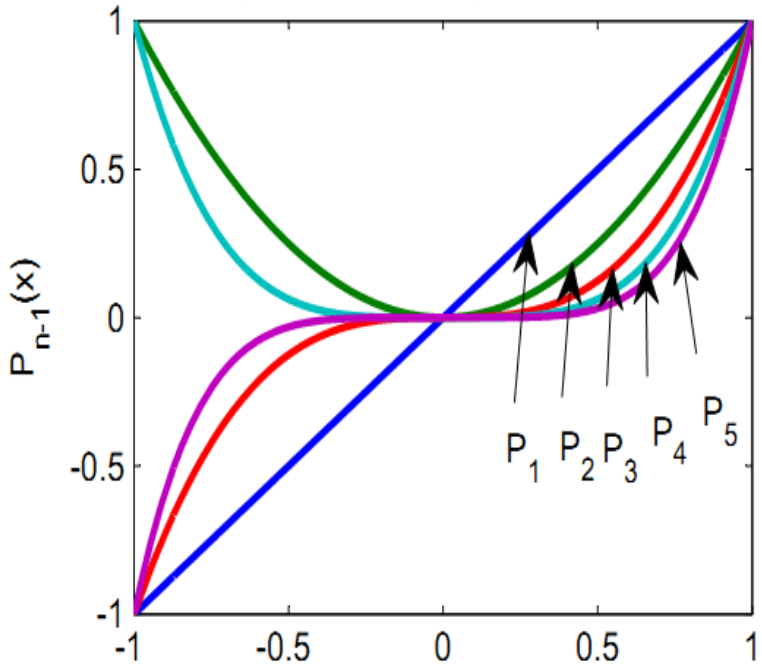
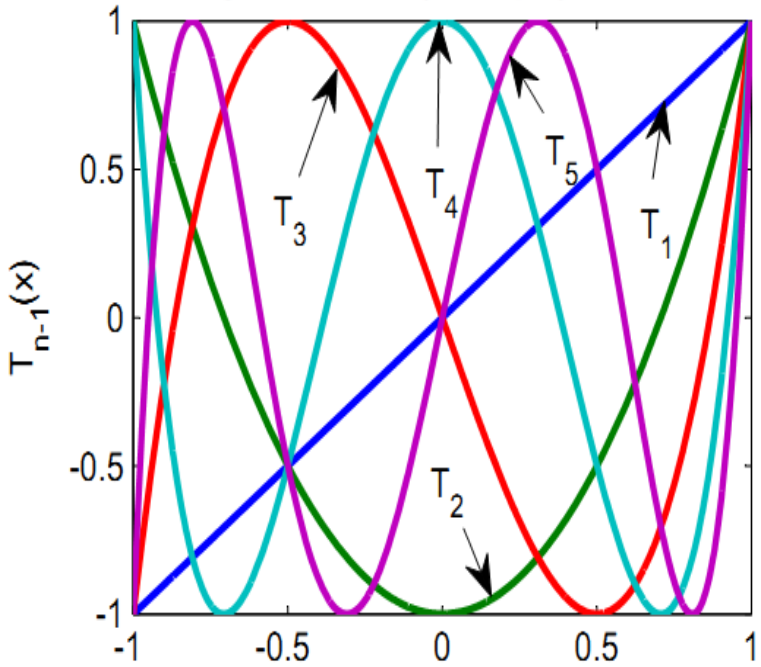


Figure 1b. Chebyshev polynomials.



Comments about unidimensional grid points and basis functions

- For the ordinary polynomial family, the basis functions look very similar on $\mathbb{R}_+$.
- Approximation methods using ordinary polynomials may fail because they cannot distinguish between similarly shaped polynomial terms such as x^2 and x^4.
- In contrast, for the Chebyshev polynomial family, basis functions have very different shapes and are easy to distinguish.

Chebyshev polynomials for approximation

Let us illustrate the use of Chebyshev polynomials for approximation by way of example.

Example

Let $f(x)$ be a function defined on an interval $[-1,1]$, and let us approximate this function with a Chebyshev polynomial function of degree two, i.e.,

$$f(x) \approx \widehat{f}(x;b) = b_1 + b_2 x + b_3\left(2x^2 - 1\right).$$

We compute $b \equiv (b_1, b_2, b_3)$ so that $\widehat{f}(\cdot;b)$ and f coincide in three extrema of Chebyshev polynomials, namely, $\{-1,0,1\}$,

$$\begin{aligned}
\widehat{f}(-1;b) &= b_1 + b_2\cdot(-1) + b_3\left(2\cdot(-1)^2 - 1\right) = f(-1)\\
\widehat{f}(0;b) &= b_1 + b_2\cdot 0 + b_3\left(2\cdot 0^2 - 1\right) = f(0)\\
\widehat{f}(1;b) &= b_1 + b_2\cdot 1 + b_3\left(2\cdot 1^2 - 1\right) = f(1).
\end{aligned}$$

Example

(cont.) This leads us to a system of three linear equations with three unknowns that has a unique solution

$$\begin{bmatrix} b_1 \\ b_2 \\ b_3 \end{bmatrix} = \begin{bmatrix} 1 & -1 & 1 \\ 1 & 0 & -1 \\ 1 & 1 & 1 \end{bmatrix}^{-1} \begin{bmatrix} f(-1) \\ f(0) \\ f(1) \end{bmatrix}$$

$$= \begin{bmatrix} \frac{1}{4} & \frac{1}{2} & \frac{1}{4} \\ -\frac{1}{2} & 0 & \frac{1}{2} \\ \frac{1}{4} & -\frac{1}{2} & \frac{1}{4} \end{bmatrix} \begin{bmatrix} f(-1) \\ f(0) \\ f(1) \end{bmatrix} = \begin{bmatrix} \frac{f(-1)}{4} + \frac{f(0)}{2} + \frac{f(1)}{4} \\ -\frac{f(-1)}{2} + \frac{f(1)}{2} \\ \frac{f(-1)}{4} - \frac{f(0)}{2} + \frac{f(1)}{4} \end{bmatrix}.$$

- It is possible to use Chebyshev polynomials with other grids, but the grid of extrema (or zeros) of Chebyshev polynomials is a perfect match.

Multidimensional grid points and basis functions

- In Step 1 of the Euler equation algorithm, we must specify a method for approximating, representing, and interpolating two-dimensional functions.
- A tensor-product method constructs multidimensional grid points and basis functions using all possible combinations of unidimensional grid points and basis functions.
- As an example, let us approximate the capital decision function K.
- First, we take two grid points for each state variable, namely, $\{k_1, k_2\}$ and $\{\theta_1, \theta_2\}$, and we combine them to construct two-dimensional grid points, $\{(k_1, \theta_1), (k_1, \theta_2), (k_2, \theta_1), (k_2, \theta_2)\}$.
- Second, we take two basis functions for each state variable, namely, $\{1, k\}$ and $\{1, \theta\}$, and we combine them to construct two-dimensional basis functions $\{1, k, \theta, k\theta\}$.
- Third, we construct a flexible functional form for approximating K,

$$\widehat{K}(k, \theta; b) = b_1 + b_2 k + b_3 \theta + b_4 k\theta. \tag{3}$$

Multidimensional grid points and basis functions (cont.)

- Finally, we identify the four unknown coefficients $(b_1, b_2, b_3, b_4) \equiv b$ such that $K(k, \theta)$ and $\widehat{K}(k, \theta; b)$ coincide exactly in the four grid points constructed.
- That is, we write $\mathcal{B}b = w$, where

$$\mathcal{B} = \begin{bmatrix} 1 & k_1 & \theta_1 & k_1\theta_1 \\ 1 & k_1 & \theta_2 & k_1\theta_2 \\ 1 & k_2 & \theta_1 & k_2\theta_1 \\ 1 & k_2 & \theta_2 & k_2\theta_2 \end{bmatrix}, \qquad b = \begin{bmatrix} b_1 \\ b_2 \\ b_3 \\ b_4 \end{bmatrix}, \qquad w = \begin{bmatrix} K(k_1, \theta_1) \\ K(k_1, \theta_2) \\ K(k_2, \theta_1) \\ K(k_2, \theta_2) \end{bmatrix}$$

- If $\mathcal{B}$ has full rank, then coefficients vector b is uniquely determined by $b = \mathcal{B}^{-1}w$.
- The obtained approximation can be used to interpolate the capital decision function in each point off the grid.

Numerical integration

- For integration, we consider first a simple two-node Gauss-Hermite quadrature method that approximates an integral of a function of a Normally distributed variable $\epsilon \sim \mathcal{N}\left(0, \sigma^2\right)$ with a weighted average of just two values $\epsilon_1 = -\sigma$ and $\epsilon_2 = \sigma$ that happen with probability $\omega_1 = \omega_2 = \frac{1}{2}$, i.e.,

$$\int_{-\infty}^{\infty} G\left(\epsilon\right) w\left(\epsilon\right) d\epsilon \approx G\left(\epsilon_1\right) \omega_1 + G\left(\epsilon_2\right) \omega_2 = \frac{1}{2}\left[G\left(-\sigma\right) + G\left(\sigma\right)\right],$$

where G is a bounded continuous function, and w is a density function of a Normal distribution, i.e.,

$$\begin{aligned} u_1\left(c_m, \ell_m\right) = \frac{1}{2}\beta \{ & u_1\left(c'_{m,\sigma}, \ell'_{m,\sigma}\right)\left(1 - \delta + \theta'_{m,\sigma} f_1\left(k'_m\left(b\right), \ell'_{m,\sigma}\right)\right) \\ & + u_1\left(c'_{m,-\sigma}, \ell'_{m,-\sigma}\right)\left(1 - \delta + \theta'_{m,-\sigma} f_1\left(k'_m\left(b\right), \ell'_{m,-\sigma}\right)\right)\} \end{aligned}$$

- Alternatively, we can use a three-node Gauss-Hermite quadrature method, which uses nodes $\epsilon_1 = 0$, $\epsilon_2 = \sigma\sqrt{\frac{3}{2}}$, $\epsilon_3 = -\sigma\sqrt{\frac{3}{2}}$ and weights $\omega_1 = \frac{2\sqrt{\pi}}{3}$, $\omega_2 = \omega_3 = \frac{\sqrt{\pi}}{6}$.

Numerical integration (cont.)

- Another possibility is to approximate integrals using Monte Carlo integration, e.g., Parameterized Expectation Algorithm (PEA) by den Haan and Marcet (1990).
- We can make J random draws and approximate an integral with a simple average of the draws,

$$\int_{-\infty}^{\infty} G(\epsilon) w(\epsilon) d\epsilon \approx \frac{1}{J} \sum_{j=1}^{J} G(\epsilon_j).$$

Numerical integration (cont.)

Let us compare the above integration methods using an example.

Example

Consider a quadratic function $G(\epsilon) = b_1 + b_2\epsilon + b_3\epsilon^2$, where $\epsilon \sim \mathcal{N}(0, \sigma^2)$.
(i) An exact integral is $\mathcal{I} \equiv \int_{-\infty}^{\infty} (b_1 + b_2\epsilon + b_3\epsilon^2)\, w(\epsilon)\, d\epsilon = b_1 + b_3\sigma^2$;
(ii) A two-node Gauss-Hermite quadrature integration method yields $\mathcal{I} \approx b_1 + b_3\sigma^2$;
(iii) A one-node Gauss-Hermite quadrature integration method yields $\mathcal{I} \approx b_1$;
(iv) A Monte Carlo integration method yields $\mathcal{I} \approx b_1 + b_2\left[\frac{1}{J}\sum_{j=1}^{J}\epsilon_j\right] + b_3\left[\frac{1}{J}\sum_{j=1}^{J}\epsilon_j^2\right]$.
(v) Quasi Monte Carlo methods for integration - the error bounds are provided in Rust (1987).

Numerical integration (cont.)

- Note that the quadrature method with two nodes delivers the exact value of the integral.
- Even with just one node, the quadrature method can deliver accurate integral if G is close to linear (which is often the case in real business cycle models), i.e., $b_3 \approx 0$.
- To assess the accuracy of Monte Carlo integration, let us use $\sigma = 0.01$, which is consistent with the magnitude of fluctuations in real business cycle models.
  – Let us concentrate just on the term $\frac{1}{J}\sum_{j=1}^{J}\epsilon_j$ for which the expected value and standard deviation are $E\left[\frac{1}{J}\sum_{j=1}^{J}\epsilon_j\right] = 0$ and $std\left[\frac{1}{J}\sum_{j=1}^{J}\epsilon_j\right] = \frac{\sigma}{\sqrt{J}}$, respectively.
  – The standard deviation depends on the number of random draws: with one random draw, it is 0.01 and with 1,000,000 draws, it is $\frac{0.01}{\sqrt{1000000}} = 10^{-5}$.

Numerical integration (cont.)

- The last number represents an (expected) error in approximating the integral and restricts the overall accuracy of solutions that can be attained by a solution algorithm using Monte Carlo integration.
- Why is Monte Carlo integration inefficient in this context?
- This is because we compute expectations as do econometricians, who do not know the true density function of the data-generating process and have no choice but to estimate such a function from noisy data using a regression.
- However, when solving an economic model, we do know the process for shocks. Hence, we can construct the "true" density function and we can use such a function to compute integrals very accurately, which is done by the Gauss-Hermite quadrature method.
- This is done in Judd, Maliar and Maliar (2011) who develop generalized stochastic simulation method (GSSA) that attains high accuracy by combining stochastic simulation for constructing the domain and accurate deterministic integration methods.

Optimization methods

- To solve nonlinear equations with respect to the unknown parameters vectors b^k, b^c, b^ℓ.
- This can be done with Newton-style optimization methods; see, e.g., Judd (1992).
- Such methods compute first and second derivatives of an objective function with respect to the unknowns and move in the direction of gradient descent until a solution is found.
- Newton methods are fast and efficient in small problems but become increasingly expensive when the number of unknowns increases.
- In high-dimensional applications, we may have thousands of parameters in approximating functions, and the cost of computing derivatives may be prohibitive.
- In such applications, derivative-free optimization methods are an effective alternative.
- A useful choice is a fixed-point iteration method that finds a root of $x = F(x)$ by constructing a sequence $x^{(i+1)} = F\left(x^{(i)}\right)$.

Optimization methods (cont.)

- We illustrate this method using an example.

Example

Consider an equation $x^3 - x - 1 = 0$. Let us rewrite this equation as $x = (x+1)^{1/3}$ and construct a sequence $x^{(i+1)} = (x^{(i)} + 1)^{1/3}$ starting from $x^{(0)} = 1$. This yields a sequence $x^{(1)} = 1.26$, $x^{(2)} = 1.31$, $x^{(3)} = 1.32$,... which converges to a solution.

- The advantage of fixed-point iteration is that it can iterate in this simple manner on objects of any dimensionality, for example, on a vector of the polynomial coefficients.
- The cost of this procedure does not grow considerably with the number of the polynomial coefficients.

Nonconvergence of fixed-point iteration

- The shortcoming of fixed point iteration is that it does not always converge.

Example

If we wrote the above equation as $x = x^3 - 1$ and implemented fixed-point iteration $x^{(i+1)} = \left(x^{(i)}\right)^3 - 1$, we would obtain a sequence that diverges to $-\infty$ starting from $x^{(0)} = 1$.

Evaluating accuracy of solutions

- Our solution procedure has two stages. In Stage 1, a method attempts to compute a numerical solution to a model.
- Provided that it succeeds, we proceed to Stage 2, in which we subject a candidate solution to a tight accuracy check.
- We specifically construct a set of points $\{k_i, \theta_i\}_{i=1,...,I}$ that covers an area in which we want the solution to be accurate, and we compute unit-free residuals in the model's equations:

$$\begin{aligned}
\mathcal{R}^{BC}(k_i, \theta_i) &= \frac{(1-\delta) k_i + \theta_i f(k_i, \ell_i)}{c_i + k_i'} - 1, \\
\mathcal{R}^{EE}(k_i, \theta_i) &= \beta E\left\{ \frac{u_1(c_i', \ell_i')}{u_1(c_i, \ell_i)} \left[1 - \delta + \theta_i' f_1(k_i', \ell_i')\right] \right\} - 1, \\
\mathcal{R}^{MUL}(k_i, \theta_i) &= \frac{u_1(c_i, \ell_i) \theta_i f_2(k_i, \ell_i)}{u_2(c_i, \ell_i)} - 1,
\end{aligned}$$

where $\mathcal{R}^{BC}$, $\mathcal{R}^{EE}$ and $\mathcal{R}^{MUL}$ are the residuals in the budget constraint, Euler equation, and FOC for the marginal utility of leisure.

Evaluating accuracy of solutions

- In the exact solution, residuals are zero, so we judge the quality of approximation by how far these residuals are away from zero.
- We should never evaluate residuals on points used for computing a solution in Stage 1 (in particular, for some methods the residuals in the grid points are zeros by construction) but we do so on a new set of points constructed for Stage 2.
- We consider two alternative sets of I points:
  - a fixed rectangular grid
  - a stochastic simulation.
- We report two accuracy measures, namely, the average and maximum absolute residuals across both the optimality conditions and I test points in $\log 10$ units, for example, $\mathcal{R}^{BC}(k_i, \theta_i) = -2$ means that a residual of $10^{-2} = 1\%$, and $\mathcal{R}^{BC}(k_i, \theta_i) = -4.5$ means $10^{-4.5} = 0.00316\%$.
- Judd et al (2017) show how to construct bounds on approximation errors from the residuals.

Economically significant accuracy

- If either a solution method fails to converge in Stage 1 or the quality of a candidate solution in Stage 2 is economically inacceptable, we modify the algorithm's design.
- For example, change the number and placement of grid points, approximating functions, integration method, fitting method, etc.
- We repeat the computations until a satisfactory solution is produced.
- We do not want the solution to be more accurate than needed - this is costly!

Challenges of economic dynamics

Curse of dimensionality

- It turned out that not only analytical but also numerical solutions can be expensive (or infeasible) to obtain for many models of interest.
- Curse of dimensionality: the complexity of a problem grows exponentially with the size:
  - assume that there are N capital stocks;
  - take 10 grid points for each capital stock;
  - we obtain 10^N grid points for N capital stocks, e.g., $N = 10 \Rightarrow 10^{10}$ grid points!
- (a) Tensor product grids $\Longrightarrow$ Curse of dimensionality!
  (b) Product quadrature integration $\Longrightarrow$ Curse of dimensionality!
  (c) Newton's solver (Jacobian, Hessian) $\Longrightarrow$ Curse of dimensionality!
- Economic models can easily become intractable even with supercomputers.

For large problems, we need state-of-art numerical methods, as well as powerful hardware and software.

Challenges of economic dynamics

Multiple solutions, numerical instability and non-convergence

- Optimal control problems can be formulated as dynamic programming (DP) problems and described by Bellman equation. For these problems, we can show the existence of solutions and convergence.
- Equilibrium problems do not always admit a DP formulation. Such problems lead to systems of non-linear equations that may have multiple solutions.
- The convergence is not guaranteed for equilibrium problems.
- Furthermore, inverse problems implied by some models can be ill conditioned.

We need numerical techniques that can compute multiple equilibria and to select a particular equilibrium of interest - little progress so far!

Challenges of economic dynamics

Estimating parameters in economic models with the data:

- Nested fixed point analysis requires us to compute a solution to an economic model is computed within an estimation procedure, for example,
  - Pakes and McGuire (2001) used stochastic simulation for the estimation of IO models
  - Smets and Wouters (2003, 2007) used perturbation solutions for the estimation of a new Keynesian model
  - Fernández-Villaverde and Rubio-Ramírez (2007), and Winschel and Krätzig (2010) estimate parameters in non-linear macroeconomic models
- In an estimation procedure, we may need to solve a model under different parameters values 50,000 times or so.

We need very fast and efficient numerical algorithms, as well as parallel computing (multiple cores, GPU computing and supercomputers) - very challenging applications!

Part 1: Solution Methods for State-Dependent Models

Lilia Maliar and Serguei Maliar

CEF 2017 Workshop

Envelope Condition Method

Envelope condition method (ECM)

Conventional DP approaches are expensive (rely on numerical maximization and solvers).

How one can make DP approaches more tractable?

1. Carroll (EL, 2005): Endogenous Grid Method (EGM)
2. Maliar and Maliar (EL, 2013): Envelope Condition Method (ECM)
   - *ECM uses* **"forward"** *recursion and differs from conventional Bellman operator*
   - *constructs policy functions using* **envelope condition** *(EC) instead of* **first-order condition** *(FOC)*
   - *solves for* **derivatives of value function** *instead of (in addition to)* **value function** *itself*

For growth model, ECM can solve Bellman equation by using only direct calculation - *no need of numerical solver or maximization!*

Results:

1. ECM for default risk models:
   for a version of Arellano's (2008) model, ECM is about 50x time faster than conventional VFI!
2. ECM for large-scale problems:
   ECM can solves a multi-country models with at least up to 20 state variables and compete in accuracy and speed with the state-of-the-art Euler equation methods
3. Convergence theorems for ECM:
   Our formal results show that, unfortunately, ECM is not necessarily a contraction mapping, unlike conventional Bellman operator.

Model of default risk of Arellano (2008)

A country solves

$$\max_{\{B_{t+1},c_t\}_{t=0,\ldots,\infty}} E_0 \sum_{t=0}^{\infty} \beta^t u(c_t)$$

$$\text{s.t. } c_t = y_t + B_t - q(B_{t+1}, y_t) B_{t+1}$$

$$\log(y_t) = \rho \log(y_{t-1}) + \varepsilon_t^y$$

where $E[\varepsilon^y] = 0$, $E\left[(\varepsilon^y)^2\right] = \eta_y^2$ and (B_0, y_0) is given;
c_t, y_t and B_t are consumption, capital and bonds, respectively;
$q(B_{t+1}, y_t)$ is price of your bonds depending on quantity B_{t+1} & state y_t.

- You may default by setting at $B_t = 0$ ($B_t < 0$ if you are a borrower) $\Rightarrow$ you will be punished.
- $q(\underset{(-)}{B_{t+1}}, \underset{(+)}{y_t})$ increases with the probability of default (investors compute your probability of default in all states (B_t, y_t) conditional on realization y_{t+1}).

How is the model of default risk solved?

For this presentation, consider Arellano's model without default

$$\max_{\{B_{t+1},c_t\}_{t=0,\ldots,\infty}} E_0 \sum_{t=0}^{\infty} \beta^t u(c_t)$$
$$\text{s.t. } c_t = y_t + B_t - qB_{t+1}$$
$$\log(y_t) = \rho \log(y_{t-1}) + \varepsilon_t^y$$

where $E[\varepsilon^y] = 0$, $E\left[(\varepsilon^y)^2\right] = \eta_y^2$ and $q = \frac{1}{1+r}$ is determined by the world interest rate on borrowing.

Conventional DP approaches

Conventional VFI on Bellman equation iterates backward:

$$V(B,y) = \max_{c,B'} \left\{ u(c) + \beta E\left[V(B',y')\right] \right\}$$

$$\text{s.t.} \quad c = y + B - qB'$$

Take a grid of points for (B,y), assume some (FUTURE!) V and find maximum of

$$\max_{B'} \left\{ u(y + B - qB') + \beta E\left[V(B',y')\right] \right\}$$

- Discretization of state space: *discretize state space* (B,y) *into a large number of points.*
- Parametric dynamic programming: *approximate* V *with a parametric function.*

Conventional DP approaches (cont.)

For example, assume a Cobb-Douglas utility function and polynomial approximation of V

$$\max_{B'} \left\{ \frac{(y + B - qB')^{1-\gamma}}{1-\gamma} + \beta E \left[b_0 + b_1 B' + b_2 y' + b_3 \left(B'\right)^2 + ... + b_k \left(B'\right)^d \right] \right\}$$

Solving a maximization problem in each grid point (on multiple iterations) is expensive!

Conventional DP approaches (cont.)

If V is differentiable, instead of a maximization problem, we can find a solution to FOCs

$$\max_{B'} \left\{ u(y + B - qB') + \beta E \left[V(B', y') \right] \right\}$$

Find the derivative and set it to zero

$$u'(y + B - qB')q = \beta E \left[V_1(B', y') \right]$$

We need to find B' that solves this equation

$$(y + B - qB')^{-\gamma} q + \beta E \left[b_0 + b_1 B' + b_2 y' + b_3 \left(B'\right)^2 + ... + b_k \left(B'\right)^d \right] = 0$$

This nonlinear equation that must be solved w.r.t B' in each grid point (B, y)

Solving a non-linear equation in each grid point (on multiple iterations) is also expensive!

Conventional DP approaches (cont).

- Conventional DP approaches are tractable only for relatively simple problems.
- What alternatives are available to conventional DP approaches?

Endogenous grid method (EGM) of Carroll (2005)

- Consider again the FOC

$$(y + B - qB')^{-\gamma} q + \beta E \left[b_0 + b_1 B' + b_2 y' + b_3 \left(B'\right)^2 + ... + b_k \left(B'\right)^d \right] = 0$$

- Note that given (B, y), it is difficult to solve for B'
- But given (B', y), we can solve for B explicitly!

$$B = \left(-\frac{\beta}{q} E \left[b_0 + b_1 B' + b_2 y' + b_3 \left(B'\right)^2 + ... + b_k \left(B'\right)^d \right] \right)^{-\frac{1}{\gamma}} + qB' - y$$

- Instead of $B'(B, y)$, we characterize the solution by inverse $B(B', y)$
- That is, we construct a grid for (B', y) (this gives the name "endogenous grid" to Carroll's method) and solve for $B(B', y)$

Envelope condition method of Maliar and Maliar (2013)

- Instead of iteration on FOC using future V, iterate on envelope condition assuming (CURRENT!) V:

$$V(B,y) = \max_{c,B'} \left\{ u(c) + \beta E\left[V(B',y') \right] \right\}$$

$$\text{s.t.} \quad c = y + B - qB'$$

- Envelope condition $V_1(B,y) = u'(c)$
- For example, if $u(c) = \frac{c^{1-\gamma}}{1-\gamma}$, we have

$$\begin{aligned} \text{Assume } V \quad : \quad & V(B,y) \Rightarrow V_1(B,y) = c^{-\gamma} \\ \Rightarrow \quad & c = V_1(B,y)^{-1/\gamma} \text{ and } B' = \frac{1}{q}(c - y - B) \\ \Rightarrow \quad & V(B,y) = \max_{c,B'} \left\{ u(c) + \beta E\left[V(B',y') \right] \right\} \end{aligned}$$

Note: We find everything analytically! We avoid a numerical solver and numerical optimization.

Envelope condition method

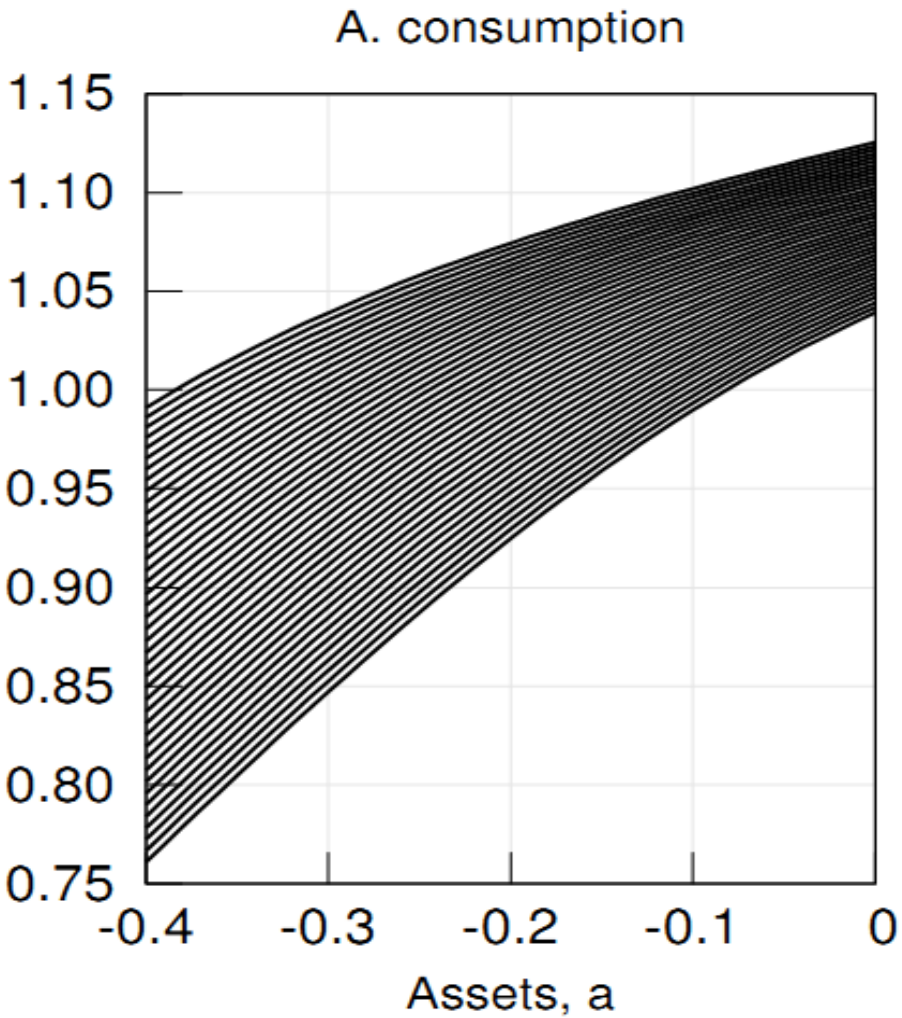
- Solving for value function V (ECM-VF) versus its derivative V_1 (ECM-DVF)
- For example, if $u(c) = \frac{c^{1-\gamma}}{1-\gamma}$, we have

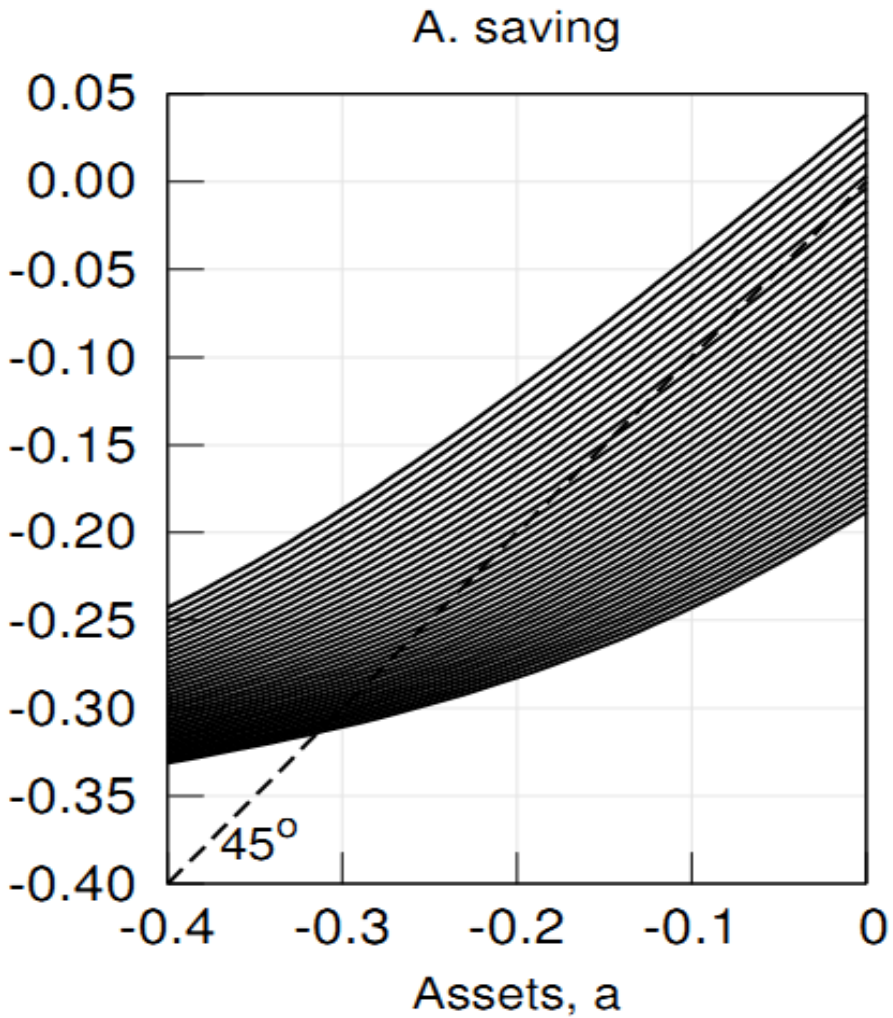
$$\begin{aligned} \text{Assume } V_1(\text{not } V) \quad &: \quad V_1(B,y) = c^{-\gamma} \\ &\Rightarrow \quad c = V_1(B,y)^{-1/\gamma} \text{ and } B' = \frac{1}{q}(c - y - B) \\ &\Rightarrow \quad u(c) = \beta E\left[V_1(B',y')\right] = V_1(B,y) \end{aligned}$$

- Why can this be a good idea?
- Assume a second-degree polynomial function for $V(B,y) \approx b_0 + b_1 B + b_2 y + b_3 B^2 + b_4 By + b_5 y^2$.
- Then, $V_1(B,y) \approx b_1 + 2b_3 B + b_4 y$ is a first-degree polynomial.

We "lose" one polynomial degree when differentiating V to compute V_1 which reduces accuracy.

Numerical examples





$\epsilon = 10^{-5}$: 11.5sec(s) for ECM, 24.3sec(s) for EGM, 510.8 sec(s) for VFI

Conclusion

- Conventional DP approaches are expensive and intractable even for moderately large problems
- Carroll (2005) introduced a far more efficient endogenous grid method (EGM)
- We introduced a competing envelope condition method (ECM)
- In the studied examples, EGM and ECM have similar performance
- In other applications, one method can have advantage over the other
- We show that ECM can be used to solve large problems and has accuracy and speed comparable to state-of-the-art Euler equation methods

Computer codes

"*ECM_and_EGM_MM_2013.zip*"

- *Envelope condition and endogenous grid method for growth model with valued leisure.*

"*7_methods_for_growth_model_AMMT_2016.zip*". *Comparison:*

- *conventional value and policy iteration*
- *envelope condition value and policy iteration*
- *envelope condition method iterating on derivative of value function*
- *endogenous grid method*
- *conventional Euler equation method.*

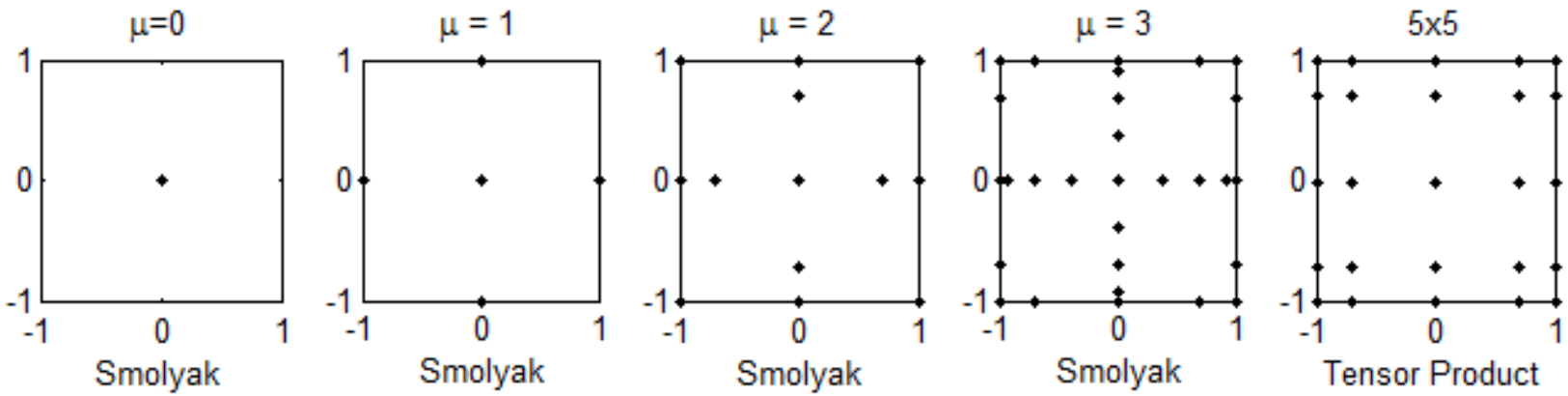
Smolyak Method

Introduction

- In a seminal work, Sergey Smolyak (1963) introduced **a sparse grid technique** for representing, interpolating and integrating multidimensional functions.
- The Smolyak technique builds on non-product rules and does not suffer from the curse of dimensionality (for smooth functions).
- **Idea of the Smolyak method:**
  - Not all tensor product terms are equally important for the quality of approximation.
  - Low-order terms are more important than high-order terms (this is like Taylor series).
  - The Smolyak technique orders all tensor-product elements by their potential importance and selects a relatively small number of the most important elements.
- A parameter, called a *level of approximation* (like the order of Taylor expansion), controls how many tensor-product elements are included into the Smolyak grid.
- By increasing the level of approximation μ, we add new elements and

Introduction

Examples of Smolyak grids under the approximation levels $\mu = 0, 1, 2, 3$ for the two-dimensional case.



Tensor-product grid with 5^d points vs. Smolyak grid

d	Tensor-product grid with 5^d points	Smolyak grid		
		$\mu = 1$	$\mu = 2$	$\mu = 3$
1	5	3	5	9
2	25	5	13	29
10	9,765,625	21	221	1581
20	95,367,431,640,625	41	841	11,561

- The number of points in the Smolyak grids grows polynomially with dimensionality d.
  - for $\mu = 1$, we have $1 + 2d$ elements (grows linearly);
  - for $\mu = 2$, we have $1 + 4d + (4d\,(d-1))/2$ elements (grows quadratically).
- A relatively small number of Smolyak grid points contrasts sharply with a huge number of tensor-product grid points.

Introduction

Our results: toward more efficient Smolyak interpolation

1. *Efficient construction of Smolyak polynomials.*
   - The nested-set construction of Smolyak polynomials is inefficient: it first creates a long list of repeated elements and then eliminates the repeated elements from the list.
   - We construct Smolyak polynomials using disjoint sets $\Longrightarrow$ **we avoid costly repetitions of elements.**
2. *A Lagrange-style technique for computing coefficients.*
   - The conventional Smolyak method computes polynomial coefficients using a formula with a large number of nested loops.
   - We compute the coefficients by precomputing a solution to the inverse problem $\Longrightarrow$ **a simple, general and cheap technique.**
3. *Anisotropic grid: different approximation levels for different variables.*
   - The conventional Smolyak method is symmetric (with the same number of grids and polynomial functions for all variables).
   - We develop an anisotropic version of the Smolyak method $\Longrightarrow$ **we can vary the quality of approximation across variables.**

Introduction

Our results: adapting Smolyak method to economic applications

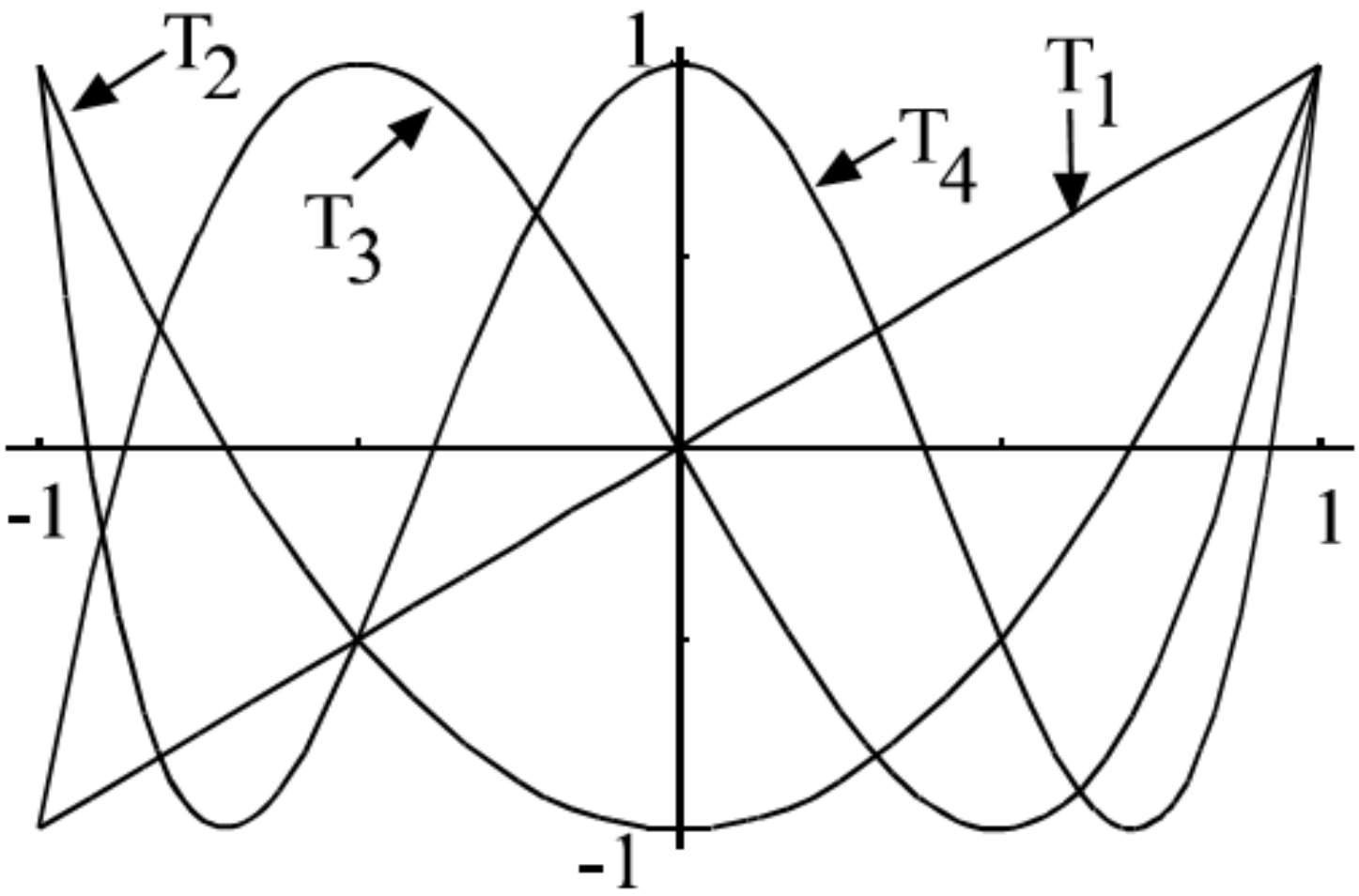
4. *Adaptive domain.*
   - The conventional Smolyak method constructs grid points in a normalized multidimensional hypercube $[-1, 1]^d$.
   - **We show how to effectively adapt the Smolyak hypercube domain to the high-probability set of the given model**.
5. *Iterative procedure*.
   - The conventional Smolyak method of Krueger and Kubler (2004) and Malin et al. (2011) uses *time iteration*: given functional forms for future variables, they solve for current variables using a numerical solver.
   - **We replace time-iteration with a fixed-point iteration which is cheap and simple to implement**. The fixed-point iteration involves just straightforward computations and avoids the need for a numerical solver under time iteration (this modification, although minor in substance, is still important for reducing the cost)**.**

Conventional Smolyak grid using nested sets

Unidimensional nested sets

- Construct sets of points $i = 1, 2, ...$ that satisfy two conditions:
  - *Condition 1.* Sets $i = 1, 2, ...$ have $m(i) = 2^{i-1} + 1$ points for $i \geq 2$ and $m(1) \equiv 1$.
  - *Condition 2.* Each subsequent set $i + 1$ contains all points of the previous set i. Such sets are called *nested*.
- There are many ways to construct the sets of points, satisfying Conditions 1 and 2.
- As an example, let us consider grid points $\left\{-1, \frac{-1}{\sqrt{2}}, 0, \frac{1}{\sqrt{2}}, 1\right\}$ in the interval $[-1, 1]$ and create 3 nested sets of points:
  $i = 1 : S_1 = \{0\}$;
  $i = 2 : S_2 = \{-1, 0, 1\}$;
  $i = 3 : S_3 = \left\{-1, \frac{-1}{\sqrt{2}}, 0, \frac{1}{\sqrt{2}}, 1\right\}$.

Extrema of Chebyshev polynomials



Conventional Smolyak grid using nested sets

Tensor products of unidimensional nested sets

	$S_{i_1} \setminus S_{i_2}$	$i_2 = 1$	$i_2 = 2$	$i_2 = 3$
		0	$-1, 0, 1$	$-1, \frac{-1}{\sqrt{2}}, 0, \frac{1}{\sqrt{2}}, 1$
$i_1 = 1$	0	$(0,0)$	$(0,-1), (0,0), (0,1)$	$(0,-1), (0,\frac{-1}{\sqrt{2}}), (0,0), (0,\frac{1}{\sqrt{2}}), (0,1)$
$i_1 = 2$	-1 0 1	$(-1,0)$ $(0,0)$ $(1,0)$	$(-1,-1), (-1,0), (-1,1)$ $(0,-1), (0,0), (0,1)$ $(1,-1), (1,0,), (1,1)$	$(-1,-1), (-1,\frac{-1}{\sqrt{2}}), (-1,0), (-1,\frac{1}{\sqrt{2}}), (-1,1)$ $(0,-1), (0,\frac{-1}{\sqrt{2}}), (0,0), (0,\frac{1}{\sqrt{2}}), (1,0)$ $(1,-1), (1,\frac{-1}{\sqrt{2}}), (1,0), \left(1,\frac{1}{\sqrt{2}}\right) (1,1)$
$i_1 = 3$	-1 $\frac{-1}{\sqrt{2}}$ 0 $\frac{1}{\sqrt{2}}$ 1	$(-1,0)$ $\left(\frac{-1}{\sqrt{2}},0\right)$ $(0,0)$ $\left(\frac{1}{\sqrt{2}},0\right)$ $(1,0)$	$(-1,-1), (-1,0), (-1,1)$ $(\frac{-1}{\sqrt{2}},-1), (\frac{-1}{\sqrt{2}},0), (\frac{-1}{\sqrt{2}},1)$ $(0,-1), (0,0), (1,0)$ $(\frac{1}{\sqrt{2}},-1), (\frac{1}{\sqrt{2}},0), (\frac{1}{\sqrt{2}},1)$ $(1,-1), (1,0), (1,1)$	$(-1,-1), (-1,\frac{-1}{\sqrt{2}}), (-1,0), (-1,\frac{1}{\sqrt{2}}), (-1,1)$ $(\frac{-1}{\sqrt{2}},-1), (\frac{-1}{\sqrt{2}},\frac{-1}{\sqrt{2}}), (\frac{-1}{\sqrt{2}},0), (\frac{-1}{\sqrt{2}},\frac{1}{\sqrt{2}}), (\frac{-1}{\sqrt{2}},1)$ $(0,-1), (0,\frac{-1}{\sqrt{2}}), (0,0), (0,\frac{1}{\sqrt{2}}), (1,0)$ $(\frac{1}{\sqrt{2}},-1), (\frac{1}{\sqrt{2}},\frac{-1}{\sqrt{2}}), (\frac{1}{\sqrt{2}},0), (\frac{1}{\sqrt{2}},\frac{-1}{\sqrt{2}}), (\frac{1}{\sqrt{2}},1)$ $(1,-1), (1,\frac{-1}{\sqrt{2}}), (1,0), \left(1,\frac{1}{\sqrt{2}}\right) (1,1)$

Conventional Smolyak grid using nested sets

Smolyak sparse grid

- Smolyak (1963) rule used to select tensor products:

$$d \leq i_1 + i_2 \leq d + \mu,$$

  where $\mu \in \{0, 1, 2, ...\}$ is the approximation level, and d is the dimensionality (in our case, $d = 2$).
- In terms of the above table, the sum of indices of a column i_1 and a raw i_2, must be between d and $d + \mu$.
- Let $\mathcal{H}^{d,\mu}$ denote the Smolyak grid for a problem with dimensionality d and approximation level μ.

Conventional Smolyak grid using nested sets

Smolyak sparse grid: $d = 2$.

- If $\mu = 0 \Longrightarrow 2 \leq i_1 + i_2 \leq 2$. The only cell that satisfies this restriction is $i_1 = 1$ and $i_2 = 1 \Longrightarrow$ the Smolyak grid has just one grid point

$$\mathcal{H}^{2,0} = \{(0,0)\}.$$

- If $\mu = 1 \Longrightarrow 2 \leq i_1 + i_2 \leq 3$. The 3 cells that satisfy this restriction: (a) $i_1 = 1$, $i_2 = 1$; (b) $i_1 = 1$, $i_2 = 2$; (c) $i_1 = 2$, $i_2 = 1$, and the corresponding 5 Smolyak grid points are

$$\mathcal{H}^{2,1} = \{(0,0),(-1,0),(1,0),(0,-1),(0,1)\}.$$

- If $\mu = 2 \Longrightarrow 2 \leq i_1 + i_2 \leq 4$. There are 6 cells satisfy this restriction $\Longrightarrow$ 13 Smolyak grid points:

$$\mathcal{H}^{2,2} = \{(-1,1),(0,1),(1,1),(-1,0),(0,0),(1,0),(-1,-1),$$
$$(0,-1),(1,-1),(\frac{-1}{\sqrt{2}},0),(\frac{1}{\sqrt{2}},0),(0,\frac{-1}{\sqrt{2}}),(0,\frac{1}{\sqrt{2}})\}.$$

Conventional Smolyak polynomials using nested sets

Let $\mathcal{P}^{d,\mu}$ denote a Smolyak polynomial function in dimension d, with approximation level μ,

$$\mathcal{P}^{d,\mu}\left(x_1,...,x_d;b\right) = \sum_{\max(d,\mu+1)\leq|i|\leq d+\mu}(-1)^{d+\mu-|i|}\binom{d-1}{d+\mu-|i|}p^{|i|}\left(x_1,...,x_d\right),$$

where $p^{|i|}\left(x_1,...,x_d\right)$ is the sum of $p^{i_1,...,i_d}\left(x_1,...,x_d\right)$ with $i_1+...+i_d=|i|$ defined as

$$p^{i_1,...,i_d}\left(x_1,...,x_d\right)=\sum_{\ell_1=1}^{m(i_1)}...\sum_{\ell_d=1}^{m(i_d)}b_{\ell_1...\ell_d}\psi_{\ell_1}\left(x_1\right)\cdots\psi_{\ell_d}\left(x_d\right),$$

where $m\left(i_1\right),...,m\left(i_d\right)=$ number of basis functions in dimensions $1,...,d$; $m\left(i\right)\equiv 2^{i-1}+1$ for $i\geq 2$ and $m\left(1\right)\equiv 1$; $\psi_{\ell_1}\left(x_1\right),...,\psi_{\ell_d}\left(x_d\right)=$ unidimensional basis functions; $\ell_d=1,...,m\left(i_d\right)$; and $b_{\ell_1...\ell_d}$ are polynomial coefficients.

Inefficiency of conventional Smolyak interpolation

- **Inefficiency:** First, we create a list of tensor products with many repeated elements and then, we eliminate the repetitions.
- **Repetitions of grid points.**
  - $\mathcal{H}^{2,1}$: $(0,0)$ is listed 3 times $\Longrightarrow$ must eliminate 2 grid points out of 7.
  - $\mathcal{H}^{2,2}$: must eliminate 12 repeated points out of 25 points.
  - *But grid points must be constructed just once (fixed cost), so repetitions are not so important for the cost.*
- **Repetitions of basis functions**.
  - $\mathcal{P}^{2,1}$ lists 7 basis functions from sets $\{1\}$, $\{1, \psi_2(x), \psi_3(x)\}$, $\{1, \psi_2(y), \psi_3(y)\}$ and eliminates 2 repeated functions $\{1\}$ by assigning a weight (-1) to $p^{|2|}$.
  - $\mathcal{P}^{2,2}$: must eliminate 12 repeated basis functions out of 25.
  - *Smolyak polynomials must be constructed many times (in every grid point, integration node and time period) and each time we suffer from repetitions.*
- **The number of repetitions increases in μ and d** $\Longrightarrow$ important for high-dimensional applications.

Smolyak method with Lagrange interpolation

We now present an alternative variant of the Smolyak method.

- First, instead of nested sets, we use disjoint sets, which allows us to avoid repetitions.
- Second, we find the coefficients using Lagrange-style interpolation. This technique works for any basis function and not necessarily orthogonal ones. Most of the computations can be done up-front (precomputed).
- Our version of the Smolyak method will be more simple and intuitive and easier to program.

Step 1. Smolyak grid using disjoint sets

Unidimensional grid points using disjoint sets

- We construct the Smolyak grid using disjoint sets.
- We consider grid points $\left\{-1, \frac{-1}{\sqrt{2}}, 0, \frac{1}{\sqrt{2}}, 1\right\}$ in the interval $[-1, 1]$ and create 3 unidimensional sets of elements (grid points), A_1, A_2, A_3, which are disjoint, i.e., $A_i \cap A_j = \{\varnothing\}$ for any i and j.
  $i = 1 : A_1 = \{0\}$;
  $i = 2 : A_2 = \{-1, 1\}$;
  $i = 3 : A_3 = \left\{\frac{-1}{\sqrt{2}}, \frac{1}{\sqrt{2}}\right\}$.

Step 1. Smolyak grid using disjoint sets

Tensor products of unidimensional disjoint sets of points

	$A_{i_1} \setminus A_{i_2}$	$i_2 = 1$	$i_2 = 2$	$i_2 = 3$
		0	$-1, 1$	$\frac{-1}{\sqrt{2}}, \frac{1}{\sqrt{2}}$
$i_1 = 1$	0	$(0,0)$	$(0,-1), (0,1)$	$\left(0, \frac{-1}{\sqrt{2}}\right), \left(0, \frac{1}{\sqrt{2}}\right)$
$i_1 = 2$	-1 1	$(-1,0)$ $(1,0)$	$(-1,-1), (-1,1)$ $(1,-1), (1,1)$	$\left(-1, \frac{-1}{\sqrt{2}}\right), \left(-1, \frac{1}{\sqrt{2}}\right)$ $\left(1, \frac{-1}{\sqrt{2}}\right), \left(1, \frac{1}{\sqrt{2}}\right)$
$i_1 = 3$	$\frac{-1}{\sqrt{2}}$ $\frac{1}{\sqrt{2}}$	$\left(\frac{-1}{\sqrt{2}}, 0\right)$ $\left(\frac{1}{\sqrt{2}}, 0\right)$	$\left(\frac{-1}{\sqrt{2}}, -1\right), \left(\frac{-1}{\sqrt{2}}, 1\right)$ $\left(\frac{1}{\sqrt{2}}, -1\right), \left(\frac{1}{\sqrt{2}}, 1\right)$	$\left(\frac{-1}{\sqrt{2}}, \frac{-1}{\sqrt{2}}\right), \left(\frac{-1}{\sqrt{2}}, \frac{1}{\sqrt{2}}\right)$ $\left(\frac{1}{\sqrt{2}}, \frac{-1}{\sqrt{2}}\right), \left(\frac{1}{\sqrt{2}}, \frac{1}{\sqrt{2}}\right)$

We select elements that belong to the cells with the sum of indices of a column and a row, $i_1 + i_2$, between d and $d + \mu$. This leads to the same Smolyak grids as before. However, in our case, no grid points are repeated.

Smolyak grid using disjoint sets

Smolyak sparse grid

- We use the same Smolyak rule for constructing multidimensional grid points

$$d \leq i_1 + i_2 \leq d + \mu$$

  That is, we select elements that belong to the cells in the above table for which the sum of indices of a column and a row, $i_1 + i_2$, is between d and $d + \mu$.
- This leads to the same Smolyak grids $\mathcal{H}^{2,0}$, $\mathcal{H}^{2,1}$ and $\mathcal{H}^{2,2}$ as under the construction built on nested sets. However, in our case, no grid points are repeated.

Step 2. Smolyak polynomials using disjoint sets

Disjoint sets of basis functions

The same construction as the one we used for constructing the grid points.

$i = 1 : A_1 = \{1\};$
$i = 2 : A_2 = \{\psi_2(x), \psi_3(x)\};$
$i = 3 : A_3 = \{\psi_4(x), \psi_5(x)\}.$

Step 2. Smolyak polynomials using disjoint sets

Tensor products of unidimensional disjoint sets of basis functions

	$A_{i_1} \setminus A_{i_2}$	$i_2 = 1$	$i_2 = 2$	$i_2 = 3$
		1	$\psi_2(y), \psi_3(y)$	$\psi_4(y), \psi_5(y)$
$i_1 = 1$	1	1	$\psi_2(y), \psi_3(y)$	$\psi_4(y), \psi_5(y)$
$i_1 = 2$	$\psi_2(x)$ $\psi_3(x)$	$\psi_2(x)$ $\psi_3(x)$	$\psi_2(x)\psi_2(y), \psi_2(x)\psi_3(y)$ $\psi_3(x)\psi_2(y), \psi_3(x)\psi_3(y)$	$\psi_2(x)\psi_4(y), \psi_2(x)\psi_5(y)$ $\psi_3(x)\psi_4(y), \psi_3(x)\psi_5(y)$
$i_1 = 3$	$\psi_4(x)$ $\psi_5(x)$	$\psi_4(x)$ $\psi_5(x)$	$\psi_4(x)\psi_2(y), \psi_4(x)\psi_3(y)$ $\psi_5(x)\psi_2(y), \psi_5(x)\psi_3(y)$	$\psi_4(x)\psi_4(y), \psi_4(x)\psi_5(y)$ $\psi_5(x)\psi_4(y), \psi_5(x)\psi_5(y)$

For example, for $\mu = 1$, we get

$$\mathcal{P}^{2,1}(x, y; b) = b_{11} + b_{21}\psi_2(x) + b_{31}\psi_3(x) + b_{12}\psi_2(y) + b_{13}\psi_3(y).$$

Step 3. Lagrange-style interpolation for finding coefficients

- Simply find the coefficients so that a polynomial with M basis functions passes through M given grid points.
- Let $f : [-1,1]^d \longrightarrow \mathbb{R}$ be a smooth function.
- Let $\mathcal{P}(\cdot; b)$ be a polynomial function, $\mathcal{P}(x; b) = \sum_{n=1}^{M} b_n \Psi_n(x)$, where $\Psi_n : [-1,1]^d \longrightarrow \mathbb{R}$ is a d-dimensional basis function; $b \equiv (b_1, ..., b_M)$ is a coefficient vector.
- We construct a set of M grid points $\{x_1, ..., x_M\}$ within $[-1,1]^d$, and we compute b so that the true function, f, and its approximation, $\mathcal{P}(\cdot; b)$ coincide in all grid points:

$$\begin{bmatrix} f(x_1) \\ \cdots \\ f(x_M) \end{bmatrix} = \begin{bmatrix} \widehat{f}(x_1; b) \\ \cdots \\ \widehat{f}(x_M; b) \end{bmatrix} = \overbrace{\begin{bmatrix} \Psi_1(x_1) & \cdots & \Psi_M(x_1) \\ \cdots & \ddots & \cdots \\ \Psi_1(x_M) & \cdots & \Psi_M(x_M) \end{bmatrix}}^{\equiv \mathcal{B}} \cdot \begin{bmatrix} b_1 \\ \cdots \\ b_M \end{bmatrix}$$

Lagrange-style interpolation

- Provided that the matrix of basis functions $\mathcal{B}$ has full rank, we have a system of M linear equations with M unknowns that admits a unique solution for b

$$\begin{bmatrix} b_1 \\ \cdots \\ b_M \end{bmatrix} = \begin{bmatrix} \Psi_1(x_1) & \cdots & \Psi_M(x_1) \\ \cdots & \ddots & \cdots \\ \Psi_1(x_M) & \cdots & \Psi_M(x_M) \end{bmatrix}^{-1} \begin{bmatrix} f(x_1) \\ \cdots \\ f(x_M) \end{bmatrix}.$$

  By construction, approximation $\mathcal{P}(\cdot; b)$ coincides with true function f in all grid points, i.e., $\widehat{f}(x_n; b) = f(x_n)$ for all $x_n \in \{x_1, ..., x_M\}$.
- For orthogonal basis functions, matrix $\mathcal{B}$ is well-conditioned.

Lagrange-style interpolation

Example: $d = 2$ and $\mu = 1$.

Just compute 5 coefficients in Smolyak polynomial:
$P^{2,1}(x, y; b) = b_{11} + b_{21}x + b_{31}\left(2x^2 - 1\right) + b_{12}y + b_{13}\left(2y^2 - 1\right)$ to match function f in 5 Smolyak grid points $\{(0,0), (-1,0), (1,0), (0,-1), (0,1)\}$

$$\begin{bmatrix} b_{11} \\ b_{21} \\ b_{31} \\ b_{12} \\ b_{13} \end{bmatrix} = \begin{bmatrix} 1 & 0 & -1 & 0 & -1 \\ 1 & -1 & 1 & 0 & -1 \\ 1 & 1 & 1 & 0 & -1 \\ 1 & 0 & -1 & -1 & 1 \\ 1 & 0 & -1 & 1 & 1 \end{bmatrix}^{-1} \cdot \begin{bmatrix} f(0,0) \\ f(-1,0) \\ f(1,0) \\ f(0,-1) \\ f(0,1) \end{bmatrix}$$

$$= \begin{bmatrix} \frac{f(-1,0)+f(1,0)+f(0,-1)+f(0,1)}{4} \\ \frac{-f(-1,0)+f(1,0)}{2} \\ -\frac{f(0,0)}{2} + \frac{f(-1,0)+f(1,0)}{4} \\ \frac{-f(0,-1)+f(0,1)}{2} \\ -\frac{f(0,0)}{2} + \frac{f(0,-1)+f(0,1)}{4} \end{bmatrix}.$$

Anisotropic grid

- The conventional Smolyak method treats all dimensions symmetrically: it uses the same number of grid points and basis functions for all variables.
- In economic applications, it may be of value to give different treatments to different variables.
- *Why?*
  - Decisions functions may have more curvature in some variables than in others.
  - Some variables may have a larger range of values than the others.
  - Some variables may be more important than the others.

Anisotropic grid

- Let μ_i be an approximation level in dimension i.
- Let $\mu = \left(\mu_1, ..., \mu_d\right)$.
- Let $\mu^{\max} = \max\left\{\mu_1, ..., \mu_d\right\}$
- Note that $\mu_j = i_j^{\max} - 1$ where $i_j^{\max}$ is the maximum index of the sets considered for dimension j.
- Smolyak grid is called *asymmetric (anisotropic)* if there is at least one dimension j such that $i_j \neq i_k$ for $\forall k \neq j$.
- $\mathcal{H}^{d,\left(\mu_1,...,\mu_d\right)} \equiv$ a d-dimensional anisotropic Smolyak grid of approximation levels $\mu = \left(\mu_1, ..., \mu_d\right)$.
- $\mathcal{P}^{d,\left(\mu_1,...,\mu_d\right)} \equiv$ the corresponding Smolyak polynomial.

Tensor products of sets of unidimensional elements

	$A_{i_1} \backslash A_{i_2}$	$i_2 = 1$	$i_2 = 2$
		0	$-1, 1$
$i_1 = 1$	0	$(0,0)$	$(0,-1),(0,1)$
$i_1 = 2$	-1 1	$(-1,0)$ $(1,0)$	$(-1,-1),(-1,1)$ $(1,-1),(1,1)$
$i_1 = 3$	$\frac{-1}{\sqrt{2}}$ $\frac{1}{\sqrt{2}}$	$\left(\frac{-1}{\sqrt{2}},0\right)$ $\left(\frac{1}{\sqrt{2}},0\right)$	$\left(\frac{-1}{\sqrt{2}},-1\right),\left(\frac{-1}{\sqrt{2}},1\right)$ $\left(\frac{1}{\sqrt{2}},-1\right),\left(\frac{1}{\sqrt{2}},1\right)$
$i_1 = 4$	$\frac{-\sqrt{2+\sqrt{2}}}{2}$ $\frac{-\sqrt{2-\sqrt{2}}}{2}$ $\frac{\sqrt{2-\sqrt{2}}}{2}$ $\frac{\sqrt{2+\sqrt{2}}}{2}$	$\left(\frac{-\sqrt{2+\sqrt{2}}}{2},0\right)$ $\left(\frac{-\sqrt{2-\sqrt{2}}}{2},0\right)$ $\left(\frac{\sqrt{2-\sqrt{2}}}{2},0\right)$ $\left(\frac{\sqrt{2+\sqrt{2}}}{2},0\right)$	$\left(\frac{-\sqrt{2+\sqrt{2}}}{2},-1\right),\left(\frac{-\sqrt{2+\sqrt{2}}}{2},1\right)$ $\left(\frac{-\sqrt{2-\sqrt{2}}}{2},-1\right),\left(\frac{-\sqrt{2-\sqrt{2}}}{2},1\right)$ $\left(\frac{\sqrt{2-\sqrt{2}}}{2},-1\right),\left(\frac{\sqrt{2-\sqrt{2}}}{2},1\right)$ $\left(\frac{\sqrt{2+\sqrt{2}}}{2},-1\right),\left(\frac{\sqrt{2+\sqrt{2}}}{2},1\right)$

Anysotropic Smolyak sets

- The Smolyak rule: select elements that satisfy

$$d \leq i_1 + i_2 \leq d + \mu^{\max}$$

- If $\mu = (1, 0)$, then $\mu^{\max} = 1$ and $2 \leq i_1 + i_2 \leq 3$. The 3 cells that satisfy this restriction are (a) $i_1 = 1$, $i_2 = 1$; (b) $i_1 = 1$, $i_2 = 2$; (c) $i_1 = 2$, $i_2 = 1$,

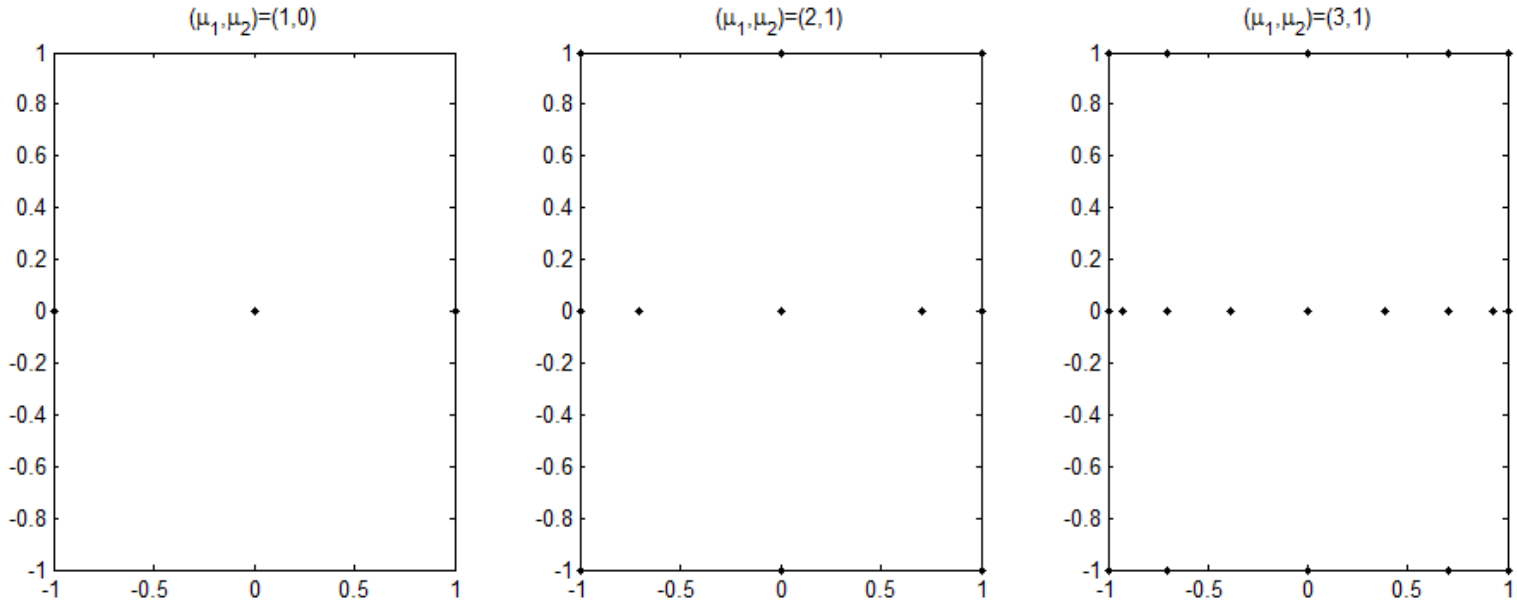
$$\mathcal{H}^{2,\{1,0\}} = \{(0,0), (-1,0), (1,0)\}.$$

- If $\mu = (2, 1)$, then $\mu^{\max} = 2$ and $2 \leq i_1 + i_2 \leq 4$, there are 5 cells that satisfy this restriction (a) $i_1 = 1$, $i_2 = 1$; (b) $i_1 = 1$, $i_2 = 2$; (c) $i_1 = 2$, $i_2 = 1$; (d) $i_1 = 1$, $i_2 = 3$; (e) $i_1 = 2$, $i_2 = 2$; and 11 points:

$$\mathcal{H}^{2,\{2,1\}} = \{(-1,1), (0,1), (1,1), (-1,0), (0,0), (1,0), (-1,-1), (0,-1), (1,-1), (\frac{-1}{\sqrt{2}}, 0), (\frac{1}{\sqrt{2}}, 0)\}.$$

- If $\mu = (3, 1)$, then $\mu^{\max} = 3$ and $2 \leq i_1 + i_2 \leq 5$, there are 19 points.

Anisotropic grids: an illustration



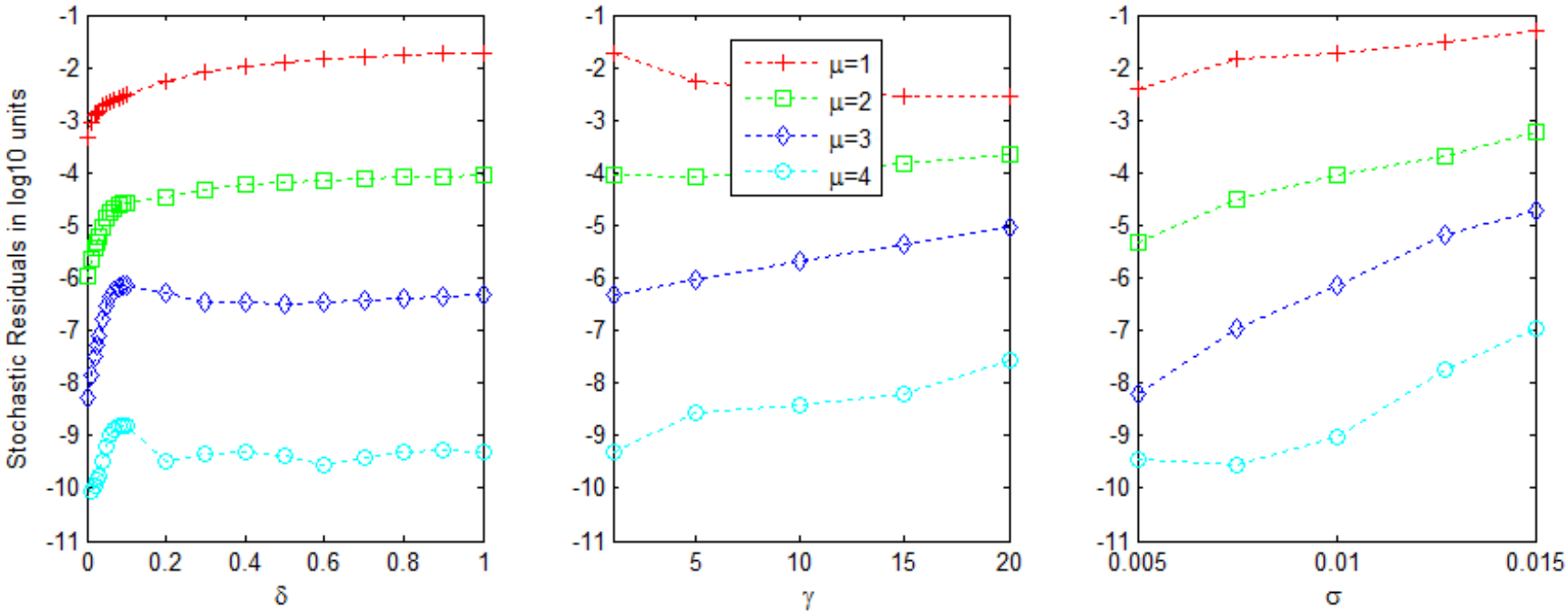
Results for the representative agent model

- CRRA utility function: $u(c) = \frac{c^{1-\gamma}-1}{1-\gamma}$;
- Cobb-Douglas production function: $f(k) = k^{\alpha}$, with $\alpha = 1/3$;
- AR(1) process: $\ln \theta' = \rho \ln \theta + \sigma \varepsilon$, with $\rho = 0.95$
- Discount factor: $\beta = 0.99$.
- Benchmark values: $\delta = 1$, $\gamma = 1$ and $\sigma = 0.01$.
- Then, we consider variations in δ, γ and σ one-by-one holding the remaining parameters at the benchmark values.

$$\begin{aligned} \delta &= \{0,\ 0.1,\ 0.2,\ 0.3,\ 0.4,\ 0.5,\ 0.6,\ 0.7,\ 0.8,\ 0.9,\ 1\}, \\ \gamma &= \{1,\ 5,\ 10,\ 15,\ 20\}, \\ \sigma &= \{0.001,\ 0.005,\ 0.01,\ 0.02,\ 0.03,\ 0.04,\ 0.05\}. \end{aligned}$$

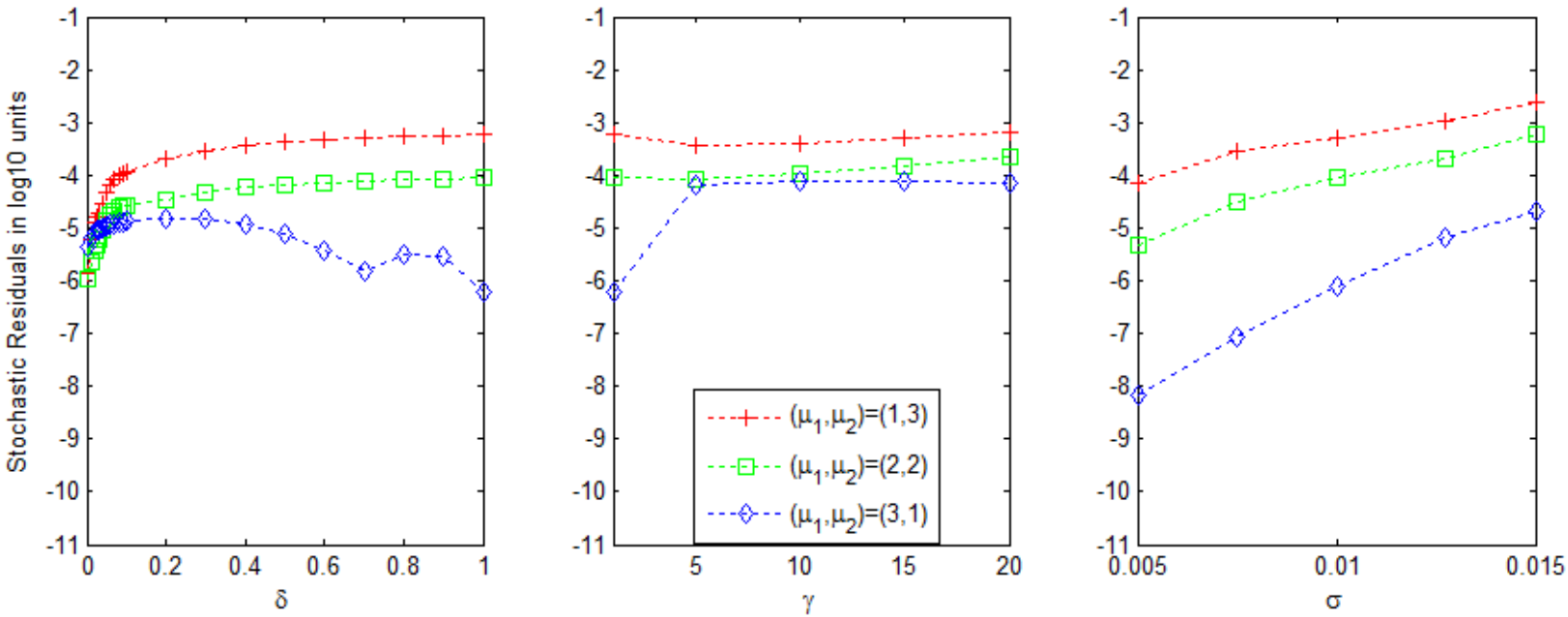
Conventional (isotropic) sparse grids under different approximation levels

- Consider approximation levels $\mu = 1, 2, 3, 4$.

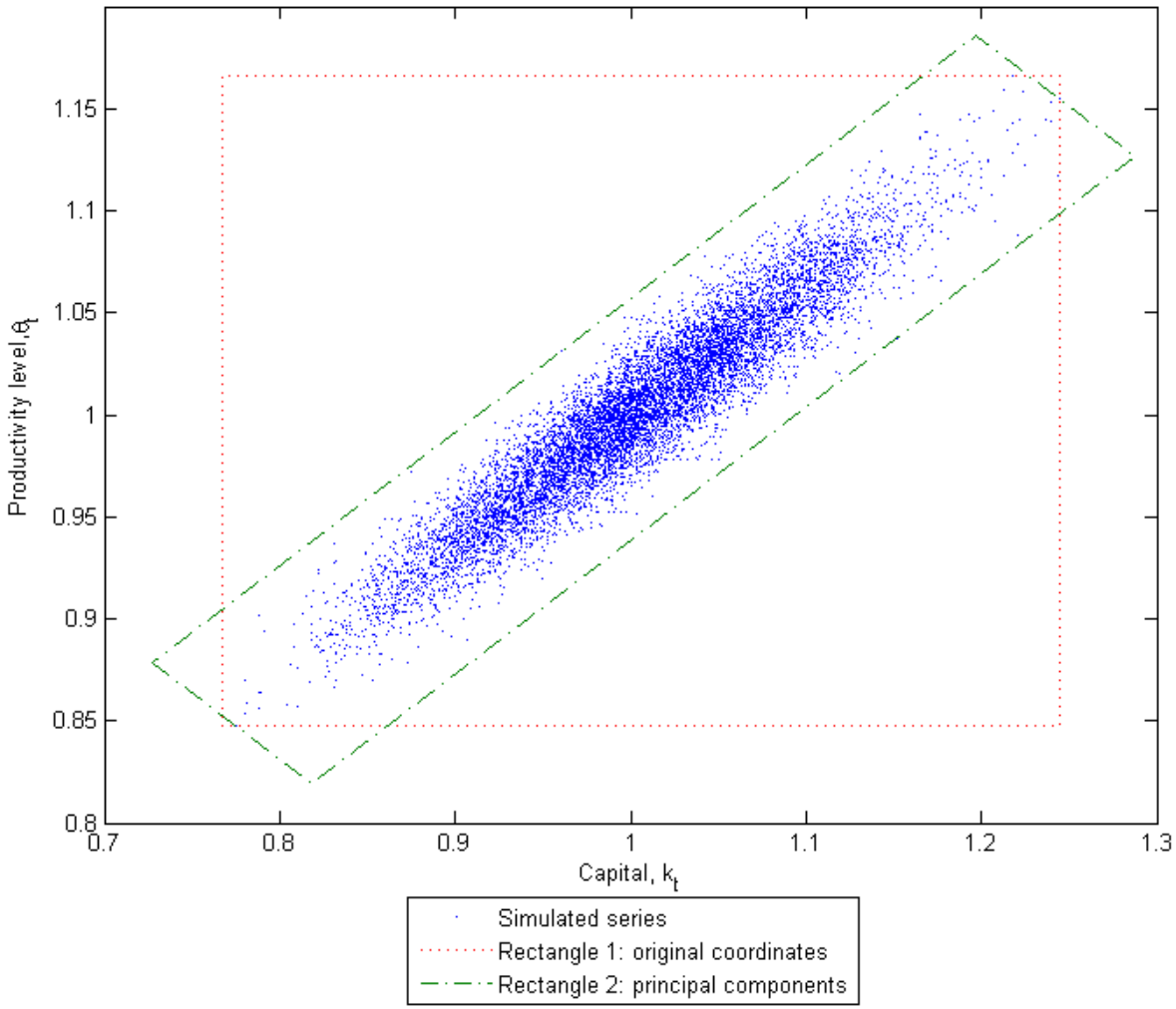


Anisotropic sparse grids

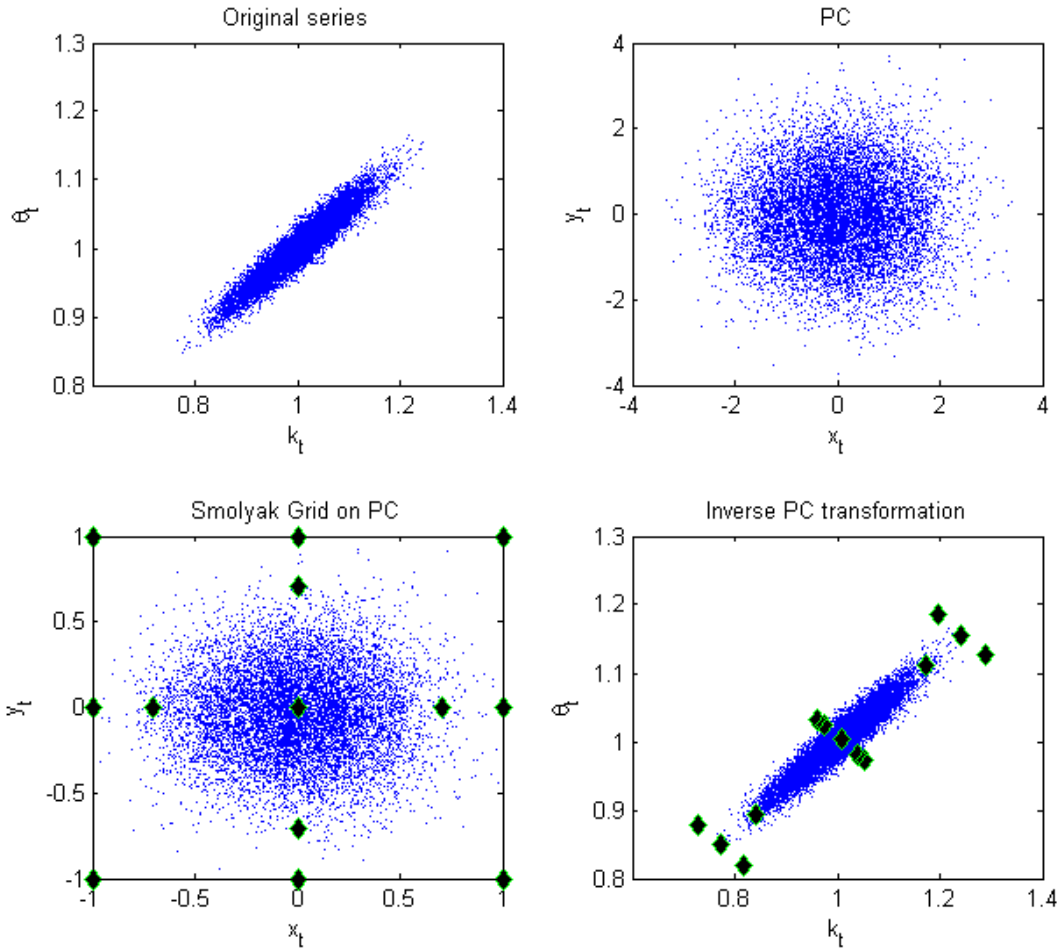
- Consider 2 anisotropic cases: $\mu = (3, 1)$ and $\mu = (1, 3)$.
- There are 9 elements in the first dimension and 3 elements in the second dimension $\Longrightarrow$ 15 grid points and 15 basis functions.



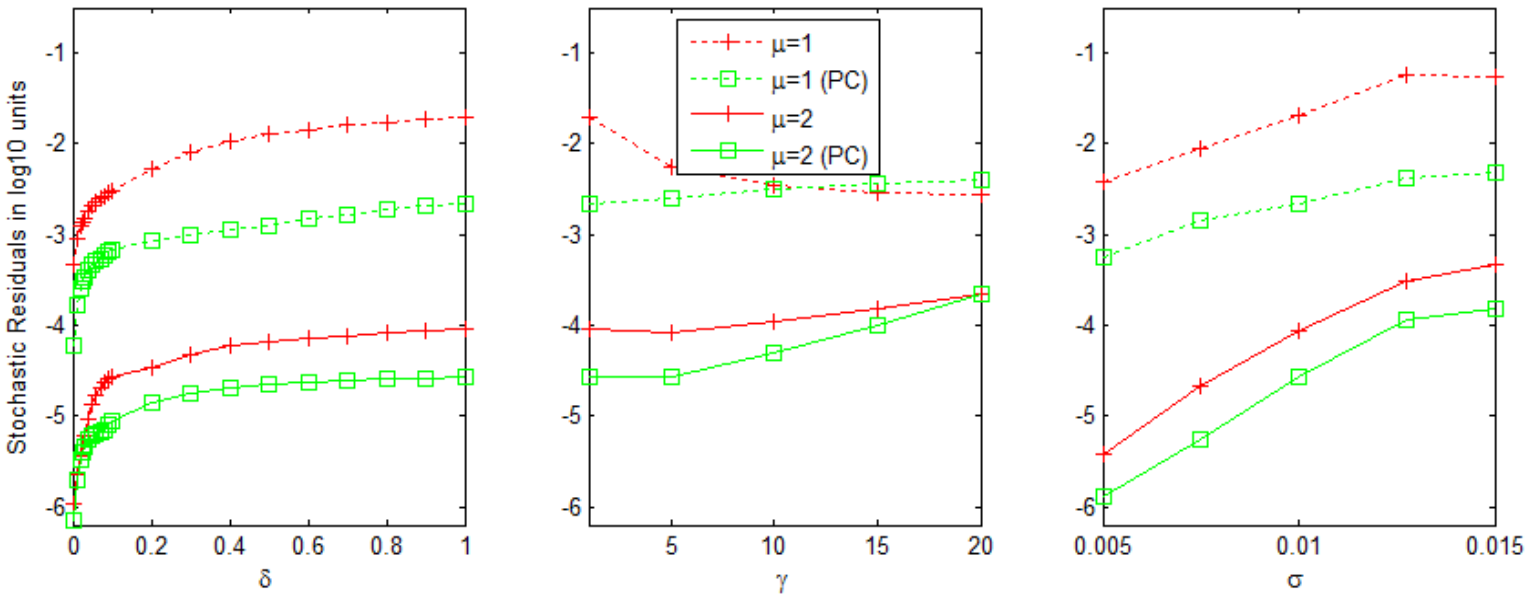
Adaptive Domain: Conventional hypercube vs. a hypercube obtained after the change of variables



Smolyak grid on principal components



Adaptive domain



Multicountry model

$$\max_{\left\{c_t^h, k_{t+1}^h\right\}_{t=0,...,\infty}^{h=1,...,N}} E_0 \sum_{h=1}^{N} \tau^h \left(\sum_{t=0}^{\infty} \beta^t u^h \left(c_t^h \right) \right)$$

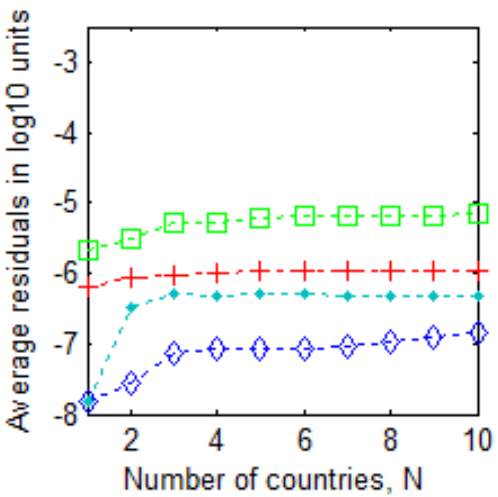
$$\text{s.t.} \quad \sum_{h=1}^{N} c_t^h = \sum_{h=1}^{N} \left[\theta_t^h f^h \left(k_t^h \right) + k_t^h \left(1 - \delta \right) - k_{t+1}^h \right],$$

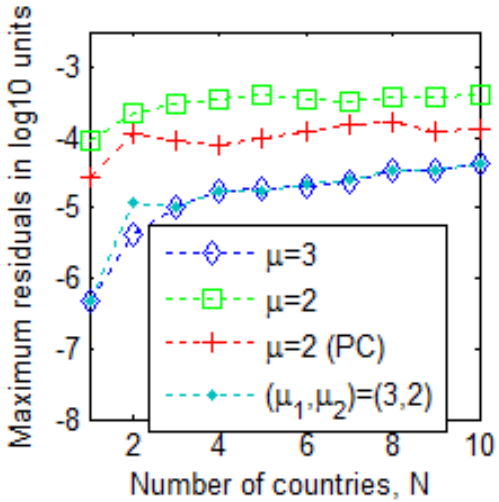
$$\ln \theta_t^h = \rho \ln \theta_{t-1}^h + \epsilon_t^h,$$

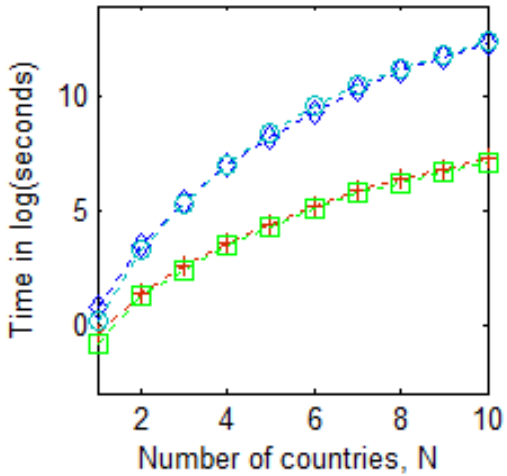
- c_t^h, k_t^h, a_t^h, u^h, f^h and τ^h = consumption, capital, productivity level, utility function, production function, welfare weight of a country h;
- $\epsilon_t^h \equiv \epsilon_t + \varpi_t^h$, $\epsilon_t \sim N\left(0, \sigma\right)$ is a common-for-all-countries shocks, $\varpi_t^h \sim N\left(0, \sigma\right)$ is a country-specific productivity shocks;
- Thus, $\left(\epsilon_t^1, ..., \epsilon_t^N\right)^\top \sim \mathcal{N}\left(0_N, \Sigma\right)$, with $0_N \in \mathbb{R}^N$,

$$\Sigma = \begin{pmatrix} 2\sigma^2 & ... & \sigma^2 \\ ... & ... & ... \\ \sigma^2 & ... & 2\sigma^2 \end{pmatrix} \in \mathbb{R}^{N \times N}.$$

Results for the multicountry model







Conclusion

- The Smolyak method is designed to deal with high-dimensional problems, but its cost still grows rapidly with dimensionality, especially if we target a high quality of approximation.
- We propose a variant of the Smolyak method that has a better performance (lower cost and higher accuracy).
  - We introduce formula for Smolyak polynomials that avoids repetitions and eliminates unnecessary function evaluations.
  - We propose a simple Lagrange-style technique for finding the polynomial coefficients.
  - We develop an anisotropic version of the Smolyak grid that takes into account an asymmetric structure of variables in economic model.
  - As a solution domain, we use a minimum hypercube that encloses the high-probability set of a given economic model.
- The above four improvements are related to Smolyak interpolation. Our last improvement is concerned with an iterative procedure for solving dynamic economic models. We propose to use fixed-point iteration instead of time iteration.

Computer codes

"*Smolyak_Anisotropic_JMMV_2014.zip*"

- *Smolyak method with anisotropic grid for one and multi-country models*
- *Smolyak method operating on a hypercube enclosing the ergodic set*

GSSA, EDS and cluster grids

Hypercube versus ergodic set

- Conventional projection methods including the Smolyak method operate on exogenously given hypercube.
- However, many areas of the hypercube have low probability of occurrence - we might not need to know the solution in low probability areas.
- Stochastic simulation methods construct the solution on a set of simulated points where the solution "lives".
- This can saves on cost a low!

Illustrative example: a representative-agent model

The representative-agent neoclassical stochastic growth model:

$$\max_{\{k_{t+1},c_t\}_{t=0}^{\infty}} E_0 \sum_{t=0}^{\infty} \beta^t u\left(c_t\right)$$

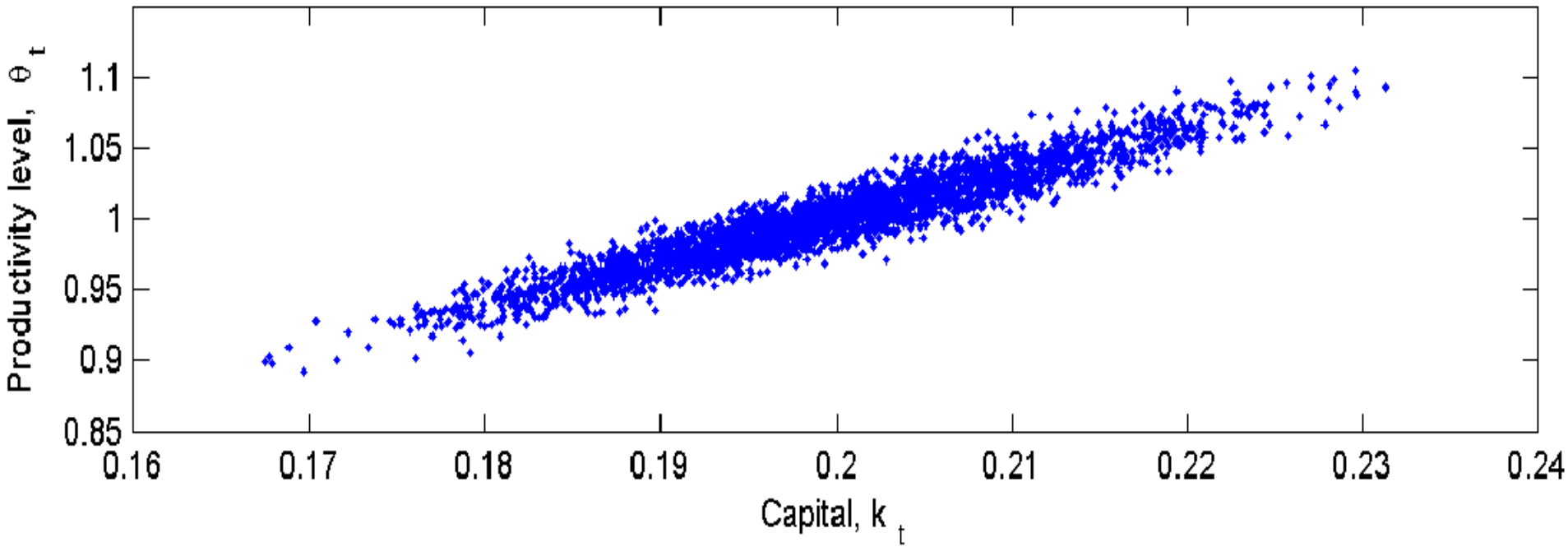
$$\text{s.t.} \quad c_t + k_{t+1} = \left(1-\delta\right) k_t + \theta_t f\left(k_t\right),$$

$$\ln \theta_{t+1} = \rho \ln \theta_t + \epsilon_{t+1}, \quad \epsilon_{t+1} \sim \mathcal{N}\left(0, \sigma^2\right),$$

where initial condition $\left(k_0, \theta_0\right)$ is given;
$u\left(\cdot\right)$ = utility function; $f\left(\cdot\right)$ = production function;
c_t = consumption; k_{t+1} = capital; θ_t = productivity;
β = discount factor; δ = depreciation rate of capital;
ρ = autocorrelation coefficient of the productivity level;
σ = standard deviation of the productivity shock.

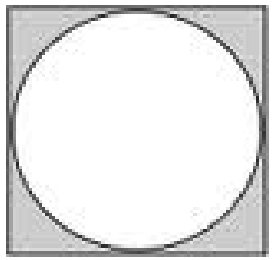
Advantage of stochastic simulation method

Advantage of stochastic simulation method: "Grid" is adaptive: we solve the model only in the area of the state space that is visited in simulation.



Reduction in cost in a 2-dimensional case

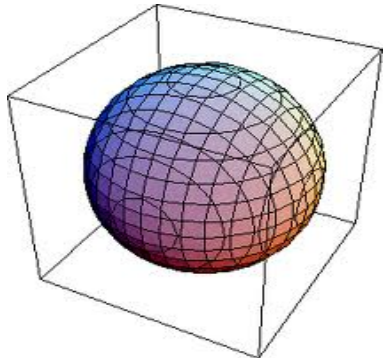
- How much can we save on cost using the ergodic-set domain comparatively to the hypercube domain?
- Suppose the ergodic set is a circle (it was an ellipse in the figure).
- In the 2-dimensional case, a circle inscribed within a square occupies about 79% of the area of the square.
- The reduction in cost is proportional to the shaded area in the figure.



- It does not seem to be a large gain.

Reduction in cost in a d-dimensional case

- In a 3-dimensional case, the gain is larger (a volume of a sphere of diameter 1 is 52% of the volume of a cube of width 1)



- In a d-dimensional case, the ratio of a hypersphere's volume to a hypercube's volume

$$\mathcal{V}^d = \begin{cases} \frac{(\pi/2)^{\frac{d-1}{2}}}{1 \cdot 3 \cdot \ldots \cdot d} & \text{for } d = 1, 3, 5... \\ \frac{(\pi/2)^{\frac{d}{2}}}{2 \cdot 4 \cdot \ldots \cdot d} & \text{for } d = 2, 4, 6... \end{cases}.$$

- $\mathcal{V}^d$ declines very rapidly with dimensionality of state space. When $d = 10 \Rightarrow \mathcal{V}^d = 3 \cdot 10^{-3}$ (0.3%). When $d = 30 \Rightarrow \mathcal{V}^d = 2 \cdot 10^{-14}$.
- We face a tiny fraction of cost we would have faced on the hypercube.

Shortcomings of stochastic simulation approach

1. Simulated points are not an efficient choice for constructing a grid:
   1. there are many closely situated and hence, redundant points;
   2. there are points outside the high-probability area.
2. Simulated points are not an efficient choice for the purpose of integration – accuracy of Monte Carlo integration is low, e.g., Parameterized Expectation Algorithm (PEA) by den Haan and Marcet (1990).

$$E_t\left[y_{t+1}\right] \approx \overline{y}_{t+1} \equiv \sum_{\tau=1}^{n} y_{\tau+1}$$

$$\text{Suppose } std\left(y_{\tau+1}\right) = 1\%$$

$$n = 1 \text{ draws } \Rightarrow std\left(\overline{y}_{t+1}\right) = 1\%$$

$$n = 100 \text{ draws } \Rightarrow std\left(\overline{y}_{t+1}\right) = 0.1\%$$

$$n = 10,000 \text{ draws } \Rightarrow std\left(\overline{y}_{t+1}\right) = 0.01\%$$

Monte Carlo method has a slow, $\sqrt{n}$, rate of convergence.

Inefficiency of Monte Carlo integration

Why is Monte Carlo integration inefficient?

- Because we compute expectations from noisy simulated data as do econometricians who do not know true density of DGP.
- But we do know the true density of DGP (we define productivity ourselves, $\ln \theta_{t+1} = \rho \ln \theta_t + \epsilon_{t+1}$).
- We can compute integrals far more accurately using quadrature methods based on true density of DGP!

Generalized Stochastic Simulation Algorithm

Stochastic simulation methods and their shortcomings

- The aim of GSSA is to improve the performance of stochastic simulation methods.
- The regression step requires
  - to fit an approximating function to the simulated data (regression);
  - to evaluate conditional expectations (integration).
- We show that both regression and integration have two problems:
  - *Problem 1:* In regression, polynomial terms are highly correlated (multicollinearity), and the standard LS technique fails $\Rightarrow$ **numerical instability.**
  - *Problem 2:* Monte Carlo integration is very inaccurate $\Rightarrow$ **the overall accuracy of solutions is low.**

With GSSA, we correct the above two shortcomings

- We stabilize the stochastic simulation procedure:
  - *we build the regression step on approximation methods designed for dealing with multicollinearity*
- We attain high accuracy of solutions:
  - *we generalize the stochastic simulation algorithm to include accurate Gauss Hermite quadrature and monomial integration methods*
- The generalized stochastic simulation algorithm (GSSA) is
  - *numerically stable*
  - *comparable in accuracy to most accurate methods in the literature*
  - *tractable in problems with high dimensionality (hundreds of state variables)*
  - *very simple to program*

Addressing Problem 1: Attaining numerical stability

1. As an approximating function, we use ordinary polynomial $\Psi(k, a; b) = b_0 + b_1 \ln k_t + b_2 \ln a_t + ... + b_n (\ln a_t)^L$.
2. We use approximation methods that can handle collinear data and dampen movements in b.
   - LS using SVD, Tikhonov regularization;
   - Least absolute deviations (LAD) methods (primal and dual linear programming problems);
   - Principal components (truncated SVD) method.
3. Other factors that can affect numerical stability of GSSA:
   - Data normalization.
   - The choice of a family of basis functions.
   - The choice of policy functions to parameterize.

GSSA algorithm I

Stage 1.

- Initialization:
  - Choose an initial guess $b^{(1)}$.
  - Choose a simulation length, T, draw a sequence of productivity shocks, $\{\epsilon_t\}_{t=1,\ldots,T}$, and compute $\{a_t\}_{t=1,\ldots,T+1}$ from the process for a_t.
  - Choose the initial state (k_0, a_0) for simulations.
- Step 1. At iteration p, use $b^{(p)}$ to simulate the model T periods forward,

$$\begin{aligned} k_{t+1} &= \widehat{K}\left(k_t, a_t; b^{(p)}\right), \\ c_t &= (1-\delta)\, k_t + a_t f\left(k_t\right) - k_{t+1}. \end{aligned}$$

GSSA algorithm II

- Step 2. For $t = 0, ..., T-1$, define y_t to be an approximation of the conditional expectation using J integration nodes and weights, $\{\epsilon_{t+1,j}\}_{j=1,...,J}$ and $\{\omega_{t,j}\}_{j=1,...,J}$, respectively:

$$y_t = \sum_{j=1}^{J} \left\{ \omega_{t,j} \cdot \left(\beta \frac{u'\left(c_{t+1,j}\right)}{u'\left(c_t\right)} \left[1 - \delta + a_{t+1,j}\ f'\left(k_{t+1}\right)\right]\ \ k_{t+1} \right) \right\}, \tag{1}$$

where $c_{t+1,j}$, the value of c_{t+1} if the innovation in productivity is $\epsilon_{t+1,j}$, is defined for $j = 1, ..., J$ by

$$\begin{aligned} a_{t+1,j} &\equiv a_t^{\rho} \exp\left(\epsilon_{t+1,j}\right), \\ k_{t+2,j} &\equiv \widehat{K}\left(\widehat{K}\left(k_t, a_t; b^{(p)}\right), a_t^{\rho} \exp\left(\epsilon_{t+1,j}\right); b^{(p)}\right), \\ c_{t+1,j} &\equiv \left(1-\delta\right) k_{t+1} + a_{t+1,j} f\left(k_{t+1}\right) - k_{t+2,j}. \end{aligned}$$

GSSA algorithm III

- Step 3. Find $\widehat{b}$ that minimizes the errors ε_t in the regression equation according to some norm, $\|\cdot\|$,

$$y_t = \widehat{K}\left(k_t, a_t; b\right) + \varepsilon_t. \tag{2}$$

- Step 4: Check for convergence and end Stage 1 if

$$\frac{1}{T}\sum_{t=1}^{T}\left|\frac{k_{t+1}^{(p)} - k_{t+1}^{(p+1)}}{k_{t+1}^{(p)}}\right| < \varpi, \tag{3}$$

where $\left\{k_{t+1}^{(p)}\right\}_{t=1}^{T}$ and $\left\{k_{t+1}^{(p+1)}\right\}_{t=1}^{T}$ are the capital series obtained on iterations p and $p+1$, respectively.

- Step 5. Compute $b^{(p+1)}$ for iteration $(p+1)$ using fixed-point iteration

$$b^{(p+1)} = (1-\xi)\, b^{(p)} + \xi \widehat{b}, \tag{4}$$

where $\xi \in (0, 1]$ is a damping parameter. Go to Step 1.

GSSA algorithm IV

Stage 2.

- Construct a new set of T^{test} points $\left\{k_\tau, a_\tau\right\}_{\tau=0}^{T^{test}}$ for testing the accuracy of the solution obtained in Stage 1 (this can be a set of simulation points constructed with a new random draw or some deterministic set of points).
- Re-write the Euler equation at (k_τ, a_τ) in a unit-free form,

$$\mathcal{R}\left(k_\tau, a_\tau\right) \equiv E_\tau\left\{\beta \frac{u'\left(c_{\tau+1}\right)}{u'\left(c_\tau\right)}\left[1-\delta+a_{\tau+1} f'\left(k_{\tau+1}\right)\right]\right\}-1. \quad (5)$$

- For each point (k_τ, a_τ), compute $\mathcal{R}\left(k_\tau, a_\tau\right)$ by using a high-quality integration method in evaluating the conditional expectation in (5).
- If the economic significance of these errors is small, we accept the candidate b.
- Otherwise, we tighten up Stage 1 (use a more flexible approximating function, increase simulation length, improve integration method, choose more demanding norm) when computing $\widehat{b}$ in Step 3.

Methodology and parameterization

- Production function: $f(k_t) = k_t^{\alpha}$ with $\alpha = 0.36$.
- Utility function: $u(c_t) = \frac{c_t^{1-\gamma}-1}{1-\gamma}$ with $\gamma \in \{0.1, 1, 10\}$.
- Process for shocks: $\rho = 0.95$ and $\sigma = 0.01$.
- Discount factor: $\beta = 0.99$.
- Depreciation rate: $\delta = 1$ and $\delta = 0.02$.
- Under $\gamma = 1$ and $\delta = 1 \Longrightarrow$ *closed-form solution*.
- Accuracy is measured by an Euler-equation error,

$$\mathcal{R}(k_t, a_t) \equiv E_t \left[\beta \frac{c_{t+1}^{-\gamma}}{c_t^{-\gamma}} \left(1 - \delta + \alpha a_{t+1} k_{t+1}^{\alpha-1}\right) \right] - 1,$$

expressed in log10 units.

Results for the model with the closed-form solution

Full depreciation of capital, $\delta = 1$.

	$\mathcal{R}_{mean}$	CPU	$\mathcal{R}_{mean}$	CPU	$\mathcal{R}_{mean}$	CPU
Polyn. degree	*OLS, Ordinary Unnormalized*		*OLS, Ordinary Normalized*		*OLS, Hermite Unnormalized*	
1st	-3.52	0.8 sec	-3.52	1 sec	-3.52	1 sec
2nd	-5.46	3.1 sec	-5.46	3 sec	-5.46	4 sec
3rd	-	-	-6.84	5 sec	-6.84	6 sec
4th	-	-	-	-	-7.94	8 sec
5th	-	-	-	-	-9.09	10 sec
	Ordinary, LS-SVD Normalized		*Ordinary, LAD-PP Normalized*		*Ordinary, RLS-Tikh.* $\eta = 10^{-7}$	
1st	-3.52	1 sec	-3.52	16 sec	-3.52	1 sec
2nd	-5.46	3 sec	-5.55	1.5 min	-5.46	3 sec
3rd	-6.84	5 sec	-6.97	4.1 min	-5.85	4 sec
4th	-7.94	6 sec	-8.16	6.4 min	-6.12	7 sec
5th	-9.12	10 sec	-9.10	9.3 min	-6.22	11 sec

Results for the model without a closed-form solution

Partial depreciation of capital, $\delta = 0.02$.

	$\mathcal{R}_{mean}$	CPU
Polyn. degree	*MC(1)* $T = 10,000$	
1st	-4.26	1 sec
2nd	-4.42	11 sec
3rd	-4.32	25 sec
4th	-4.31	47 sec
5th	-4.23	80 sec

- We attain stability but now high-degree polynomials do not lead to more accurate solution. Why?
- Recall that low accuracy of Monte Carlo integration restricts the overall accuracy of solutions, e.g., PEA by den Haan and Marcet (1990).

Addressing Problem 2: Deterministic integration methods

Our GSSA relies on accurate Gauss Hermite quadrature integration

$$\int_{\mathbb{R}^N} g(\varepsilon) w(\varepsilon) d\varepsilon \approx \sum_{j=1}^{J} \omega_j g(\varepsilon_j),$$

where $\{\varepsilon_j\}_{j=1}^{J}$ = integration nodes, $\{\omega_j\}_{j=1}^{J}$ = integration weights.

Example

a) A two-node Gauss-Hermite quadrature method, $Q(2)$, uses nodes $\epsilon_1 = -\sigma$, $\epsilon_2 = \sigma$ and weights $\omega_1 = \omega_2 = \frac{1}{2}$.
b) A three-node Gauss-Hermite quadrature method, $Q(3)$, uses nodes $\epsilon_1 = 0$, $\epsilon_2 = \sigma\sqrt{\frac{3}{2}}$, $\epsilon_3 = -\sigma\sqrt{\frac{3}{2}}$ and weights $\omega_1 = \frac{2\sqrt{\pi}}{3}$, $\omega_2 = \omega_3 = \frac{\sqrt{\pi}}{6}$.
c) A one-node Gauss-Hermite quadrature method, $Q(1)$, uses a zero node, $\epsilon_1 = 0$, and a unit weight, $\omega_1 = 1$.

Quadrature integration in the studied model

For $t = 0, ..., T-1$, we approximation the conditional expectation as

$$y_t = \sum_{j=1}^{J} \left\{ \omega_j \cdot \left(\beta u' \left(c_{t+1,j} \right) \left[1 - \delta + a_{t+1,j} \; f' \left(k_{t+1} \right) \right] \right) \right\},$$

where $c_{t+1,j}$, the value of c_{t+1} if the innovation in productivity is ϵ_j, is defined for $j = 1, ..., J$ by

$$\begin{aligned} a_{t+1,j} &\equiv a_t^{\rho} \exp\left(\epsilon_j\right), \\ c_{t+1,j} &\equiv F\left(k_{t+1}, a_t^{\rho} \exp\left(\epsilon_j\right); b^{(p)} \right). \end{aligned}$$

where $\{\epsilon_j\}_{j=1,...,J}$ and $\{\omega_j\}_{j=1,...,J}$ are J integration nodes and weights, respectively.

Results for the model with partial depreciation of capital

	$\mathcal{R}_{mean}$	CPU	$\mathcal{R}_{mean}$	CPU	$\mathcal{R}_{mean}$	CPU
Polyn. degree	*MC(1)* $T = 10,000$		*MC(2000)* $T = 10,000$		*MC(1)* $T = 100,000$	
1st	-4.26	1 sec	-4.40	20.6 min	-4.39	4 sec
2nd	-4.42	11 sec	-6.04	28.5 min	-4.87	1.3 min
3rd	-4.32	25 sec	-6.15	36.6 min	-4.86	3.1 min
4th	-4.31	47 sec	-6.08	55.6 min	-4.72	5.7 min
5th	-4.23	80 sec	-6.07	1.27 h	-4.71	10.4 min
	Q(1) $T = 100$		*Q(2)* $T = 10,000$		*Q(10)* $T = 10,000$	
1st	-4.36	3 sec	-4.36	16 sec	-4.36	20 sec
2nd	-6.05	4 sec	-6.13	27 sec	-6.13	34 sec
3rd	-6.32	5 sec	-7.48	35 sec	-7.48	44 sec
4th	-6.24	6 sec	-8.72	44 sec	-8.72	54 sec
5th	-6.04	7 sec	-8.91	51 sec	-8.91	63 sec

RLS-TSVD with $\kappa = 10^7$

Multi-dimensional problems: Gauss Hermite product rules

In multi-dimensional problem, we can use Gauss Hermite product rules.

Example

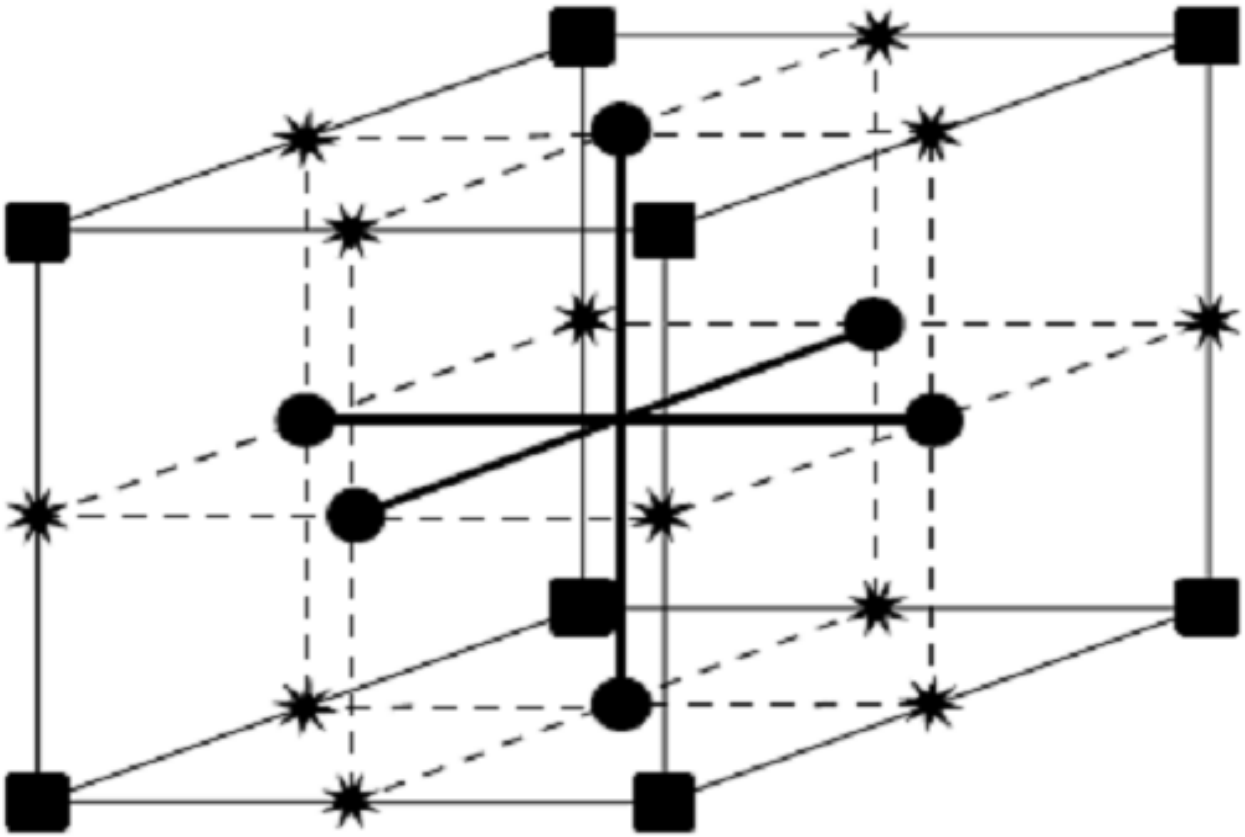
Let $\epsilon_{t+1}^{h} \sim \mathcal{N}\left(0, \sigma^2\right)$, $h = 1, 2, 3$ be uncorrelated random variables. A two-node Gauss-Hermite product rule, $Q\left(2\right)$, (obtained from the two-node Gauss-Hermite rule) has 2^3 nodes, which are as follows:

	$j=1$	$j=2$	$j=3$	$j=4$	$j=5$	$j=6$	$j=7$	$j=8$
$\epsilon_{t+1,j}^{1}$	σ	σ	σ	σ	$-\sigma$	$-\sigma$	$-\sigma$	$-\sigma$
$\epsilon_{t+1,j}^{2}$	σ	σ	$-\sigma$	$-\sigma$	σ	σ	$-\sigma$	$-\sigma$
$\epsilon_{t+1,j}^{3}$	σ	$-\sigma$	σ	$-\sigma$	σ	$-\sigma$	σ	$-\sigma$

where weights of all nodes are equal, $\omega_{t,j} = 1/8$ for all j.

The cost of product rules increases exponentially, 2^N, with the number of exogenous state variables, N. Such rules are not practical when the dimensionality is high.

Deterministic integration



Types of nodes: the center; the circles (6 centers of faces); the stars (12 centers of edges); the squares (8 vertices).

Monomial non-product integration formulas

Monomial formulas are a cheap alternative for multi-dimensional problem (there is a variety of such formulas differing in accuracy and cost).

Example

Let $\epsilon_{t+1}^{h} \sim \mathcal{N}\left(0, \sigma^2\right)$, $h = 1, 2, 3$ be uncorrelated random variables. Consider the following monomial (non-product) integration rule with $2 \cdot 3$ nodes:

	$j=1$	$j=2$	$j=3$	$j=4$	$j=5$	$j=6$
$\epsilon_{t+1,j}^{1}$	$\sigma\sqrt{3}$	$-\sigma\sqrt{3}$	0	0	0	0
$\epsilon_{t+1,j}^{2}$	0	0	$\sigma\sqrt{3}$	$-\sigma\sqrt{3}$	0	0
$\epsilon_{t+1,j}^{3}$	0	0	0	0	$\sigma\sqrt{3}$	$-\sigma\sqrt{3}$

where weights of all nodes are equal, $\omega_{t,j} = 1/6$ for all j.

Monomial rules are practical for problems with very high dimensionality, for example, with $N = 100$, this rule has only $2N = 200$ nodes.

The multi-country model

The planner maximizes a weighted sum of N countries' lifetime utilities

$$\max_{\left\{\left\{c_t^h, k_{t+1}^h\right\}_{h=1}^N\right\}_{t=0}^{\infty}} E_0 \sum_{h=1}^{N} \lambda^h \left(\sum_{t=0}^{\infty} \beta^t u^h \left(c_t^h \right) \right)$$

subject to

$$\sum_{h=1}^{N} c_t^h + \sum_{h=1}^{N} k_{t+1}^h = \sum_{h=1}^{N} k_t^h \left(1-\delta\right) + \sum_{h=1}^{N} a_t^h f^h \left(k_t^h \right),$$

where λ^h is country h's welfare weight.
Productivity of country h follows the process

$$\ln a_{t+1}^h = \rho \ln a_t^h + \epsilon_{t+1}^h,$$

where $\epsilon_{t+1}^h \equiv \varsigma_{t+1} + \varsigma_{t+1}^h$ with $\varsigma_{t+1} \sim \mathcal{N}\left(0, \sigma^2\right)$ is identical for all countries and $\varsigma_{t+1}^h \sim \mathcal{N}\left(0, \sigma^2\right)$ is country-specific.

Results for the multi-country model

Numb. of countr.	Polyn. degree	Numb. of coeff.	$\mathcal{R}_{mean}$	CPU	$\mathcal{R}_{mean}$	CPU
			RLS-Tikh.,$\eta = 10^{-5}$		*RLS-TSVD*, $\kappa = 10^7$	
			MC(1), $T = 10,000$		M2, $T = 1000$	
N=2	1st	5	-4.70	4.2 min	-4.65	37 sec
	2nd	15	-4.82	19.3 min	-6.01	6.8 min
	3rd	35	-4.59	57 min	-7.09	10.4 min
	4th	70	-4.57	2.6 hours	-7.99	16.3 min
	5th	126	-4.53	6.8 hours	-8.00	34.8 min
			RLS-Tikh.,$\eta = 10^{-5}$		*RLS-Tikh.*, $\eta = 10^{-5}$	
			MC(1), $T = 10,000$		*Q(1)*, $T = 1000$	
N=20	1st	41	-4.55	6.5 min	-4.75	56 sec
	2nd	861	-3.88	2.1 hours	-5.40	18 min
N=200	1st	401	-3.97	37.2 min	-4.59	16.8 min

When N=200, for *RLS-Tikh.*,*Q(1)*, we use $T = 2000$

Epsilon Distinguishable Set and Cluster Grid Algorithms

Merging projection and stochastic simulation

EDS and CGA algorithms merge stochastic simulation and projection approaches.

What do we do?

- **Similar to stochastic simulation approach:** use simulation to identify and approximate the ergodic set.
- **Similar to projection approach:** construct a fixed grid and use quadrature integration to accurately solve the model on that grid.
- We use integration and optimization methods that are tractable in high-dimensional problems: *non-product monomial integration formulas and derivative-free solvers*.

Epsilon-distinguishable set (EDS) algorithm

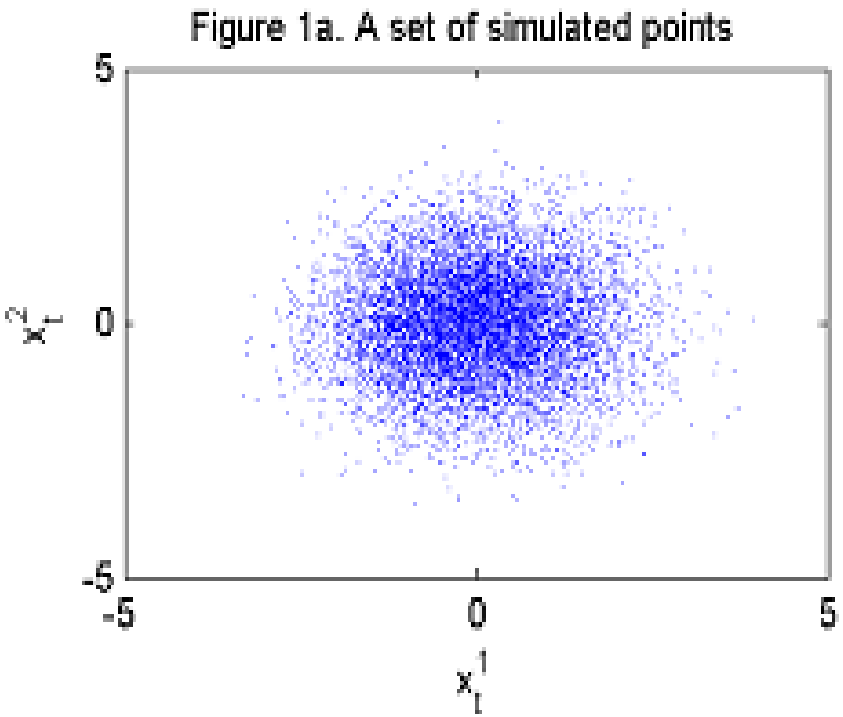
The key novel piece of our analysis

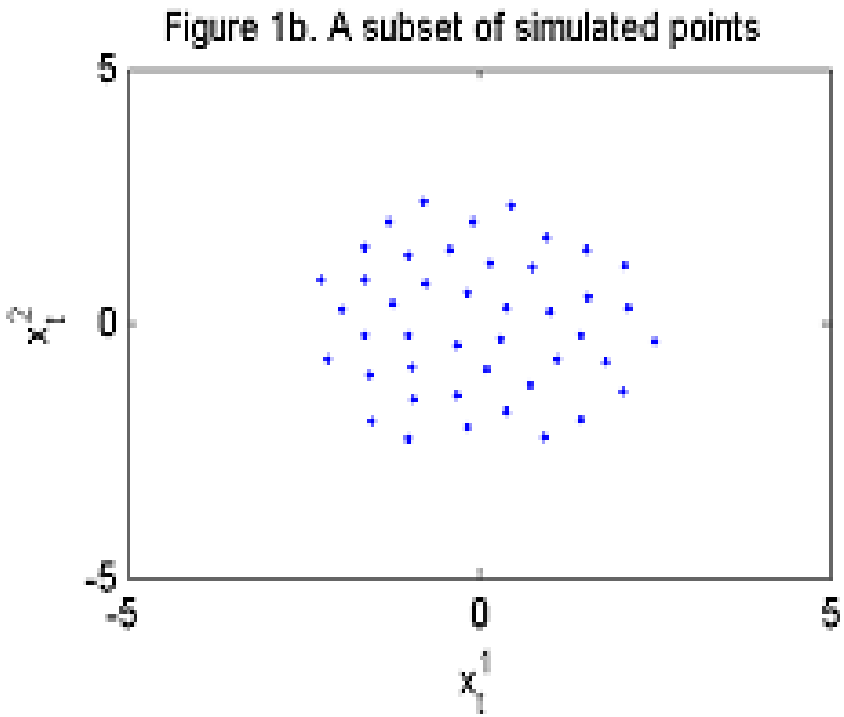
The EDS grid construction:

- we select *an ε-distinguishable subset of simulated points* that covers the support of the ergodic measure roughly uniformly.
- "*ε-distinguishable set (EDS)*" = a set of points situated at the distance at least ε from one another, where $\varepsilon > 0$ is a parameter.

A grid of points covering support of the ergodic measure

An illustration of an *ε-distinguishable set.*





A class of stochastic processes

Suppose we know the solution to the model.
A class of discrete-time stochastic processes:

$$x_{t+1} = \varphi\left(x_t, \epsilon_{t+1}\right), \qquad t = 0, 1, ...,$$

$\epsilon \in E \subseteq \mathbb{R}^p$ = vector of p independent and identically distributed shocks;
$x \in X \subseteq \mathbb{R}^d$ = vector of d (exogenous and endogenous) state variables;
x is endowed with its relative Borel σ-algebra denoted by $\mathbb{X}$.

- *Example,* $k_{t+1} = K\left(k_t, \theta_t\right)$ *and* $\theta_{t+1} = \theta_t^{\rho} \exp\left(\epsilon_{t+1}\right)$.

Assumption 1. *There exists a unique ergodic set* $\mathcal{A}^*$ *and the associated ergodic measure* μ.
Assumption 2. *The ergodic measure* μ *admits a representation in the form of a density function* $g : X \to \mathbb{R}^+$ *such that* $\int_{\mathcal{A}} g\left(x\right) dx = \mu\left(\mathcal{A}\right)$ *for every* $\mathcal{A} \subseteq \mathbb{X}$.

A two-step EDS technique

A two-step procedure for forming a discrete approximation to the ergodic set.

1. We identify an area of the state space that contains nearly all the probability mass.
2. We cover this area with a finite set of points that are roughly evenly spaced.

An essentially ergodic set

We define a high-probability area of the state space using the level set of the density function g.

Def. *A set* $\mathcal{A}^{\eta} \subseteq \mathcal{A}^{*}$ *is called a* η*-level ergodic set if* $\eta > 0$ *and*

$$\mathcal{A}^{\eta} \equiv \{x \in X : g(x) \geq \eta\}.$$

- The mass of $\mathcal{A}^{\eta}$ under the density $g(x)$ is equal to $p(\eta) \equiv \int_{g(x) \geq \eta} g(x)\, dx$.
- If $p(\eta) \approx 1$, then $\mathcal{A}^{\eta}$ contains all X except for points where the density is lowest.
- In this case, $\mathcal{A}^{\eta}$ is called an *essentially ergodic set*.

Law of iterated logarithm

LIL: The ergodic measure can be approximated by simulation.
P = random draws $x_1, ..., x_n \subseteq \mathbb{R}^d$ generated with $\mu : \mathbb{R}^d \to \mathbb{R}^+$.
$C(P; J)$ = counts the number of points from P in a given $J \subseteq \mathbb{R}^d$.
$\mathcal{J}$ = intersection of all subintervals $\Pi_{i=1}^{d} [-\infty, v_i)$, where $v_i > 0$.

Proposition: (Law of iterated logarithm). For every dimensionality d and every continuous function μ, we have

$$\lim_{n\to\infty} \left\{ \sup_{J\in\mathcal{J}} \left| \frac{C(P; J)}{n} - \mu(J) \right| \cdot \left(\frac{2n}{\log\log n} \right)^{1/2} \right\} = 1, \quad \text{a.e.}$$

Proof: See Kiefer (1961, Theorem 2).

That is, the empirical distribution function $\widehat{\mu}(J) \equiv \frac{C(P;J)}{n}$ converges asymptotically to the true distribution function $\mu(J)$ for every $J \in \mathcal{J}$ at the rate given by $\left(\frac{2n}{\log\log n} \right)^{1/2}$.

Multivariate kernel density estimation

(Algorithm $\mathcal{A}^{\eta}$): Selection of points within an essentially ergodic set.

Step 1. Simulate $x_{t+1} = \varphi(x_t, \epsilon_{t+1})$ for T periods.
Step 2. Select each κth point to get a set P of n points $x_1, ..., x_n \in X \subseteq \mathbb{R}^d$.
Step 3. Estimate the density function $\widehat{g}(x_i) \approx g(x_i)$ for all $x_i \in P$.
Step 4. Remove all points for which the density is below η.

To estimate the density function $\widehat{g}$ from the simulated data, we use a multivariate kernel algorithm

$$\widehat{g}(x) = \frac{1}{n(2\pi)^{d/2}\overline{h}^d} \sum_{i=1}^{n} \exp\left[-\frac{D(x, x_i)}{2\overline{h}^2}\right],$$

where $\overline{h}$ is the bandwidth parameter, and $D(x, x_i)$ is the distance between x and x_i.

- The complexity of Algorithm $\mathcal{A}^{\eta}$ is $O(n^2)$ because it requires to compute pairwise distances between all the sample points.
- We remove 5% of the sample which has the lowest density.

Constructing EDS

Def. *Let* (X, D) *be a bounded metric space. A set* P^{ε} *consisting of points* $x_1^{\varepsilon}, ..., x_M^{\varepsilon} \in X \subseteq \mathbb{R}^d$ *is called* ε*-distinguishable if* $D\left(x_i^{\varepsilon}, x_j^{\varepsilon}\right) > \varepsilon$ *for all* $1 \leq i, j \leq M : i \neq j$*, where* $\varepsilon > 0$ *is a parameter.*

(**Algorithm** P^{ε}**): Construction of an** EDS.
Let P be a set of n point $x_1, ..., x_n \in X \subseteq \mathbb{R}^d$.
Let P^{ε} begin as an empty set, $P^{\varepsilon} = \{\varnothing\}$.
Step 1. Select $x_i \in P$. Compute $D\left(x_i, x_j\right)$ to all x_j in P.
Step 2. Eliminate from P all x_j for which $D\left(x_i, x_j\right) < \varepsilon$.
Step 3. Add x_i to P^{ε} and eliminate it from P.
Iterate on Steps 1-3 until all points are eliminated from P.

Proposition: The complexity of Algorithm P^{ε} is of order $O\left(nM\right)$.

Measuring distance between points

- Both estimating the density and constructing an EDS requires us to measure the distance between simulated points.
- Generally, variables in economic models have different measurement units and are correlated.
- This affects the distance between the simulated points and hence, affects the resulting EDS.
- Therefore, prior to using Algorithm $\mathcal{A}^{\eta}$ and Algorithm P^{ε}, we normalize and orthogonalize the simulated data using *Principal Component* transformation.

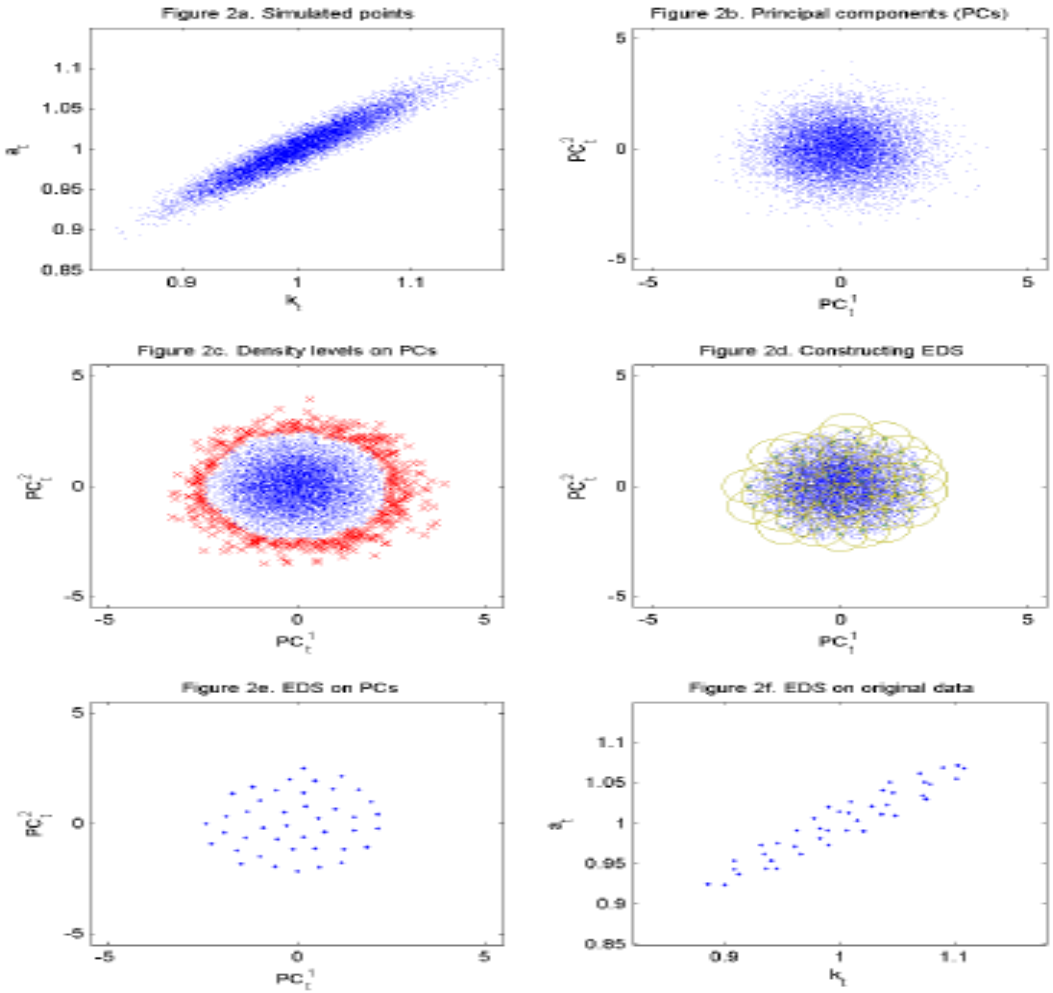
Principal component transformation

- Let $X \in \mathbb{R}^{n \times d}$ be simulated data normalized to zero mean and unit variance.
- Perform the singular value decomposition of X, i.e., $X = UQV^{\top}$, where $U \in \mathbb{R}^{n \times d}$ and $V \in \mathbb{R}^{d \times d}$ are orthogonal matrices, and $Q \in \mathbb{R}^{d \times d}$ is a diagonal matrix.
- Perform a linear transformation of X using $\text{PC} \equiv XV$.
- $\text{PC} = \left(\text{PC}^1, ..., \text{PC}^d\right) \in \mathbb{R}^{n \times d}$ are *principal components* (PCs) of X, and are orthogonal (uncorrelated), i.e., $\left(\text{PC}^{\ell'}\right)^{\top} \text{PC}^{\ell} = 0$ for any $\ell' \neq \ell$.
- Distance between two observations x_i and x_j is the Euclidean distance between their PCs

$$D\left(x_i, x_j\right) = \left[\sum_{\ell=1}^{d} \left(\text{PC}_i^{\ell} - \text{PC}_j^{\ell}\right)^2\right]^{1/2},$$

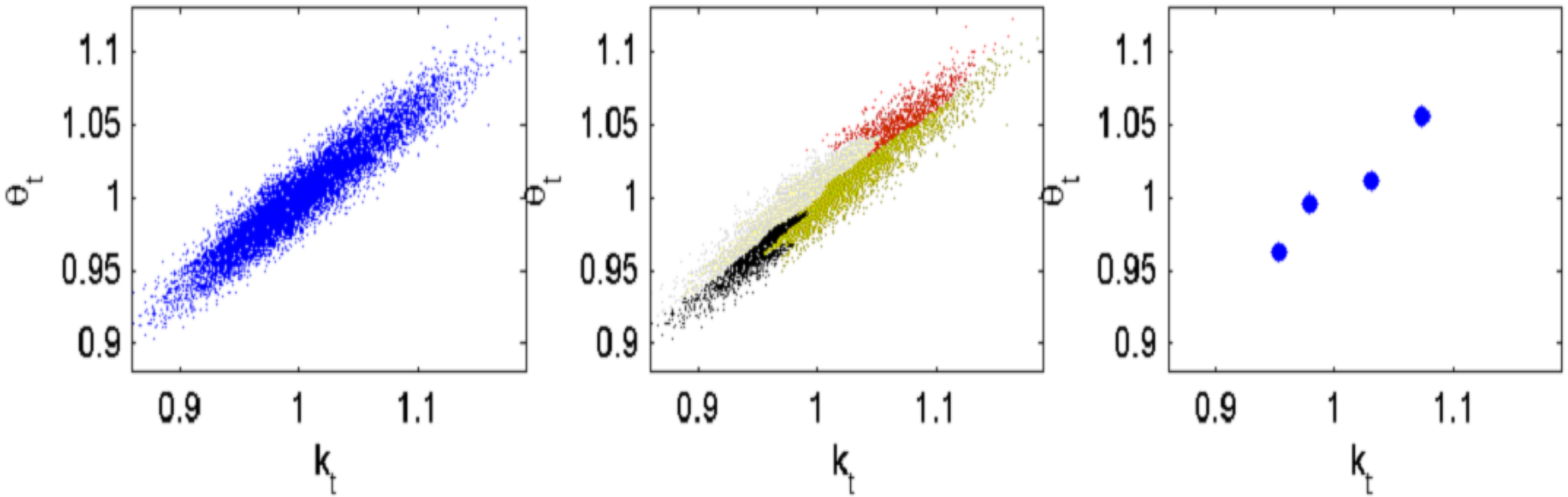
where PC^1, ..., PC^d are normalized to unit variance.

Illustrating the EDS technique



Cluster grid – another procedure for approximating the ergodic set

- Instead of constructing an EDS, we can use methods from cluster analysis to select a set of representative points from a given set of simulated points.
- We partition the simulated data into clusters (groups of closely-located points) and replace each cluster with one representative point.



Steps of the agglomerative hierarchical clustering algorithm

Clustering algorithm

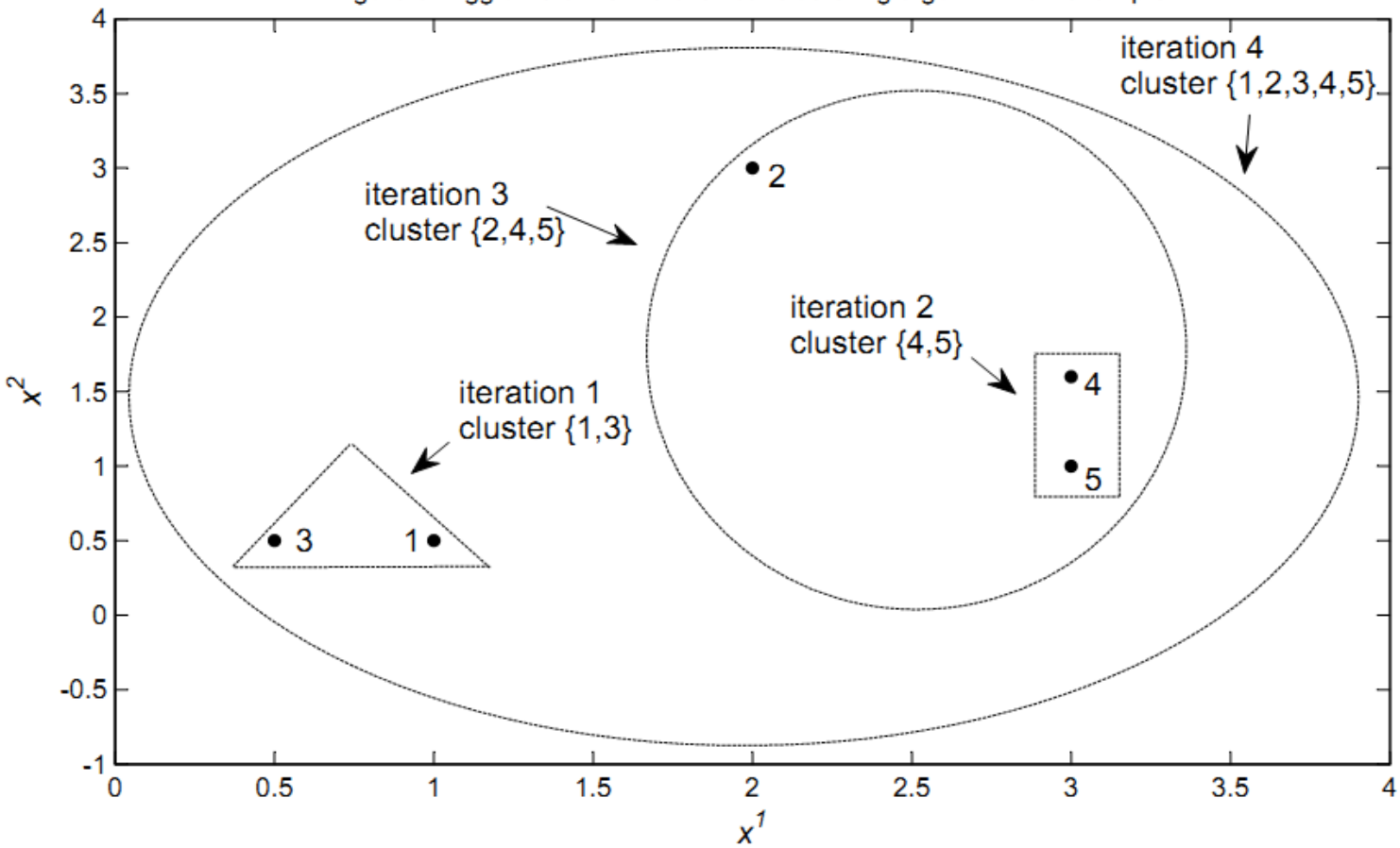
(Algorithm P^c): Agglomerative hierarchical clustering algorithm.

Initialization. Choose M, the number of clusters to be created.
In a zero-order partition $\mathcal{P}^{(0)}$, each simulated point represents a cluster.
Step 1. Compute all pairwise distances between the clusters in a partition $\mathcal{P}^{(i)}$.
Step 2. Merge a pair of clusters with the smallest distance to obtain $\mathcal{P}^{(i+1)}$.
Iterate on Steps 1 and 2. Stop when the number of clusters in the partition is M.
Represent each cluster with the closest to the cluster's center simulated point.

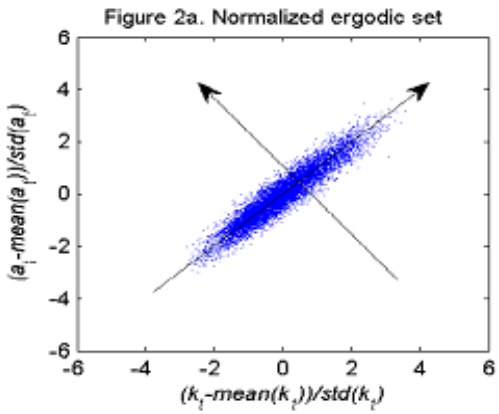
As a measure of distance between two clusters, we use Ward's measure of distance.

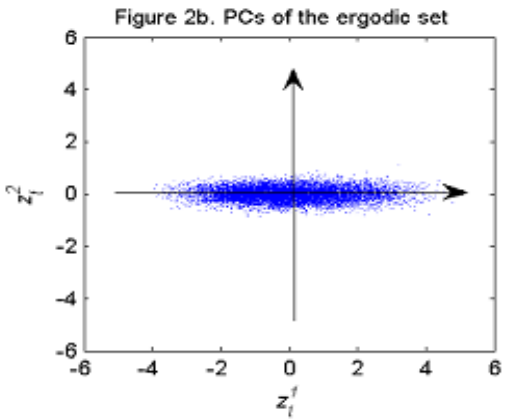
Agglomerative hierarchical clustering algorithm: an example

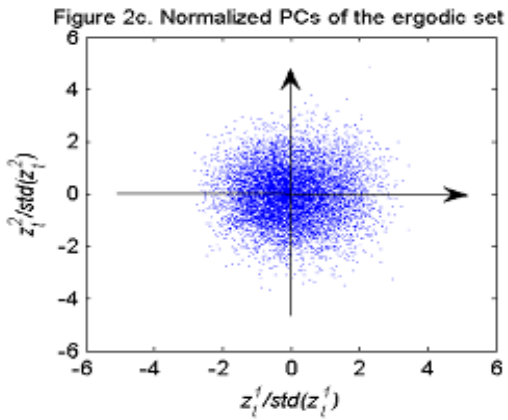
Figure 9. Agglomerative hierarchical clustering algorithm: an example.

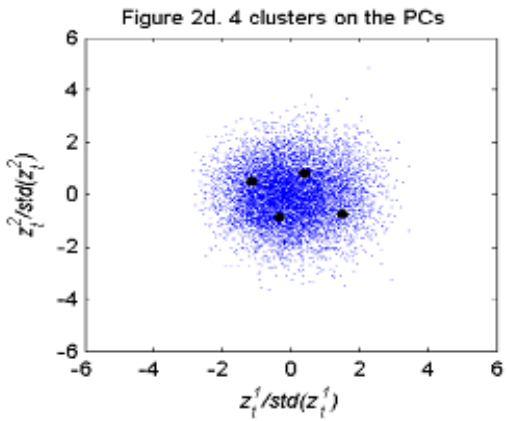


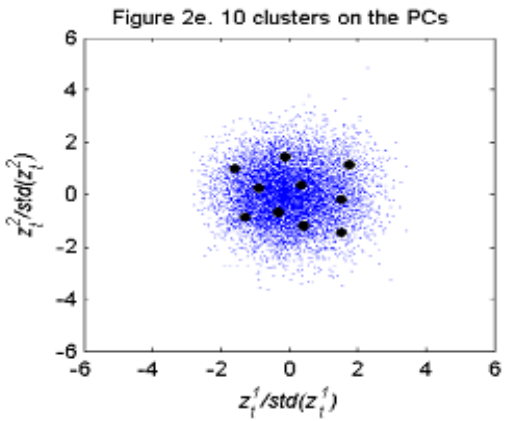
Clusters on principal components of the ergodic set

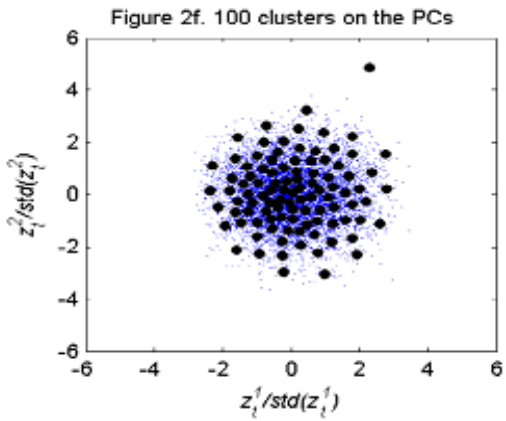












Description of the EDS algorithm iterating on Euler equation

Parameterize the RHS of the Euler equation by a polynomial $\widehat{K}(k, \theta; b)$,

$$E\left\{\beta \frac{u'(c')}{u'(c)}\left[1-\delta+\theta' f'(k')\right] k'\right\}$$

$$\approx \quad \widehat{K}(k, \theta; b) = b_0 + b_1 k + b_2 \theta + ... + b_n \theta^L$$

Step 1. Simulate $\{k_t, \theta_t\}_{t=1}^{T+1}$. Construct an EDS grid, $\{k_m, \theta_m\}_{m=1}^{M}$.

Step 2. Fix $b \equiv (b_0, b_1, b_2, ..., b_n)$. Given $\{k_m, \theta_m\}_{m=1}^{M}$ solve for $\{c_m\}_{m=1}^{M}$.

Step 3. Compute the expectation using numerical integration (quadrature integration or monomial rules)

$$\widehat{k}'_m \equiv E\left\{\beta \frac{u'(c'_m)}{u'(c_m)}\left[1-\delta+\theta'_m f'(k'_m)\right] k'_m\right\}.$$

Regress $\widehat{k}'_m$ on $\left(1, k_m, \theta_m, k_m^2, \theta_m^2, ..., \theta_m^L\right) \Longrightarrow$ get $\widehat{b}$.

Step 4. Solve for the coefficients using damping,

Representative-agent model: parameters choice

Production function: $f(k_t) = k_t^{\alpha}$ with $\alpha = 0.36$.
Utility function: $u(c_t) = \frac{c_t^{1-\gamma}-1}{1-\gamma}$ with $\gamma \in \left\{\frac{1}{5}, 1, 5\right\}$.
Process for shocks: $\ln \theta_{t+1} = \rho \ln \theta_t + \epsilon_{t+1}$ with $\rho = 0.95$ and $\sigma = 0.01$.
Discount factor: $\beta = 0.99$.
Depreciation rate: $\delta = 0.025$.
Accuracy is measured by an Euler-equation residual,

$$\mathcal{R}(k_i, \theta_i) \equiv E_i \left[\beta \frac{c_{i+1}^{-\gamma}}{c_i^{-\gamma}} \left(1 - \delta + \alpha \theta_{i+1} k_{i+1}^{\alpha-1}\right) \right] - 1.$$

Table 1. Accuracy and speed of the Euler equation EDS algorithm in the representative-agent model

Polynomial degree	Mean error	Max error	CPU (sec)
1st degree	−4.29	−3.31	24.7
2nd degree	−5.94	−4.87	0.8
3rd degree	−7.26	−6.04	0.9
4th degree	−8.65	−7.32	0.9
5th degree	−9.47	−8.24	5.5

Target number of grid points is $\overline{M} = 25$.
Realized number of grid points is $M(\varepsilon) = 27$.
Mean and Max are unit-free Euler equation errors in log10 units, e.g.,

- -4 means $10^{-4} = 0.0001$ (0.01%);
- -4.5 means $10^{-4.5} = 0.0000316$ (0.00316%).

Benchmark parameters: $\gamma = 1$, $\delta = 0.025$, $\rho = 0.95$, $\sigma = 0.01$.
In the paper, also consider $\gamma = 1/5$ (low risk aversion) and $\gamma = 5$ (high risk aversion). Accuracy and speed are similar.

Autocorrection of the EDS grid

Figure 4. Convergence of the EDS grid starting from capital series normalized to 10 steady state levels

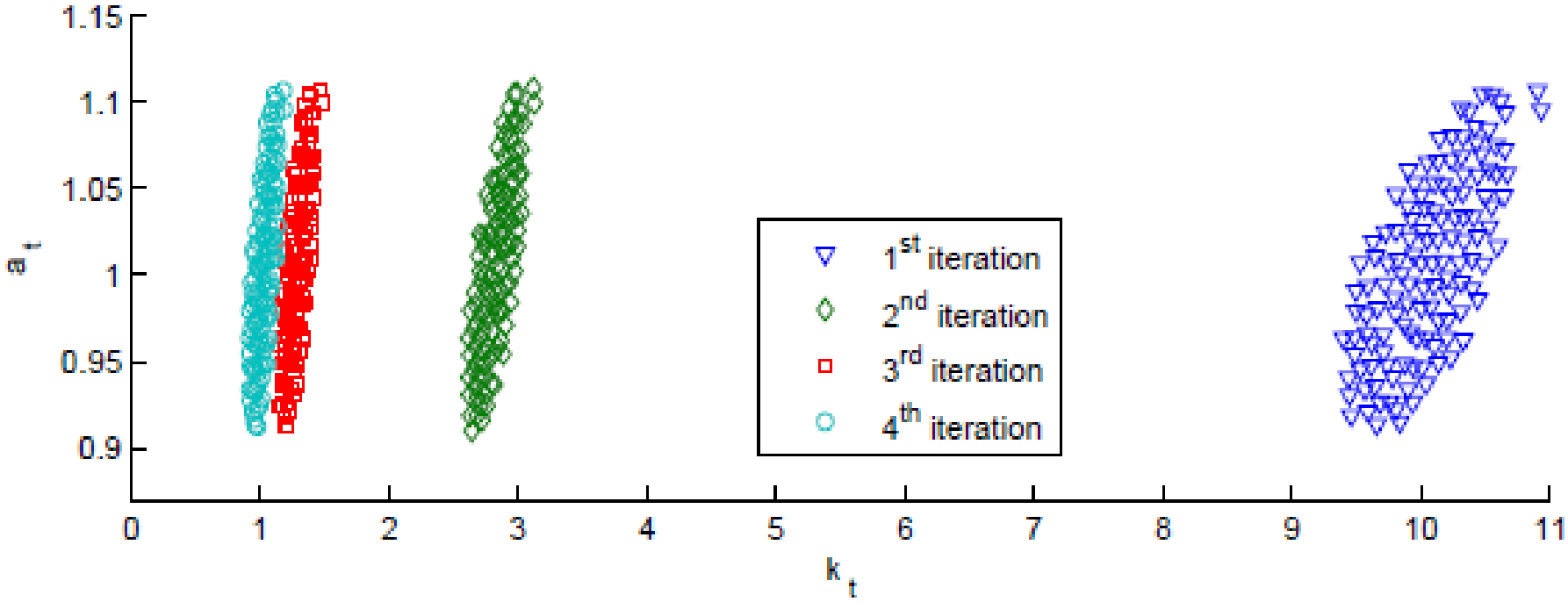


Table 2: Accuracy and speed in the one-agent model: Smolyak grid versus EDS grid

	Test on a simulation				Test on a hypercube			
Polyn.	Smolyak grid		EDS grid		Smolyak grid		EDS grid	
deg.	Mean	Max	Mean	Max	Mean	Max	Mean	Max
1st	-3.31	-2.94	-4.23	-3.31	-3.25	-2.54	-3.26	-2.38
2nd	-4.74	-4.17	-5.89	-4.87	-4.32	-3.80	-4.41	-3.25
3rd	-5.27	-5.13	-7.19	-6.16	-5.39	-4.78	-5.44	-4.11

Description of the EDS algorithm iterating on Bellman equation

Parameterize the value function by a polynomial $V(\cdot) \approx \widehat{V}(\cdot; b)$:

$$\max_{k',c} \left\{ u(c) + \beta E \left[\widehat{V}(k', \theta'; b) \right] \right\}$$

$$\approx \quad \widehat{V}(k, \theta; b) = b_0 + b_1 k + b_2 \theta + ... + b_n \theta^L.$$

Step 1. Find $\widehat{K}$ corresponding to $\widehat{V}(\cdot; b)$. Simulate $\{k_t, \theta_t\}_{t=1}^{T+1}$. Construct an EDS grid, $\{k_m, \theta_m\}_{m=1}^{M}$.

Step 2. Fix $b \equiv (b_0, b_1, b_2, ..., b_n)$. Given $\{k_m, \theta_m\}_{m=1}^{M}$ solve for $\{c_m\}_{m=1}^{M}$.

Step 3. Compute the expectation using numerical integration (quadrature integration or monomial rules)

$$V_m \equiv u(c_m) + \beta E \widehat{V}\left(k_m', \theta_m'; b\right).$$

Regress V_m on $\left(1, k_m, \theta_m, k_m^2, \theta_m^2, ..., \theta_m^L\right) \Longrightarrow$ get $\widehat{b}$.

Step 4. Solve for the coefficients using damping,

Table 3. Accuracy and speed of the Bellman equation EDS algorithm in the representative-agent model

Polynomial degree	Mean error	Max error	CPU (sec)
1st degree	–	–	–
2nd degree	−3.98	−3.11	0.5
3rd degree	−5.15	−4.17	0.4
4th degree	−6.26	−5.12	0.4
5th degree	−7.42	−5.93	0.4

Target number of grid points is $\overline{M} = 25$.
Realized number of grid points is $M(\varepsilon) = 27$.

Multi-country model

The planner maximizes a weighted sum of N countries' utility functions:

$$\max_{\left\{\left\{c_t^h, k_{t+1}^h\right\}_{h=1}^N\right\}_{t=0}^{\infty}} E_0 \sum_{h=1}^{N} v^h \left(\sum_{t=0}^{\infty} \beta^t u^h \left(c_t^h \right) \right)$$

subject to

$$\sum_{h=1}^{N} c_t^h + \sum_{h=1}^{N} k_{t+1}^h = \sum_{h=1}^{N} k_t^h \left(1-\delta\right) + \sum_{h=1}^{N} \theta_t^h f^h \left(k_t^h \right),$$

where v^h is country h's welfare weight.
Productivity of country h follows the process

$$\ln \theta_{t+1}^h = \rho \ln \theta_t^h + \epsilon_{t+1}^h,$$

where $\epsilon_{t+1}^h \equiv \varsigma_{t+1} + \varsigma_{t+1}^h$ with $\varsigma_{t+1} \sim \mathcal{N}\left(0, \sigma^2\right)$ is identical for all countries and $\varsigma_{t+1}^h \sim \mathcal{N}\left(0, \sigma^2\right)$ is country-specific.

Table 3. Accuracy and speed in the multi-country model

	Polyn. degree	M1			Q(1)		
		Mean	Max	CPU	Mean	Max	CPU
N=2	1st	−4.09	−3.19	44 sec	−4.07	−3.19	45 sec
	2nd	−5.45	−4.51	2 min	−5.06	−4.41	1 min
	3rd	−6.51	−5.29	4 min	−5.17	−4.92	2 min
N=20	1st	−4.21	−3.29	20 min	−4.17	−3.28	3 min
	2nd	−5.08	−4.17	5 hours	−4.83	−4.10	32 min
N=40	1st	−4.23	−3.31	5 hours	−4.19	−3.29	2 hours
	2nd	–	–	-	−4.86	−4.48	24 hours
N=100	1st	−4.09	−3.24	10 hours	−4.06	−3.23	36 min
N=200	1st	–	–	-	−3.97	−3.20	2 hours

M1 means monomial integration with 2N nodes; Q(1) means quadrature integration with one node in each dimension; Mean and Max are mean and maximum unit-free Euler equation errors in log10 units, respectively; CPU is running time.

A new Keynesian (NK) model

Assumptions:

- *Households* choose consumption and labor.
- Perfectly competitive *final-good firms* produce goods using intermediate goods.
- Monopolistic *intermediate-good firms* produce goods using labor and are subject to sticky price (á la Calvo, 1983).
- *Monetary authority* obeys a Taylor rule with zero lower bound (ZLB).
- *Government* finances a stochastic stream of public consumption by levying lump-sum taxes and by issuing nominal debt.
- *6 exogenous shocks and 8 state variables* $\Longrightarrow$ The model is large scale (it is expensive to solve or even intractable under conventional global solution methods that rely on product rules).

Summary of equilibrium conditions

We have

- Stochastic processes for 6 exogenous shocks.
- 8 endogenous equilibrium equations & 8 unknowns.
- 2 endogenous state variables, price dispersion and interest rate.
- Thus, there are 8 (endogenous plus exogenous) state variables.

Computational papers on ZLB

How to impose the ZLB on interest rate?

- Perturbation methods do not allow us to impose the ZLB in the solution procedure.
- The conventional approach in the literature is to disregard the ZLB when computing perturbation solutions and to impose the ZLB in simulations when running accuracy checks (that is, whenever R_t happens to be smaller than 1 in simulation, we set it at 1).
  - Christiano, Eichenbaum&Rebelo (2009)
- **In contrast,** our global EDS method does allow to impose the ZLB both in the solution and simulation procedures.

Parameter values

We calibrate the model using the results in Smets and Wouters (2003, 2007), and Del Negro, Smets and Wouters (2007).

- Preferences: $\gamma = 1$; $\vartheta = 2.09$; $\beta = 0.99$
- Intermediate-good production: $\varepsilon = 4.45$
- Fraction of firms that cannot change price: $\theta = 0.83$
- Taylor rule: $\phi_y = 0.07$; $\phi_\pi = 2.21$; $\mu = 0.82$
- Inflation target: $\pi_* \in \{1, 1.0598\}$
- Government to output ratio: $\overline{G} = 0.23$
- Stochastic processes for shocks:
  $\rho_u = 0.92$; $\rho_L = 0.25$; $\rho_B = 0.22$; $\rho_a = 0.95$; $\rho_R = 0.15$; $\rho_G = 0.95$
  $\sigma_u = 0.54\%$; $\sigma_L \in \{18.21\%, 40.54\%\}$; $\sigma_B = 0.23\%$; $\sigma_a = 0.45\%$;
  $\sigma_R = 0.28\%$; $\sigma_G = 0.38\%$

We compute 1st and 2nd perturbation solutions using Dynare, and we compute 2nd and 3rd degree EDS solutions.

Time-series solution and EDS grid

Figure 5. Simulated points and the grid for a new Keynesian model: ZLB is imposed

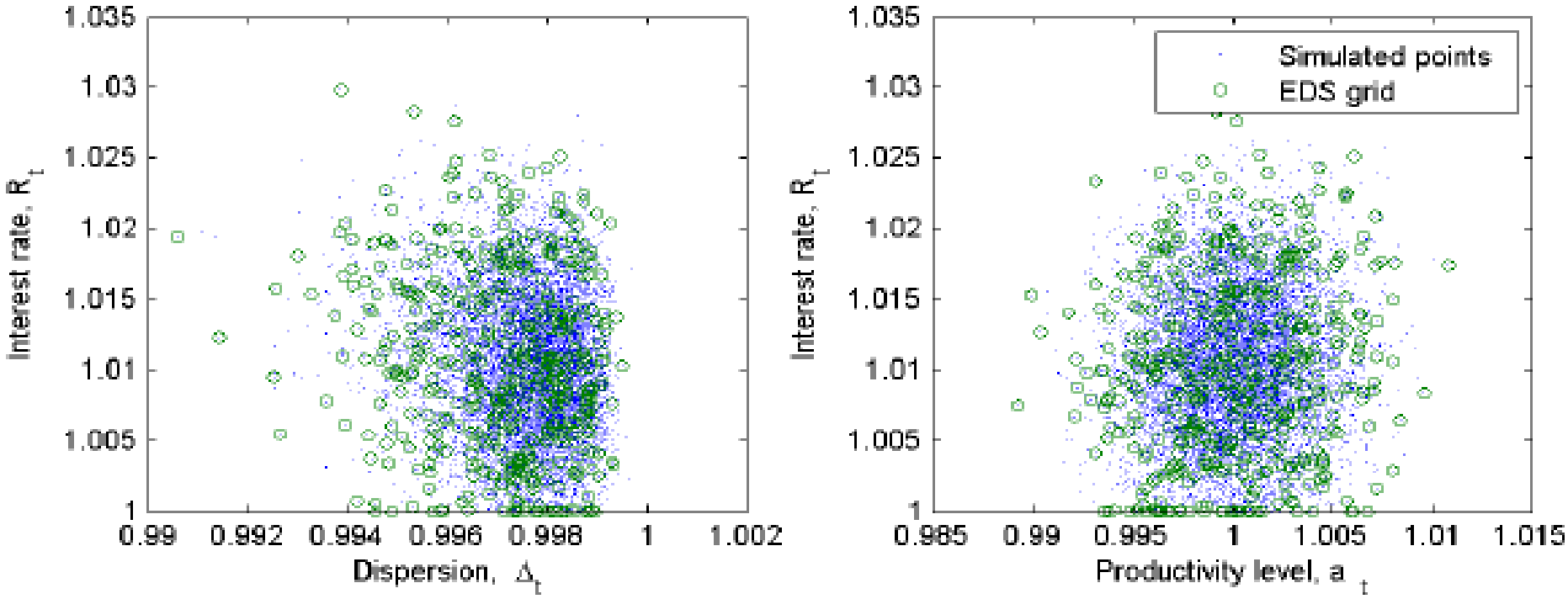


Table 4. Accuracy and speed in the NK model with 0% inflation target and 18.21% volatility of labor shock

	PER1	PER2	EDS2	EDS3
CPU	0.15		24.3	4.4
Mean	-3.03	-3.77	-3.99	-4.86
Max	-1.21	-1.64	-2.02	-2.73
R_{min}	0.9916	0.9929	0.9931	0.9927
R_{max}	1.0340	1.0364	1.0356	1.0358
$Fr_{(R\leq 1)}, \%$	2.07	1.43	1.69	1.68
$\triangle R, \%$	0.17	0.09	0.05	0
$\triangle C, \%$	1.00	0.19	0.12	0
$\triangle Y, \%$	1.00	0.19	0.12	0
$\triangle L, \%$	0.65	0.33	0.16	0
$\triangle \pi, \%$	0.30	0.16	0.11	0

PER 1 and PER 2 = 1st and 2nd order Dynare solutions; EDS2 and EDS3 = 2nd and 3rd degree EDS algorithm; Mean and Max = average and max absolute errors (in log10 units); R_{min} and R_{max} = min and max R; Fr = frequency of $R \leq 1$; $\triangle X$ = max difference from EDS3.

Table 5. Accuracy and speed in the NK model with 5.98% inflation target and 40.54% volatility of labor shock

	PER1	PER2	EDS2	EDS3
CPU	0.15		22.1	12.0
Mean	-2.52	-2.90	-3.43	-4.00
Max	-0.59	-0.42	-1.31	-1.91
R_{min}	1.0014	1.0065	1.0060	1.0060
$R_{\max}$	1.0615	1.0694	1.0653	1.0660
$Fr_{(R\leq 1)}, \%$	0	0	0	0
$\triangle R, \%$	0.63	0.39	0.25	0
$\triangle C, \%$	6.57	1.49	0.72	0
$\triangle Y, \%$	6.57	1.48	0.72	0
$\triangle L, \%$	3.16	1.30	0.54	0
$\triangle \pi, \%$	1.05	0.79	0.60	0

PER 1 and PER 2 = 1st and 2nd order Dynare solutions; EDS2 and EDS3 = 2nd and 3rd degree EDS; Mean and Max = average and max absolute errors (in log10 units); R_{min} and R_{max} = min and max R; Fr = frequency of $R \leq 1$; $\triangle X$ = max difference from EDS3.

Table 6. Accuracy and speed in the NK model with 0% inflation target, 18.21% volatility of labor shock and ZLB

	PER1	PER2	EDS2	EDS3
CPU	0.15		21.4	3.58
Mean	-3.02	-3.72	-3.57	-3.65
Max	-1.21	-1.34	-1.58	-1.81
R_{min}	1.0000	1.0000	1.0000	1.0000
$R_{\max}$	1.0340	1.0364	1.0348	1.0374
$Fr_{(R\leq 1)}, \%$	1.76	1.19	2.46	2.23
$\triangle R, \%$	0.33	0.34	0.34	0
$\triangle C, \%$	4.31	3.65	2.26	0
$\triangle Y, \%$	4.33	3.65	2.26	0
$\triangle L, \%$	3.37	3.17	2.45	0
$\triangle \pi, \%$	1.17	1.39	0.79	0

PER 1 and PER 2 = 1st and 2nd order Dynare solutions; EDS2 and EDS3 = 2nd and 3rd degree EDS; Mean and Max = average and max absolute errors (in log10 units); R_{min} and R_{max} = min and max R; Fr = frequency of $R \leq 1$; $\triangle X$ = max difference from EDS3.

Simulated series: ZLB is not imposed versus ZLB is imposed

Figure 6a. A time-series solution to a new Keynesian model: ZLB is not imposed

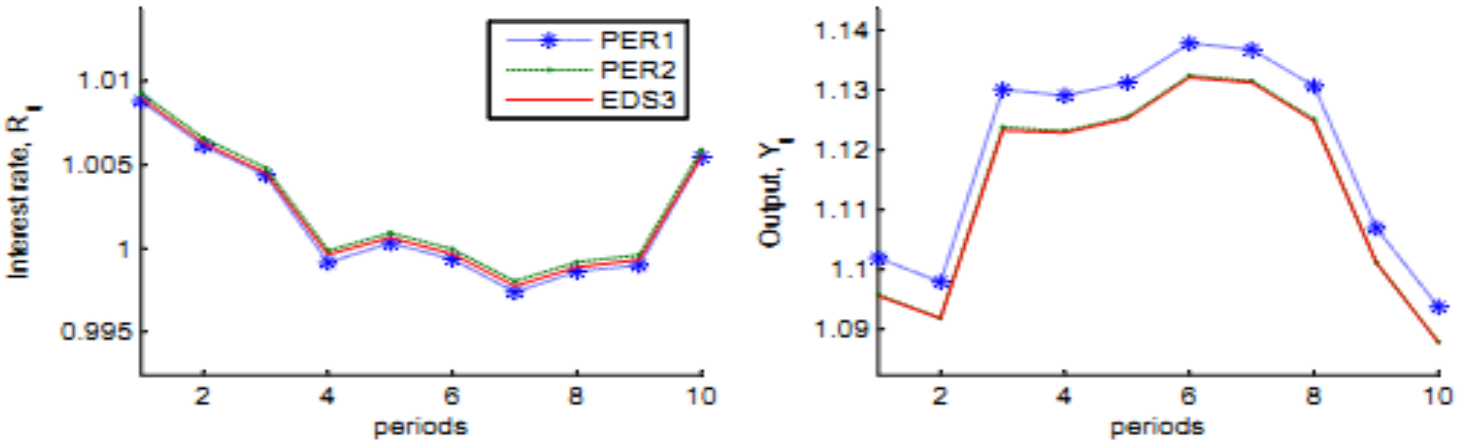
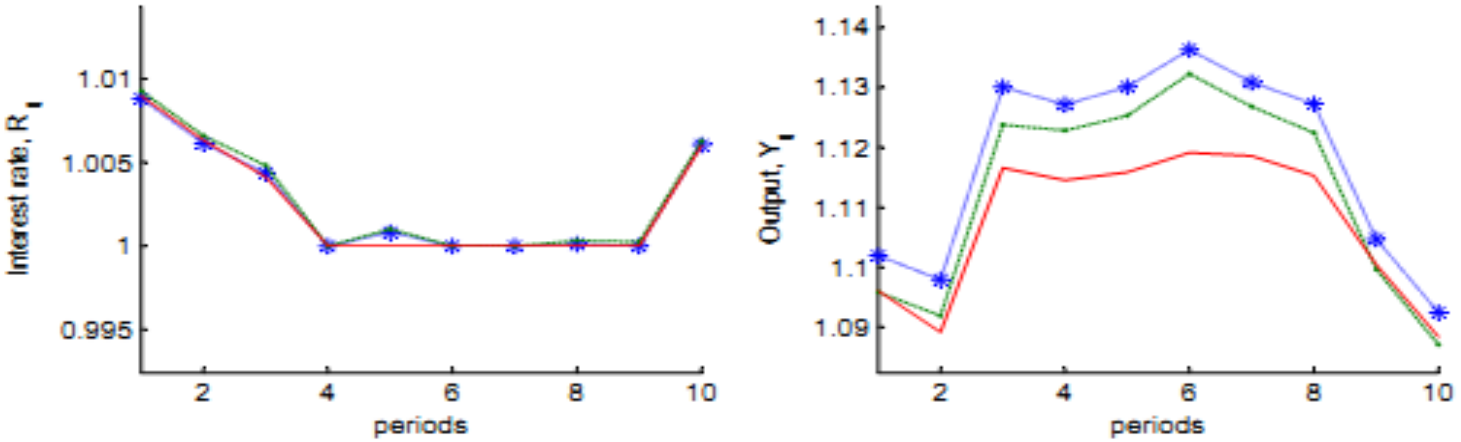


Figure 6b. A time-series solution to a new Keynesian model: ZLB is imposed



Application of the Cluster Grid Algorithm: ToTEM Model of the Bank of Canada

Large-scale models for policy analysis

Mostly used by central banks and government agencies:

- International Monetary Fund's Global Economy Model, GEM (Bayoumi et al., 2001);
- US Federal Reserve Board's SIGMA model (Erceg et al., 2006);
- Bank of Canada Terms of Trade Economic Model, ToTEM (Dorich et al. 2013);
- European Central Bank's New Area-Wide Model, NAWM (Coenen et al. 2008);
- Bank of England COMPASS model (Burgess et al., 2013);
- Swedish Riksbank's Ramses II model (Adolfson et al., 2013).

Requirements to central bank models

1. Central bank models must mimic as close as possible the actual economies in every possible dimension.
   Then, the policymakers can produce realistic simulation of alternative policies and to choose the best one.
2. Central bank models must be rich and flexible enough to describe interactions between many variables of interest, including different types of foreign and domestic consumption, investment, capital, labor, prices, exchange rate, etc.
   Central bank models may contain hundreds of equations.
   Their estimation, calibration, solution and simulation are highly nontrivial tasks.
3. Central bank models need DSGE models for policy analysis.
   Econometric models have limitations for policy analysis (Lucas critique).

Numerical methods used by central banks

- The central banks use linear (first-order) perturbation methods – Linear Taylor's expansions.
- *Advantages:*
  – computationally inexpensive;
  – simple to use;
  – can be applied to very large problems.
- *Drawbacks*:
  – insufficiently accurate in the presence of strong nonlinearities;
  – neglect second-order effects of the volatility of shocks on numerical solutions.
- Nonlinear effects can be economically significant; see Judd et al. (2017).

Questions addressed in the paper

1. How large could be the difference between local linear and global nonlinear solutions in realistically calibrated central banking models?

2. Could the limitations of the first-order perturbation analysis distort policy implications of realistic central banking models?

The answers to these questions are unknown as no one has computed nonlinear solutions to large-scale central banks' models.

Project of Bank of Canada

- To answer these questions, the Bank of Canada created a working group whose objective is to construct global solutions to their large-scale models.
- The results of this project are summarized in the form of a technical report and a research paper.
- Paper "Should Central Banks Worry about Nonlinearities of Their Large-Scale Macroeconomic Models?".
- It is presented by Vadym Lepetyuk on this conference:
  - June 28, 2017, 11:10 to 12:50, Session B02.

Full scale ToTEM model of Bank of Canada

- The Terms of Trade Economic Model (ToTEM) – the main projection and policy analysis model of the Bank of Canada.
- Small-open economy model.
- ToTEM includes 356 equations and unknowns

$\Longrightarrow$ It is too large for the existing global solution methods.

A scaled-down version of ToTEM

- We construct a scaled-down version of ToTEM, which we call a "baby ToTEM" (bToTEM) model.
- bToTEM includes 49 equations and unknowns
  $\Longrightarrow$ It is still a large-scale model.
- Production sectors: final-good production and commodity production.
- Trade: final goods, commodities, imports.
- One representative household, with differentiated labour services.
- Taylor-type interest rate rule.
- Six shocks, including exogenous ROW.

bToTEM: a serious challenge for global methods

- The models like bToTEM has not been yet studied in the literature.
- bToTEM contains 21 state variables (6 exogenous and 15 endogenous ones) $\Longrightarrow$ *curse of dimensionality*.
- Maliar and Maliar (2015) solve a new Keynesian model with 8 state variables (6 exogenous and 2 endogenous ones).
- The difference between 8 and 21 state variables is immense:
  - suppose we discretize each state variable into 10 grid points;
  - there are 10^8 and 10^{21} grid points, resp.;
  - this implies a 10^{13}-times difference in cost.
- Moreover, the system of bToTEM's equations is complex: it requires the use of numerical solvers.

A global solution by a cluster-grid algorithm

- Ergodic set method that uses an adaptive grid
  - the model is solved only in the area of the state space visited in simulation
  - Maliar and Maliar (2015)
- Merges stochastic simulation and projection approaches
  - simulation is used to identify and approximate the ergodic set
  - quadrature integration is used to accurately solve the model on a cluster grid
- Our integration and optimization methods are tractable in high-dimensional problems
  - non-product monomial integration
  - derivative-free solvers

Computer codes

"*GSSA_Two_Models.zip*" - *Generalized Stochastic Simulation Algorithm (GSSA):*

- *GSSA for one and multi-country growth models*

"*EDSCGA_Maliars_QE6_2015.zip*" - *Epsilon-distingushable set (EDS) and cluster-grid methods:*

- EDS and cluster grid methods for one and multi-country
- EDS method for a new Keynesian model with ZLB

- We have much faster code for the new Keynesian model (5 seconds!) - will provide it soon!

Precomputation of integrals

Introduction

What is precomputation in general?

- Precomputation = computation on initialization stage (Step 0), i.e., outside the main iterative cycle.
- Precomputation saves on cost because we make computations up-front rather than on each iteration.
- Precomputation increases accuracy because some integrals can be
- computed analytically.
- Thus, we will not present a new solution method but a technique that can be used in the context of existing methods to reduce their computational expense.

Numerical approximation of conditional expectations

- Solving dynamic economic models involves numerical approximation of conditional expectations:
  - Bellman equation: $V(k,z) = \max_{k',c} \{u(c) + \beta E[V(k',z')]\}$;
  - Euler equation: $u_1(c) = \beta E[u_1(c')(1-\delta+z'f_1(k'))]$.
- Expectations are recomputed each time when we update decision function, i.e., after each iteration.
- Cost of evaluating expectations increases:
  - when the number of random variables increases *(because dimensionality of integrals increases)*;
  - when more accurate methods are used *(because the number of integration nodes increases)*;
  - when models become more complex *(because numerical solvers are used more intensively which involves additional evaluations of integrals)*.

Introduction

Precomputation of conditional expectations

We introduce a simple technique - **precomputation of integrals** - that approximates integrals at initial stage of the solution procedure:

- we parameterize integrand with a polynomial function whose basis functions are separable in endogenous and exogenous state variables (e.g., ordinary polynomials);
- outside the main iterative cycle, we construct integrals for any given endogenous state variables;
- in the main iterative cycle, the values of integrals can be immediately derived using our precomputation results.

Under this procedure, we compute expectations just once, at the very beginning and never again! *Effectively, this convert a stochastic problem into a deterministic problem with the correspondent reduction in cost.*

Integrals under ordinary polynomials: an illustration

- As an example, consider a complete first-degree ordinary polynomial,

$$\mathcal{P}\left(k, z; b\right) = b_0 + b_1 k + b_2 z,$$

where $b \equiv \left(b_0, b_1, b_2\right)$ is coefficients vector; k' is capital known at present and $z' = z^{\rho} \exp\left(\epsilon'\right)$ is shock with unknown random variable ϵ'.

- We can represent conditional expectation of $\mathcal{P}\left(k', z'; b\right)$ as follows

$$\begin{aligned} E\left[\mathcal{P}\left(k', z'; b\right)\right] &= E\left[b_0 + b_1 k' + b_2 z^{\rho} \exp\left(\epsilon'\right)\right] \\ &= b_0 + b_1 k' + b_2 z^{\rho} E\left[\exp\left(\epsilon'\right)\right] = \\ &= \theta_0 + \theta_1 k' + \theta_2 z^{\rho} \equiv \mathcal{P}\left(k', z^{\rho}; \theta\right), \end{aligned}$$

where $\theta \equiv \left(\theta_0, \theta_1, \theta_2\right)$ is a new coefficient vector $\theta_0 = b_0$, $\theta_1 = b_1$ and $\theta_2 = b_2 E\left[\exp\left(\epsilon'\right)\right]$.

- The integrals in θ can be computed up-front without solving the model (i.e., precomputed).

Integrals under ordinary polynomials: an illustration

- Hence, conditional expectation of a polynomial function is given by the same polynomial function but evaluated at a different coefficients vector, i.e.,

$$E\left[\mathcal{P}\left(k', z'; b\right)\right] = \mathcal{P}\left(k', z^{\rho}; \theta\right);$$

where $\theta_0 = b_0$, $\theta_1 = b_1$ and $\theta_2 = b_2 E\left[\exp\left(\epsilon'\right)\right]$.

- With this result, conditional expectation can be evaluated as follows.
  - outside the main iterative cycle, we precompute

$$\mathcal{I} = E\left[\exp\left(\epsilon'\right)\right] = \frac{1}{\sqrt{2\pi\sigma}} \int_{-\infty}^{+\infty} \exp\left(\epsilon'\right) \exp\left(-\frac{\left(\epsilon'\right)^2}{2\sigma^2}\right) d\epsilon';$$

  - inside the main iterative cycle, we use

$$E\left[\mathcal{P}\left(k', z'; \left(b_0, b_1, b_2\right)\right)\right] = \mathcal{P}\left(k', z^{\rho}; \left(b_0, b_1, b_2\mathcal{I}\right)\right).$$

- This analysis can be easily generalized (see the paper) to
  - higher order polynomials;
  - multivariate random variables;
  - piecewise approximating functions.

Analytical construction of integrals under precomputation

- Integrals $\mathcal{I}_i$ can be constructed analytically for the case of normally distributed shock $\varepsilon' \sim \mathcal{N}\left(0, \sigma^2\right)$,

$$\begin{aligned}\mathcal{I}_i = E\left[\exp\left(\varepsilon'\right)\right] &= \frac{1}{\sqrt{2\pi\sigma^2}} \int_{-\infty}^{+\infty} \exp\left(\varepsilon'\right) \exp\left(-\frac{\left(\varepsilon'\right)^2}{2\sigma^2}\right) d\varepsilon' \\ &= \frac{1}{\sqrt{2\pi\sigma^2}} \int_{-\infty}^{+\infty} \exp\left(-\frac{\left(\varepsilon' - \sigma^2\right)^2}{2\sigma^2}\right) \exp\left(\frac{\sigma^2}{2}\right) d\varepsilon' \\ &= \exp\left(\frac{\sigma^2}{2}\right),\end{aligned}$$

- Fact used:
  - $\int_{-\infty}^{+\infty} f(x)dx = 1$ for a density function f of a normally distributed variable x with mean $I_i\sigma^2$ and variance σ^2.
- Analytical construction of integrals allows us to compute the integrals exactly.

Precomputation of integrals in the Bellman equation

- The usual Bellman equation

$$
\begin{aligned}
V(k,z) &= \max_{k',c} \left\{ u(c) + \beta E\left[V\left(k',z'\right)\right]\right\} \\
\text{s.t. } k' &= (1-\delta)k + zf(k) - c, \\
\ln z' &= \rho \ln z + \epsilon', \qquad \epsilon' \sim \mathcal{N}\left(0,\sigma^2\right),
\end{aligned}
$$

- Bellman equation with precomputation of integrals

$$
\begin{aligned}
\widehat{V}(k,z;b) &\doteq \max_{k',c} \left\{ u(c) + \beta \widehat{V}\left(k',z^{\rho};\theta\right)\right\}, \\
\text{s.t. } k' &= (1-\delta)k + zf(k) - c, \\
\theta_i &= b_i \mathcal{I}_i, \qquad i = 0,1,...,n,
\end{aligned}
$$

where $\mathcal{I}_i$ are precomputed integrals.

Precomputation of integrals in the Euler equation

Precomputation of integrals in the Euler equation requires a change of variables

- Our precomputation technique assumes that the function we parameterize is the same as the function for which we compute the expectation.
- This was true for Bellman equation: we parameterize $V(k,z)$, and we compute $E[V(k',z')]$.
- However, this is not true for a Euler equation algorithm that typically parameterizes policy functions like $C(k,z)$ or $K(k,z)$ but needs to compute $E[u_1(c')(1-\delta+z'f_1(k'))]$.
- We need to re-write the Euler equation in the way that is suitable for precomputation, namely, to parameterize the variable $u_1(c')(1-\delta+z'f_1(k'))$.

Precomputation of integrals in the Euler equation

- The usual Euler equation

$$u_1(c) = \beta E\left[u_1\left(c'\right)\left(1-\delta+z'f_1\left(k'\right)\right)\right]$$

- Introduce a new variable $q \equiv u_1(c)\left[1-\delta+zf_1(k)\right]$.
- In terms of q and q', the Euler equation is

$$\frac{q}{1-\delta+zf_1(k)} = \beta E\left[q'\right].$$

If we approximate $q = Q(k,z)$, we have the same function under the expectation, $E\left[Q\left(k',z'\right)\right]$, as required for precomputation.

- Hence, we rewrite the Euler equation as

$$\frac{\widehat{Q}(k,z;b)}{1-\delta+zf'(k)} \doteq \beta\widehat{Q}\left(k',z^{\rho};\theta\right).$$

- Again, all the effect of uncertainty on the solution is compressed into a mapping between the vectors b and θ.

Generality of precomputation of integrals

- Precomputation of integrals works under very general assumptions and can be applied to any set of equations that contains conditional expectations, including the Bellman and Euler equations.
- Precomputation of integrals is possible under many polynomial families (ordinary, Chebyshev Hermite, etc) and essentially under any process for shocks.
- Precomputation of integrals is compatible with essentially all computational techniques used by existing global solution methods, including a variety of approximating functions, solution domains, integration rules, fitting methods and iterative schemes for finding unknown parameters of approximating functions.
- Given that we must approximate integrals just once, we can use very accurate integration methods that would be intractable inside an iterative cycle.

Representative-agent model: parameters choice

Production function: $f(k_t) = k_t^{\alpha}$ with $\alpha = 0.36$.
Utility function: $u(c_t) = \frac{c_t^{1-\gamma}-1}{1-\gamma}$ with $\gamma \in \left\{\frac{1}{5}, 1, 5\right\}$.
Process for shocks: $\ln z_{t+1} = \rho \ln z_t + \epsilon_{t+1}$ with $\rho = 0.95$ and $\sigma = 0.01$.
Discount factor: $\beta = 0.99$.
Depreciation rate: $\delta = 0.025$.
Accuracy is measured by an Euler-equation residual,

$$\mathcal{R}(k_i, z_i) \equiv E_i \left[\beta \frac{c_{i+1}^{-\gamma}}{c_i^{-\gamma}} \left(1 - \delta + \alpha \theta_{i+1} k_{i+1}^{\alpha-1}\right) \right] - 1.$$

Value function iteration

Parameterize the value function by a polynomial $V(\cdot) \approx \widehat{V}(\cdot; b)$:

$$\widehat{V}(k, z; b) = b_0 + b_1 k + b_2 z + ... + b_n z^L.$$

Step 0. *Precompute integrals and construct a mapping between* b *and* θ.
Construct a grid, $\{k_m, z_m\}_{m=1}^M$.
Step 1. Fix $b \equiv (b_0, b_1, b_2, ..., b_n)$. Given $\{k_m, z_m\}_{m=1}^M$ solve for $\{c_m\}_{m=1}^M$.
Step 2. Compute the expectation using numerical integration (quadrature integration or monomial rules)

$$V_m \equiv u(c_m) + \beta \widehat{V}\left(k_m', z_m'; \theta\right).$$

Regress V_m on $\left(1, k_m, z_m, k_m^2, z_m^2, ..., z_m^L\right) \Longrightarrow$ get $\widehat{b}$.
Step 3. Solve for the coefficients using damping,

$$b^{(j+1)} = (1 - \xi) b^{(j)} + \xi \widehat{b}, \quad \xi \in (0, 1).$$

Table 1. Value function iteration in the representative-agent model

Polynomial degree	no precomputation			precomputation		
	Mean	Max	CPU	Mean	Max	CPU
1st	-	-	-	-	-	-
2nd	-3.42	-3.14	28.87	-3.42	-3.14	17.29
3rd	-4.57	-4.06	43.94	-4.57	-4.06	26.97
4th	-5.46	-5.07	55.99	-5.46	-5.07	34.21
5th	-6.49	-6.01	73.78	-6.49	-6.01	46.42

Mean and Max are unit-free Euler equation errors in log10 units, e.g.,

- -4 means $10^{-4} = 0.0001 \quad (0.01\%)$;
- -4.5 means $10^{-4.5} = 0.0000316 \quad (0.00316\%)$.

Benchmark parameters: $\gamma = 1/3$, $\delta = 0.025$, $\rho = 0.95$, $\sigma = 0.01$.
In the paper, also consider $\gamma = 3$. Accuracy and speed are similar.

Euler equation algorithm

Parameterize the RHS of the Euler equation by a polynomial $\widehat{Q}(k, z; b)$,

$$\widehat{Q}(k, z; b) = b_0 + b_1 k + b_2 z + ... + b_n z^L$$

Step 0. *Precompute integrals and construct a mapping between* b *and* θ.
Construct a grid, $\{k_m, z_m\}_{m=1}^M$.
Step 1. Fix $b \equiv (b_0, b_1, b_2, ..., b_n)$. Given $\{q_m, z_m\}_{m=1}^M$ solve for $\{k_m, c_m\}_{m=1}^M$.
Step 2. Compute the expectation using numerical integration (quadrature integration or monomial rules)

$$\widehat{q}_m = \beta \widehat{Q}\left(k'_m, z_m^\rho; \theta\right)\left[1 - \delta + z f_1\left(k_m\right)\right].$$

Regress $\widehat{q}_m$ on $\left(1, k_m, z_m, k_m^2, z_m^2, ..., z_m^L\right) \Longrightarrow$ get $\widehat{b}$.
Step 3. Solve for the coefficients using damping,

$$b^{(j+1)} = (1 - \xi) b^{(j)} + \xi \widehat{b}, \quad \xi \in (0, 1).$$

Table 2. Euler equation method in the representative-agent model

Polynomial degree	no precomputation			precomputation		
	Mean	Max	CPU	Mean	Max	CPU
1st	-3.47	-3.13	3.00	-3.47	-3.13	0.63
2nd	-4.64	-4.10	15.49	-4.64	-4.10	2.77
3rd	-5.26	-5.06	18.09	-5.26	-5.06	3.09
4th	-6.37	-5.90	22.29	-6.37	-5.90	3.62
5th	-7.34	-6.92	25.53	-7.34	-6.92	4.25

Multicountry model

The planner maximizes a weighted sum of N countries' utility functions:

$$\max_{\left\{\left\{c_t^h, k_{t+1}^h\right\}_{h=1}^N\right\}_{t=0}^\infty} E_0 \sum_{h=1}^N v^h \left(\sum_{t=0}^\infty \beta^t u^h \left(c_t^h \right) \right)$$

subject to

$$\sum_{h=1}^N c_t^h + \sum_{h=1}^N k_{t+1}^h = \sum_{h=1}^N k_t^h \left(1-\delta\right) + \sum_{h=1}^N z_t^h f^h \left(k_t^h \right),$$

where v^h is country h's welfare weight.
Productivity of country h follows the process

$$\ln z_{t+1}^h = \rho \ln z_t^h + \epsilon_{t+1}^h,$$

where $\epsilon_{t+1}^h \equiv \varsigma_{t+1} + \varsigma_{t+1}^h$ with $\varsigma_{t+1} \sim \mathcal{N}\left(0, \sigma^2\right)$ is identical for all countries and $\varsigma_{t+1}^h \sim \mathcal{N}\left(0, \sigma^2\right)$ is country-specific.

Table 3. Accuracy and speed in the multi-country model

Polynomial degree	$N = 2$			$N = 20$			$N = 30$		
	Mean	Max	CPU	Mean	Max	CPU	Mean	Max	CPU
1st	-2.77	-1.81	61	-3.12	-2.09	152	-3.15	-2.08	221
2nd	-3.88	-2.61	223	-4.36	-3.26	3303	-4.22	-3.22	13543
3rd	-4.94	-3.55	382	-	-	-	-	-	-
4th	-6.05	-4.68	574	-	-	-	-	-	-
5th	-7.15	-5.79	738	-	-	-	-	-	-

Additional results

- We show that numerical integration methods become less accurate as the degree of uncertainty increases, i.e. the standard deviation of shock increases.
- We evaluate the gains from precomputation for other numerical methods: Endogenous Grid method of Carroll (2005), Envelope Condition method of Maliar and Maliar (2013).
- We show that precomputation simplifies construction of numerical solutions to more complex models such as the model with elastic labor supply.
- Precomputation is suitable for discrete shocks. In such case, the expectations are computed exactly both with and without precomputation and all the gains from precomputation come in terms of costs reduction.
- MATLAB codes are available online.

Conclusion

- Many existing solution methods in the literature rely on parametric functions that satisfy the assumption of separability used in the present paper.
- For such methods, we can precompute integrals in the stage of initialization.
- The resulting transformed stochastic problem has the same computational complexity as a similar deterministic problem.
- Our technique of precomputation of integrals is very general and can be applied to essentially any set of equations that contains conditional expectations.
- Precomputation of integrals can save programming efforts, reduce a computational burden and increase accuracy of solutions.
- It is of special value in computationally intense applications.

Computer codes

"*Precomputation_JMMT_QE_2016.zip*" - Precomputation of integrals (= get rid off expectations before solving the model) for

- Conventional value and policy iteration
- Envelope condition value and policy iteration
- Endogenous grid method
- Conventional Euler equation method
- Multi-country model
- Aiyagari (1994) model with discrete shocks

Part 2: Solution Methods for Time-Dependent Models

Lilia Maliar and Serguei Maliar

CEF 2017 Workshop

Motivation: Why nonstationary models?

Unbalanced growth in the U.S. data

- Growth patterns appear to be highly unbalanced. For example, over the 1963-1992 period (Krusell, Ohanian, Ríos-Rull, Violante 2000):
  – output and the stock of structures increased by a factor of two;
  – the stock of equipment increased by more than seven times;
  – the number of unskilled workers slightly decreased;
  – the number of skilled workers nearly doubled;
  – the price of equipment relative to consumption (structures) went down by more than four times;
  – the skill premium was roughly stationary.
- Moreover, the growth rates are not constant over time.
- *Question:* "Can a general-equilibrium macroeconomic model (e.g. with capital-skill complementarity) explain such unbalanced growth patterns?"
- *To answer*, one may need **a framework for analyzing nonstationary and unbalanced growth models**.

Motivation: Why nonstationary models? (cont.)

Monetary policy normalization

How to normalize the monetary policy after the end of the crisis?

- Great Recession: ZLB and unconventional monetary policies (forward guidance and quantitative easing).
- Normalizing = switching back to some Taylor rule.
  Janet L. Yellen (2015) "Normalizing monetary policy: prospects and perspectives".

Questions:

- Should the Fed normalize policy **now** or **later**?
- Should the Fed normalize policy **gradually** or **all at once**?
- Should the regime shift be **announced in advance**?
- Should the policy normalization be **time** or **state dependent**?

We need a coherent nonlinear framework for analyzing **time-dependent** *models with parameter changes.*

Motivation: Why nonstationary models? (cont.)

Other examples of nonstationary applications

- deterministic trends in the data (population growth, climate changes, etc.);
- different kinds of technological progress that augment productivity of different factors, e.g., directed technical change;
- an entry into a monetary union;
- nonrecurrent policy regime switches;
- deterministic seasonals;
- changes in the consumer's tastes and habits.

In such models, the optimal value and/or decision functions nontrivially change from one period to another.

Balanced growth models

Some nonstationary models can be **converted into stationary**, for example, a class of balanced growth models.

However, the class of balanced growth models is limited:

- *King, Plosser and Rebelo (1988)* show that the standard neoclassical growth model is consistent with balanced growth only under the assumption of labor augmenting technological progress and under some additional restrictions on u and f.
- If one deviates from their assumptions, the property of balanced growth does not survive.

$\Rightarrow$ **Our goal is to solve nonstationary models without relying on the existence of a balanced growth path.**

Naive, recurrent and nonrecurrent regime changes

How does the literature model regime changes?

- **Naive solution approaches**: when solving the model, agents believe that the current regime is permanent but in policy experiments (simulation), they face regime changes.
  - *Logically inconsistent and contradicts to rational expectations.*
- Literature on **regime switches** (e.g., Davig and Leeper (2008)) provides a logically consistent way of modeling unanticipated recurrent regime switches: e.g., two recurrent regimes that happen with some probability,
  - *High and low productivity states;*
  - *UK leaves the EU but hopefully will come back.*
- But some regime changes are **nonrecurrent** (hopefully) and can be **ranked** by welfare, e.g.,
  Slavery was abolished;
  Women got the right to vote, etc.

$\Rightarrow$ **Our goal is to model nonrecurrent, time dependent regime changes**

In this talk

The rest of the talk:

1. We will explain the methodology of analyzing parameter changes using a simple growth model.
2. We will test this methodology using a growth model with balanced growth.
3. We will show a collection of nonstationary growth models with:
   - capital augmenting technological progress;
   - anticipated regime switches;
   - parameter drifting;
   - time-varying volatility with a deterministic trend;
   - seasonal adjustments;
   - estimation and calibration of parameters in an unbalanced growth model using data on the U.S. economy.
4. We will solve and simulate non-stationary transitions in a stylized new Keynesian model.

A nonstationary Markov optimization problem

A nonstationary growth model

We now introduce nonstationary Markov environment into dynamic general equilibrium modeling paradigm:

$$\max_{\{c_t, k_{t+1}\}_{t=0}^{\infty}} E_0 \left[\sum_{t=0}^{\infty} \beta^t u_t \left(c_t \right) \right] \tag{1}$$

$$\text{s.t. } c_t + k_{t+1} = \left(1 - \delta \right) k_t + f_t \left(k_t, z_t \right), \tag{2}$$

$$z_{t+1} = \varphi_t \left(z_t, \varepsilon_{t+1} \right), \tag{3}$$

– $c_t \geq 0$ and $k_t \geq 0$ are consumption and capital, resp.;
– initial condition $\left(k_0, z_0 \right)$ is given;
– $u_t : \mathbb{R}_+ \to \mathbb{R}$ and $f_t : \mathbb{R}_+^2 \to \mathbb{R}_+$ and $\varphi_t : \mathbb{R}^2 \to \mathbb{R}$ are possibly time-varying utility function, production function and law of motion for exogenous state variable z_t, resp.;
– sequence of u_t, f_t and φ_t for $t \geq 0$ is known to the agent in period $t = 0$; ε_{t+1} is i.i.d;
– $\beta \in (0, 1) =$ discount factor; $\delta \in [0, 1]$ =depreciation rate; $E_t \left[\cdot \right] =$ operator of expectation.

Why cannot we solve a nonstationary model with conventional solution methods?

A stationary growth model (dynamic-programming formulation):

$$V(k,z) = \max_{c,k'} \left\{ u(c) + \beta E\left[V\left(k', z'\right)\right]\right\}$$

$$\text{s.t. } k' = (1-\delta)k + zf(k) - c,$$

$$\ln z' = \rho \ln z + \varepsilon', \qquad \varepsilon' \sim \mathcal{N}\left(0, \sigma^2\right).$$

An interior solution satisfies the Euler equation:

$$u'(c) = \beta E\left[u'\left(c'\right)\left(1 - \delta + z' f'\left(k'\right)\right)\right].$$

- Conventional solution methods: *either iterate on Bellman equation until a fixed-point V is found or iterate on Euler equation until a fixed-point decision function $k' = K(k,z)$ is found.*
- However, if u, f, ρ and σ are time-dependent, then $V_t(\cdot) \neq V_{t+1}(\cdot)$ and $K_t(\cdot) \neq K_{t+1}(\cdot)$, i.e., no fixed-point functions V and K.
- We need to construct a sequence (path) of time-dependent value functions $\left(V_0(\cdot), V_1(\cdot), ...\right)$, decision functions $\left(K_0(\cdot), K_1(\cdot), ...\right)$.

Stochastic environment (informally)

- In the paper, we distinguish between stochastic processes that are
  - Markov;
  - stationary;
  - have stationary transition,
- Informally, a *(first-order) Markov stochastic process* is a process such that probability of an event depends not on the entire history but on the most recent past.
- A *stationary stochastic process* is a process whose unconditional probability distribution is time-invariant.
- A *stochastic process with stationary transition* is a process whose conditional probability distribution is time-invariant.

Example of the stochastic process considered

Consider an AR(1) process with time-dependent ρ_t and σ_t:

$$z_{t+1} = \rho_t z_t + \sigma_t \varepsilon_{t+1}, \qquad \varepsilon_{t+1} \sim \mathcal{N}(0,1),$$

$\rho_t \in (-1,1)$ and $\sigma_t \in (0,\infty)$ are given at $t = 0$.

- The conditional distribution $z_{t+1} \sim \mathcal{N}\left(\rho_t \overline{z}_t, \sigma_t^2\right)$ depends only on the most recent past $z_t = \overline{z}_t$ and is independent of history $(z_t, ..., z_0)$. $\Longrightarrow$ *The process is Markov*.
- Since ρ_t and σ_t change over time, the conditional probability distribution $\mathcal{N}(\rho_t \overline{z}_t, \sigma_t)$ depends not only on state $z_t = \overline{z}_t$ but also on a specific period t. $\Longrightarrow$ *The transitions are nonstationary*.
- If $\rho_t = \rho$ and $\sigma_t = \sigma$ for all t, then the conditional probability distribution $\mathcal{N}(\rho \overline{z}_t, \sigma)$ depends only on state $z_t = \overline{z}_t$ but not on time. $\Longrightarrow$ *The transitions are stationary*.
- **Note:** *Stochastic process can have stationary transition but still be nonstationary because it's unit root or explosive, e.g.,* $z_{t+1} = \rho z_t + \sigma \varepsilon_{t+1}$, with $|\rho| > 1$ $\Rightarrow$ *We do not study these cases explicitly*

Assumptions on utility and production functions

- Assumptions on the *utility function* u_t for $t \geq 0$:
  - twice continuously differentiable;
  - strictly increasing;
  - strictly concave;
  - satisfies the Inada conditions.

- Assumptions on the *production function* f_t for $t \geq 0$:
  - twice continuously differentiable;
  - strictly increasing in capital;
  - concave in capital;
  - satisfies the Inada conditions.

Assumption of a bounded objective function

Define **a pure capital accumulation process** $\left\{k_t^{\max}\right\}_{t=0}^{\infty}$ by assuming $c_t = 0$ for all t, which for each history $h_t = (z_0, ..., z_t)$ leads to

$$k_{t+1}^{\max} = f_t\left(k_t^{\max}, z_t\right),$$

$k_0^{\max} \equiv k_0$. We assume that **the objective function is bounded**:

$$E_0\left[\sum_{t=0}^{\infty} \beta^t u_t\left(k_t^{\max}\right)\right] < \infty.$$

This assumption insures that the objective function is bounded so that its maximum exists. This assumption holds when

- u_t is bounded from above for all t, i.e., $u_t(c) < \infty$ for any $c \geq 0$;
- f_t is bounded from above for all t, i.e., $f_t(k, z_t) < \infty$ for any $k \geq 0$ and $z_t \in Z_t$;
- for economies with growth as long as $k_t^{\max}$ does not grow too fast.

Thus, we can solve models with nonvanishing eternal growth, provided that the objective function is bounded.

Feasible and optimal programs

- **A feasible program** for our nonstationary economy is a pair of adapted (i.e., measurable for all t) processes $\{c_t, k_t\}_{t=0}^{\infty}$ such that given initial condition k_0 and history $h_\infty = (\varepsilon_0, \varepsilon_1 ...)$, they satisfy $c_t \geq 0$, $k_t \geq 0$ and the budget constraint for all t.

- Let $\Im^\infty$ be a set of all feasible programs for given initial capital k_0 and given history $h_\infty = (\varepsilon_0, \varepsilon_1 ...)$.
- Let us introduce the concept of a solution of the studied model.
- A feasible program $\{c_t^\infty, k_t^\infty\}_{t=0}^{\infty} \in \Im^\infty$ is called **an optimal program** if

$$E_0 \left[\sum_{t=0}^{\infty} \beta^t \left\{ u_t \left(c_t^\infty \right) - u_t \left(c_t \right) \right\} \right] \geq 0$$

for every feasible process $\{c_t, k_t\}_{t=0}^{\infty} \in \Im^\infty$.

Extended function path framework

Introducing extended function path (EFP) framework

Extended function path (EFP) framework includes two steps.

- **Solving a T-period stationary economy:** Assume that in a very remote period T, the economy becomes stationary, i.e., the utility and production functions and the laws of motions for exogenous shocks are time invariant, i.e., $u_t = u$, $f_t = f$ and $\varphi_t = \varphi$ for all $t \geq T$:
  $\Rightarrow$ we *can solve for equilibrium using conventional methods for stationary models.*
- **Constructing a function path:** Using the T-period solution as terminal condition, iterate backward on optimality conditions to construct a sequence (path) of time-dependent value and decision functions $(V_0(\cdot), V_1(\cdot), ...)$ and/or $(K_0(\cdot), K_1(\cdot), ...)$.
  $\Rightarrow$ *this is like solving OLG models.*

Extended function path (EFP) framework

Step 0. Initialization. Choose some $T \gg \tau$ and construct a T-period stationary economy such that $u_t = u$, $f_t = f$ and $\varphi_t = \varphi$ for all $t \geq T$.

Step 1. Construct a stationary economy, i.e., find a stationary capital function K satisfying:

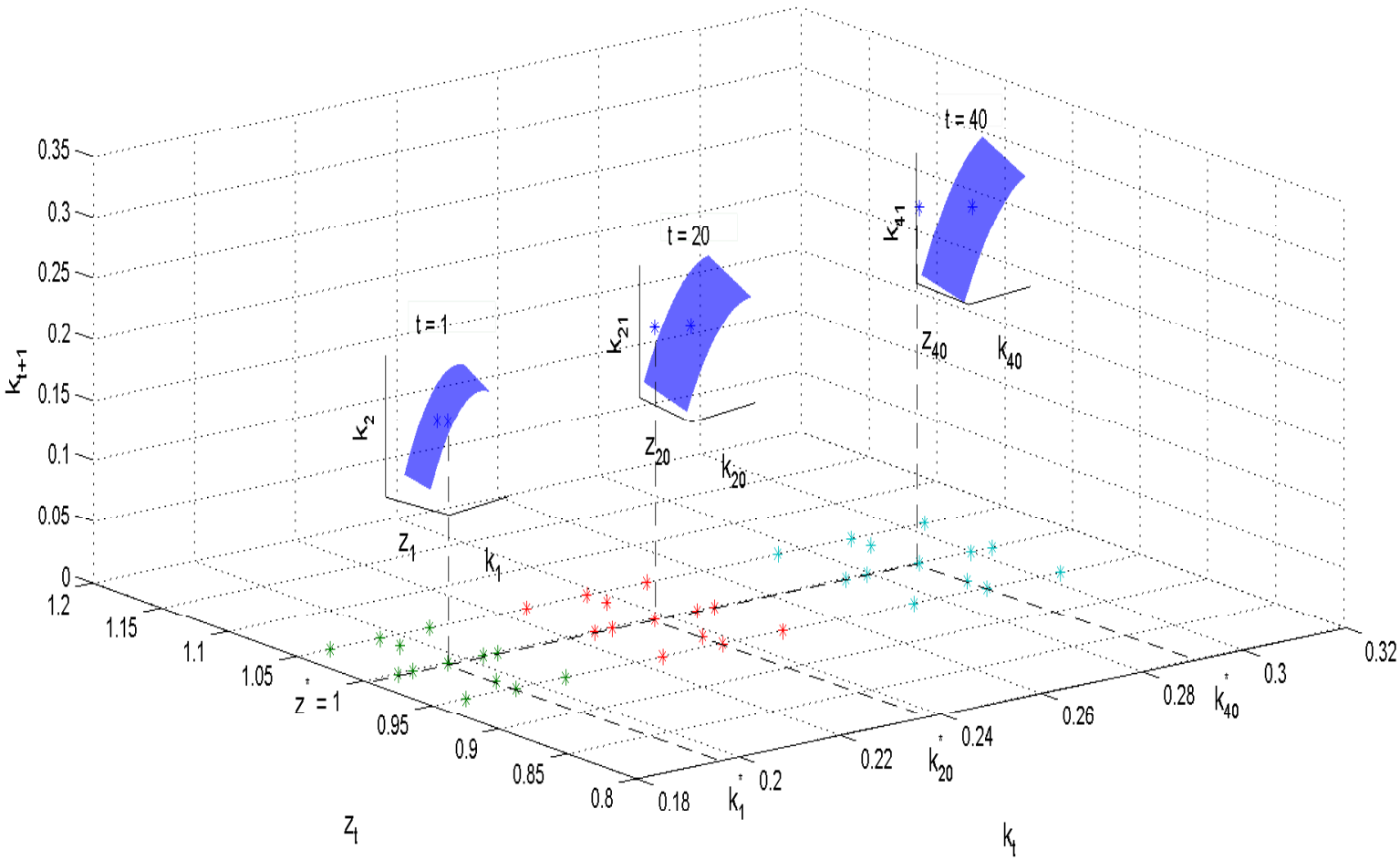
$$u'(c) = \beta E\left[u'(c')(1 - \delta + f'\left(k', \varphi\left(z, \varepsilon'\right)\right))\right]$$
$$c = (1 - \delta)\, k + f\left(k, z\right) - k'$$
$$c' = (1 - \delta)\, k' + f\left(k', \varphi\left(z, \varepsilon'\right)\right) - k''$$
$$k' = K\left(k, z\right) \text{ and } k'' = K\left(k', \varphi\left(z, \varepsilon'\right)\right).$$

Step 2. Construct a path for capital policy functions $\left(K_0, ..., K_T\right)$ that matches the terminal condition $K_T \equiv K$ and that satisfies for $t = 0, ...T - 1$:

$$u'_t(c_t) = \beta E_t\left[u'_{t+1}(c_{t+1})(1 - \delta + f'_{t+1}\left(k_{t+1}, \varphi_t\left(z_t, \varepsilon_{t+1}\right)\right))\right]$$
$$c_t = (1 - \delta)\, k_t + f_t\left(k_t, z_t\right) - k_{t+1}$$
$$c_{t+1} = (1 - \delta)\, k_{t+1} + f_{t+1}\left(k_{t+1}, \varphi_t\left(z_t, \varepsilon_{t+1}\right)\right) - k_{t+2}$$
$$k_{t+1} = K_t\left(k_t, z_t\right) \text{ and } k_{t+2} = K_{t+1}\left(k_{t+1}, \varphi_t\left(z_t, \varepsilon_{t+1}\right)\right).$$

Output: the first τ functions $\left(K_0, ..., K_\tau\right)$ constitute an approximate solution.

Example of function path constructed by EFP



Theoretical foundations of EFP framework

We provide theoretical foundations of the extended function path framework.

We prove two theorems:

- **Theorem 1 (existence)**: EFP approximations exists, is unique and possess a Markov structure.
- **Theorem 2 (turnpike)**: EFP can approximate a time-dependent solution to a nonstationary Markov model with an arbitrary degree of precision as the time horizon T increases.

Theorem 1: Nonstationary Markov program

Theorem 1 *(Optimal program of the T-period stationary economy). In the T-period stationary economy, the optimal program is given by a Markov process with possibly nonstationary transition probabilities.*

Proof. Under our assumptions on u_t, f_t and the objective function, FOCs are necessary for optimality. We will show that FOCs are also sufficient to identify the optimal program and to establish its Markov structure. Our proof is constructive: it relies on backward induction.

Step 1. – At T, the economy becomes stationary and remains stationary forever, i.e., $u_t \equiv u$, $f_t \equiv f$ and $\varphi_t \equiv \varphi$ for all $t \geq T$.
– Thus, the model's equations and decision functions are time invariant for $t \geq T$.
– It is well known that under our assumptions on u_t, f_t and the objective function, there is a unique stationary Markov capital function K that satisfies the optimality conditions.

Theorem 1: Nonstationary Markov program (cont.)

Step 2. Given the constructed T-period capital function $K_T \equiv K$, we define the capital functions $K_{T-1}, ..., K_0$ in previous periods by using backward induction. The Euler equation for period $T-1$,

$$u'_{T-1}(c_{T-1}) = \beta E_{T-1}\left[u'_T(c_T)(1-\delta+f'_T(k_T, z_T))\right],$$

where c_{T-1} and c_T are related to k_T and k_{T+1} in periods T and $T-1$ by

$$\begin{aligned} c_{T-1} &= (1-\delta)k_{T-1} + f_{T-1}(k_{T-1}, z_{T-1}) - k_T, \\ c_T &= (1-\delta)k_T + f_T(k_T, z_T) - k_{T+1}. \end{aligned}$$

– z_T follows a possibly nonstationary Markov process, i.e., $z_T = \varphi_T(z_{T-1}, \varepsilon)$.
– By construction, we have that $k_{T+1} = K_T(k_T, z_T)$ is Markov.
– Thus, we obtain a functional equation that defines k_T for each (k_{T-1}, z_{T-1}), i.e., the capital decisions at period $T-1$ are given by a state-contingent function $k_T = K_{T-1}(k_{T-1}, z_{T-1})$.
– By proceeding iteratively backward, we construct $K_{T-1}(k_{T-1}, z_{T-1}), ..., K_0(k_0, z_0)$. ■

Turnpike theorem

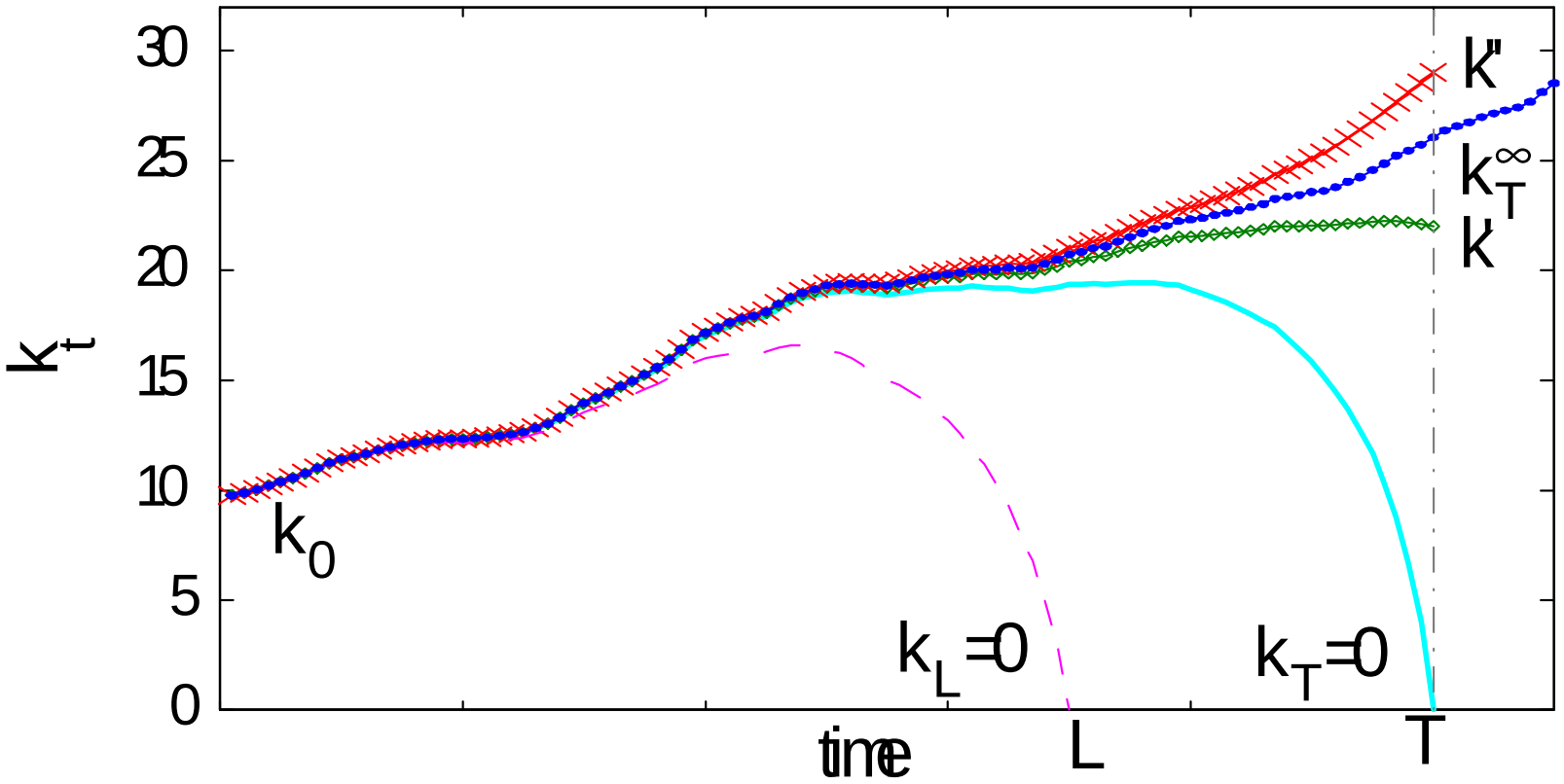
Turnpike $\equiv$ highway.

Turnpike theorem

- *Turnpike theorems*: turnpike is often the fastest route between two points which are far apart even if it is not a direct shortest route.
- Example: Driving from Los Angeles to San Francisco on highway 5.

Illustration of turnpike theorem

When you are young, you behave as if you will live forever...



Theorem 2 (turnpike theorem)

Let us fix history $h_\infty = (\varepsilon_0, \varepsilon_1 ...)$ and initial condition (k_0, z_0) and construct the productivity levels $\{z_t\}_{t=0}^T$ using the law of motion. Use the constructed functions $K_0(k_0, z_0), ..., K_T(k_T, z_T)$ to generate the optimal program $\{c_t^T, k_t^T\}_{t=0}^\infty$ for the T-period stationary economy

$$k_{t+1}^T = K_t\left(k_t^T, z_t\right),$$

where $k_0^T = k_0$, and c_t^T satisfies the budget constraint for all $t \geq 0$.

Theorem 2 *(Turnpike theorem): For any real number $\varepsilon > 0$ and any natural number τ, there exists a threshold terminal date $T(\varepsilon, \tau)$ such that for any $T \geq T(\varepsilon, \tau)$,* we have

$$\left|k_t^\infty - k_t^T\right| < \varepsilon, \quad \text{for all } t \leq \tau,$$

$\{c_t^\infty, k_t^\infty\}_{t=0}^\infty \in \Im^\infty$ = *optimal program in the nonstationary economy;*
$\{c_t^T, k_t^T\}_{t=0}^T$ = *optimal program in the T-period stationary economy.*

Remark on turnpike theorem

The convergence is uniform:

- Our turnpike theorem states that for all $T \geq T(\varepsilon, \tau)$, the constructed nonstationary Markov approximation $\{k_t^T\}$ is guaranteed to be within a given ε-accuracy range from the true solution $\{k_t^\infty\}$ during the initial τ periods.
- For periods $t > \tau$, our approximation may become insufficiently accurate and exit the ε-accuracy range.
- That is, the optimal program of the T-period stationary economy $\{k_t^T\}$
  – follows for a long time the optimal program of the nonstationary economy $\{k_t^\infty\}$ (turnpike),
  – and it deviates from turnpike only in the very last moment to meet a given terminal condition (the final destination off turnpike).

Proof of Turnpike theorem

The proof of Theorem 2 relies on three lemmas:

Lemma 1 The optimal program of the finite horizon economy with a zero terminal condition converges to the limit program.
Lemma 2 The optimal program of the T-period stationary economy converges to the same limit program.
Lemma 3 The limit program of the finite horizon economies with a zero terminal condition is optimal in the nonstationary infinite horizon economy.

$$\Downarrow \quad \text{(Theorem 2)}$$

The limit optimal program of the T-period stationary economy is optimal in the infinite horizon nonstationary economy as $T \longrightarrow \infty$.

Some related literature.

Earlier literature on nonstationary stochastic growth models

- Early literature on *stationary stochastic growth models*,
  - e.g., Brock and Gale (1969), Brock (1971), Brock and Mirman (1972, 1973), Mirman and Zilcha (1977), Brock and Majumdar (1978), Mitra and Zilcha (1981),
  - characterizes the properties of their solutions.
- Early literature on *nonstationary stochastic growth models*,
  - e.g., Majumdar and Zilcha (1987), Mitra and Nyarko (1991), Joshi (1997),
  - studies infinite-horizon, nonstationary economies similar to ours without assuming stationarity and Markov structure of the solutions;
  - but is limited to purely theoretical analysis and does not offer practical methods for constructing their nonstationary solutions in applications.
- *Our main contribution:*
  - distinguish a tractable class of nonstationary models;
  - propose a framework for studying quantitative implications of such models.

Extended path versus extended function path

EFP is related to an *"extended path"* (EP) method of Fair&Taylor (1983).

- Fair and Taylor (1983) EP method constructs a path for variables for **a larger time horizon** T than **the number of periods** τ for which an approximate solution is actually needed.
- In this respect, our EFP is similar to EP framework of Fair and Taylor (1983).
- By choosing sufficiently large T, both EFP and EP mitigate the effect of a specific terminal condition on the approximation during the initial τ periods.

In turn, the term *"path"* versus *"function path"* highlights the key difference between the EP and EFP methods:

- Fair and Taylor's (1983) EP method constructs **a path for variables** under the assumption of certainty equivalence.
- EFP method constructs **a path for decision functions** by approximating expectation functions accurately using accurate deterministic integration methods such as Gauss-Hermite quadrature

Implementation of EFP

- We implement EFP in the way that makes it tractable in complex and large-scale applications:
  - Smolyak sparse grids;
  - nonproduct monomial integration methods;
  - derivative-free solvers.
- Examples of MATLAB codes are provided in the authors' web pages.
- The running times for EFP can be reduced further if we use parallelization (our iteration, which is in line with Gauss-Jacobi method, is naturally parallelizable).

Testing EFP using a model with balanced growth

Assessing EFP accuracy in a model with balanced growth

- We assess the quality of approximations produced by EFP in the context of a model with balanced growth parameterized by

$$u_t(c) = \frac{c^{1-\gamma}}{1-\gamma}, \quad \text{and} \quad f_t(k,z) = zk^{\alpha}A_t^{1-\alpha},$$

  – $\gamma > 0$ and $\alpha \in (0,1)$;
  – $A_t = A_0 g_A^t$ = labor augmenting technological progress with an exogenous constant growth rate $g_A \geq 1$.
- Productivity is assumed to follow

$$\ln z_{t+1} = \rho \ln z_t + \sigma \varepsilon_{t+1}, \qquad \varepsilon_{t+1} \sim \mathcal{N}(0,1),$$

  $\rho \in (-1,1)$, $\sigma \in (0,\infty)$.
- This version of the model is consistent with balanced growth and can be converted into a stationary model; see King, Plosser and Rebelo (1988).
- We can solve the stationary model very accurately and use the accurate solution for comparison.

A comparison of four solution methods

We solve the nonstationary growth model using **four** alternative solution methods:

1. *"Exact solution"* is a very accurate solution to the stationary model with a balanced growth path produced by the conventional Smolyak method;
2. *"EFP solution"* is produced by the EFP method that solves the nonstationary model directly;
3. *"Naive solution"* is produced by replacing the nonstationary model with a sequence of stationary models, and it solves such models one by one.
   – it neglects a connection between the decision functions of different periods (unlike EFP);
4. *"Fair and Taylor solution"* is produced by using Fair and Taylor's (1983) method.

Fair and Taylor's (1983) extended path method

- Fair and Taylor's (1983) method relies on certainty equivalence for approximating expectation functions:

$$E_t\left[u'_{t+1}\left(c_{t+1}\right)\left(1-\delta+f'(k_{t+1},z_{t+1})\right)\right] \approx u'_{t+1}\left(c_{t+1}\right)\left(1-\delta+f'_{t+1}(k_{t+1},E_t\left[z_{t+1}\right])\right).$$

- To avoid explosive behavior, the method iterates on the economy's path at once in line with Gauss-Jacobi iteration.
- *Solution procedure:*
  (i) Guess the economy's path $\{k_1,...,k_{T+1}\}$;
  (ii) Substitute $\{k_1,...,k_{T+1}\}$ in the RHS of T Euler equations, resp., and obtain a new path $\{k_0,...,k_T\}$ in the LHS of T Euler equations;
  (iii) Iterate on the path until the convergence is achieved.

Parameterization

- For all experiments, we fix

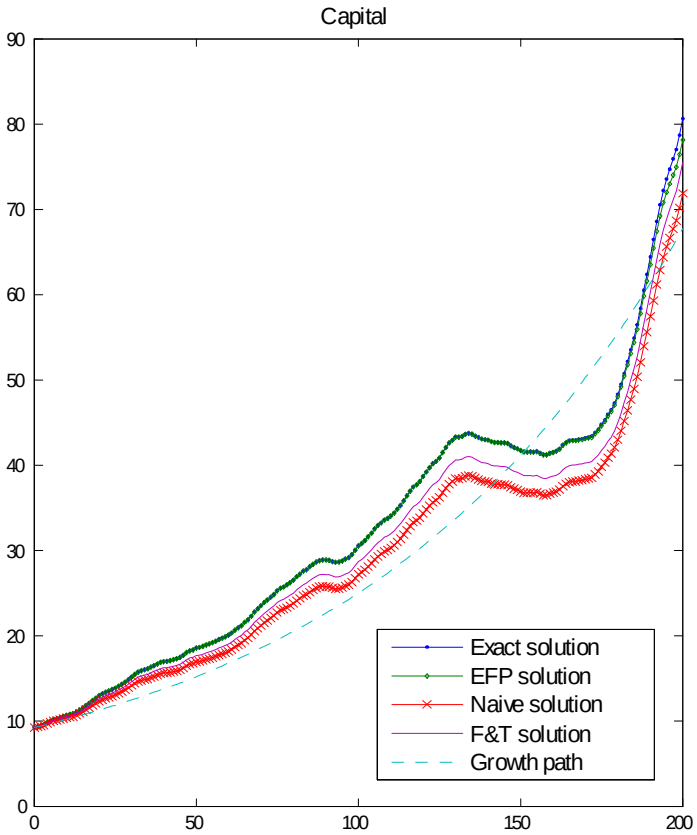
$$\alpha = 0.36, \quad \beta = 0.99, \quad \delta = 0.025, \quad \rho = 0.95.$$

- The remaining parameters are set in the benchmark case at

$$\gamma = 5, \quad \sigma_\varepsilon = 0.03, \quad g_A = 1.01, \quad T = 200.$$

  We vary these parameters across experiments.
- For all simulations, we use the same initial condition and the same sequence of productivity shocks.

Critical role of expectations in the accuracy of solutions



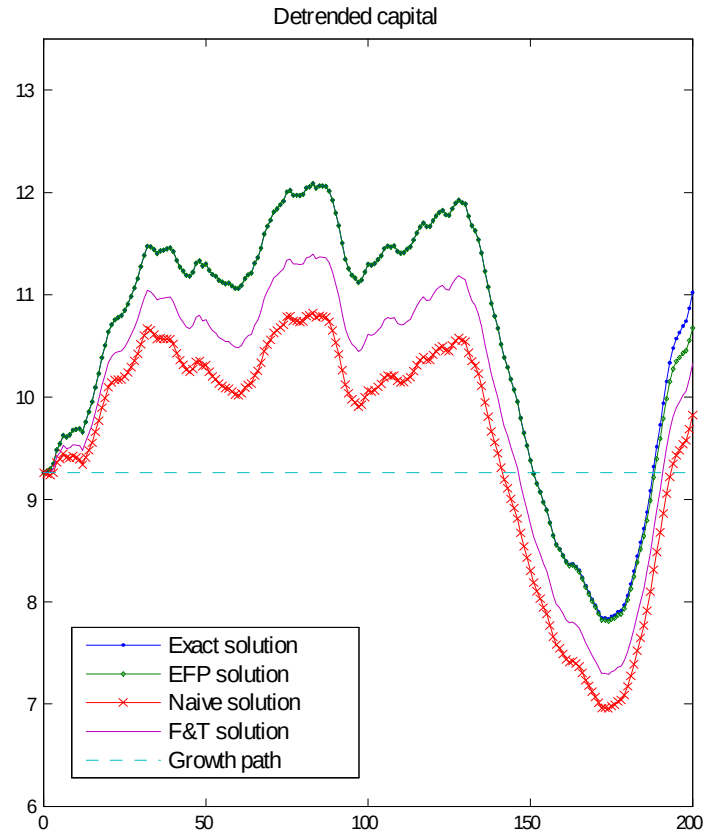


Table 1: comparison of four solution methods

	Fair-Taylor (1983) method, $\tau = 1$		Naive method	EFP method $\tau = 200$		
Terminal condition	Steady state	Steady state	-	Balanced growth	T-period stationary	
T	200	400	200	200	200	400
Maximum errors across t periods in $\log_{10}$ units						
$t \in [0, 50]$	-1.29	-1.29	-1.04	-6.82	-6.01	-6.42
$t \in [0, 100]$	-1.18	-1.18	-0.92	-6.68	-4.39	-5.99
$t \in [0, 150]$	-1.14	-1.14	-0.89	-6.66	-2.89	-5.98
$t \in [0, 175]$	-1.14	-1.13	-0.89	-6.66	-2.10	-5.98
$t \in [0, 200]$	-1.14	-1.13	-0.89	-6.66	-1.45	-5.92
Running time, in seconds						
Solution	1.2(+4)	6.1(+4)	28.9	104.9	99.1	225.9
Simulation	-	-	2.6	2.6	2.8	5.7
Total	1.2(+4)	6.1(+4)	31.5	107.6	101.9	231.6

Fair and Taylor's (1983) method

- The differences between the exact solution and Fair and Taylor's (1983) solution are around $10^{-1.6} \approx 2.5\%$ in Table 1.
- Fair and Taylor's (1983) method has relatively low accuracy because approximation of conditional expectation is inaccurate.
- Fair and Taylor's (1983) method is more accurate for models with a smaller variance of shocks and /or smaller degrees of nonlinearities.
- For example, we assess the difference between the exact solution and the Fair and Taylor's (1983) solutions for the model with $\gamma = 1$, $\sigma_\varepsilon = 0.01$, $g_A = 1.01$ and $T = 200$, and we found that such a difference is around 0.1% (this experiment is not reported).

Naive method

- The difference between the exact and naive solutions is about 10%.
- **Isn't surprising?** The naive method **does** take into account technology growth when constructing solutions.
  - it solves each t-period stationary model by assuming that today's and tomorrow's productivities are correctly given by $A_t = A_0 g_A^t$ and $A_{t+1} = A_0 g_A^{t+1}$.
- **Why does the naive method perform so poorly?**
  Because it is logically inconsistent:
  - agents are "unaware" of future permanent productivity growth;
  - they have expectations that are systematically more pessimistic than those of "aware" agents;
  - they are only confronted with parameter changes later, in simulations.
- Davig and Leeper (2009) address this problem by introducing rational expectations of regime switches.
- We address this problem in the context of nonstationary models.
- *Our conclusion:* approximating expectation functions accurately is critical for constructing accurate solutions to nonstationary models.

Terminal condition and the "tail" of simulation

- The exact and EFP solutions differ in the tail considerably; see Figure 3.
- This difference is especially well seen for the detrended time series (the right panel).
- The difference in the tail is due to the difference in the terminal conditions:
  - to construct the exact solution, we assume that the economy grows forever,
  - while to construct the EFP solution, we assume that it stops growing at T.
- If we use the same terminal conditions in both cases, then the EFP solution would be visually indistinguishable from the exact solution everywhere in the figure.
- Our turnpike theorem suggests a cheaper version of EFP in which we construct a longer function path, the EFP solution is very accurate everywhere including the tail.

Sensitivity analysis

Parameters	Model 1	Model 2	Model 3	Model 4	Model 5	Models 6	Model 7
γ	5	5	5	5	0.1	1	10
σ_ϵ	0.03	0.03	0.03	0.01	0.01	0.01	0.01
g_A	1.01	1.00	1.05	1.01	1.01	1.01	1.01
Maximum errors across t periods in $\log_{10}$ units							
$t \in [0, 50]$	-6.42	-6.31	-7.13	-6.66	-6.08	-6.24	-6.81
$t \in [0, 100]$	-5.99	-6.12	-7.05	-6.54	-5.97	-6.18	-6.36
$t \in [0, 150]$	-5.98	-6.04	-7.05	-6.52	-5.97	-6.18	-6.35
$t \in [0, 175]$	-5.98	-6.01	-7.05	-6.52	-5.97	-6.13	-6.33
$t \in [0, 200]$	-5.92	-5.99	-7.05	-6.51	-5.96	-5.88	-6.24
Running time, in seconds							
Solution	225.9	150.0	193.0	216.98	836.5	300.7	245.9
Simulation	5.6	5.7	5.8	5.66	5.6	5.6	5.7
Total	231.6	155.7	198.8	222.64	842.1	306.3	251.6

Nonstationary and Unbalanced Growth Applications

Application 1: model with capital augmenting technological progress

- Acemoglu (2002) argues that technical change may be directed toward different factors of production.
- Acemoglu (2003) explicitly incorporates capital augmenting technological progress into a deterministic model of endogenous technical change.
- However, the assumption of capital augmenting technological progress is inconsistent with a balanced growth path in the standard neoclassical stochastic growth model but only is the assumption of labor augmenting technological progress; see King, Plosser and Rebello (1988).
- We use EFP to solve a nonstationary growth model with capital augmenting technological progress that does not admit a stationary Markov equilibrium.

Application 1: model with capital augmenting technological progress (cont.)

- We assume a constant elasticity of substitution (CES) production function, and we allow for both labor and capital augmenting technological progresses,

$$F(k_t, \ell_t) = \left[\alpha(A_{k,t}k_t)^v + (1-\alpha)(A_{\ell,t}\ell_t)^v\right]^{1/v},$$

  – $A_{k,t} = A_{k,0}g_{A_k}^t$; $A_{\ell,t} = A_{\ell,0}g_{A_\ell}^t$; $v \leq 1$; $\alpha \in (0,1)$;
  – g_{A_k} and g_{A_ℓ} = rates of capital and labour augmenting technological progresses, resp.
- Labor is supplied inelastically. Let $\ell_t = 1$ for all t. The corresponding production function by $f(k_t) \equiv F(k_t, 1)$.
- The model with capital augmenting technological progress does not satisfy the assumptions in King, Plosser and Rebelo (1988) and does not admit a balanced growth path.

Parameterization for numerical experiments

- For numerical experiments, we assume:

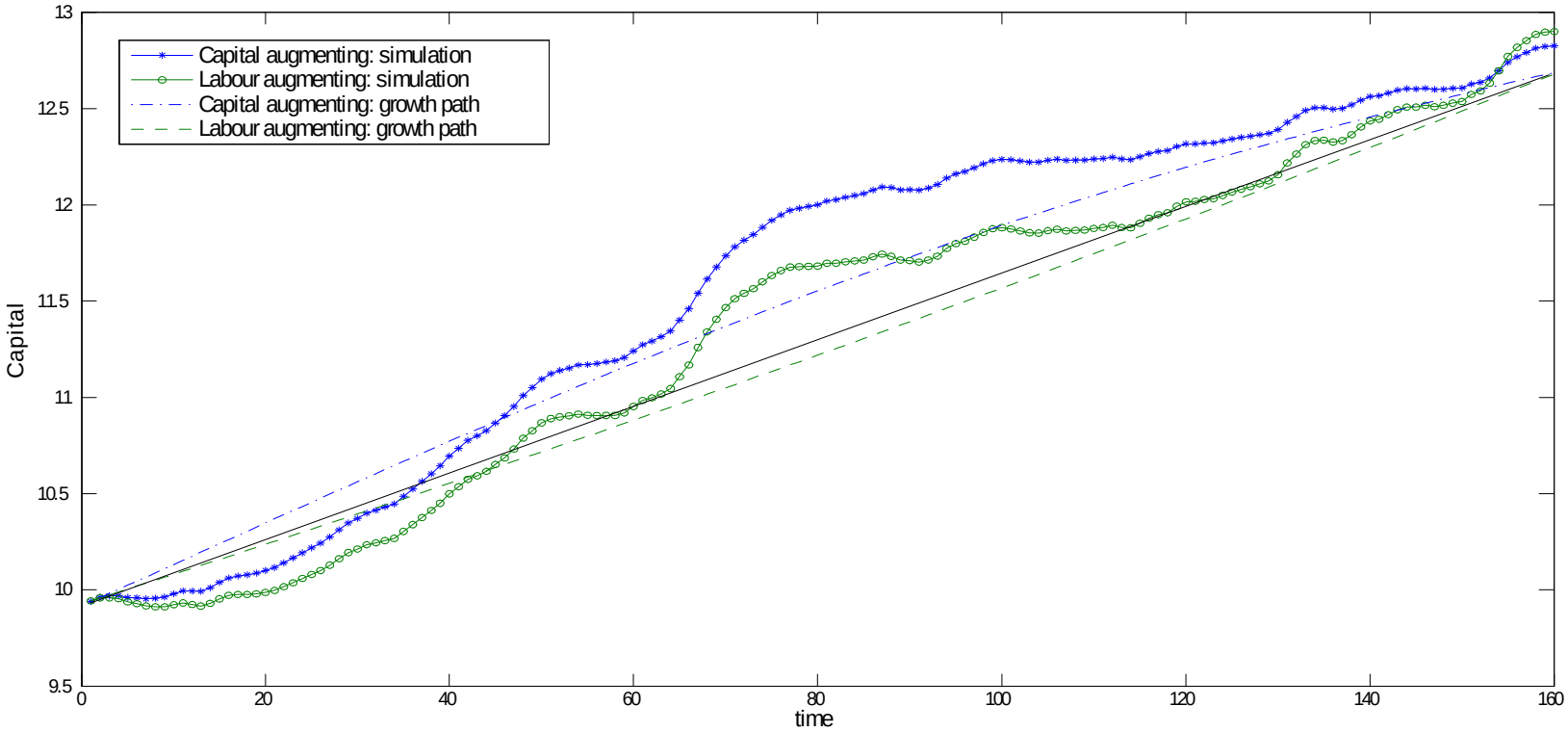
$$\begin{aligned} T &= 260, \quad \gamma = 1, \quad \alpha = 0.36, \quad \beta = 0.99, \\ \delta &= 0.025, \quad \rho = 0.95, \quad \sigma_\varepsilon = 0.01, \quad \nu = -0.42; \end{aligned}$$

  - the last value is taken in line with Antrás (2004) who estimated the elasticity of substitution between capital and labor to be in the range $[0.641, 0.892]$ that corresponds to $\nu \in [-0.12, -0.56]$.

- We solve two models:
  - the model with labor augmenting progress parameterized by $A_{\ell,0} = 1.1130$, $g_{A_\ell} = 1.00153$ and $A_{k,0} = g_{A_k} = 1$;
  - the model with capital augmenting progress parameterized by $A_{k,0} = 1$, $g_{A_k} = 0.9867$ and $A_{\ell,0} = g_{A_\ell} = 1$.

- For both models, $A_{\ell,0}$, g_{A_ℓ}, $A_{k,0}$, g_{A_k} are chosen to approximately match the initial and terminal capital stocks for time-series solutions of both models.

Capital versus labor augmenting technological progress



Application 2: A nonstationary model with a parameter shift

Recurrent regime changes

- Recent literature provides a logically consistent way of modeling unanticipated regime switches.
- Agents solve maximization problems in which regime changes are possible.
- Agents can adequately react to regime changes in simulation as implied by their decision functions;
  - e.g., *Sims and Zha (2006), Davig and Leeper (2007, 2009), Farmer, Waggoner, and Zha (2011), Foerster, Rubio-Ramírez, Waggoner and Zha (2013).*

Application 2: A nonstationary model with a parameter shift (cont.)

Recurrent unanticipated versus nonrecurrent anticipated regime changes

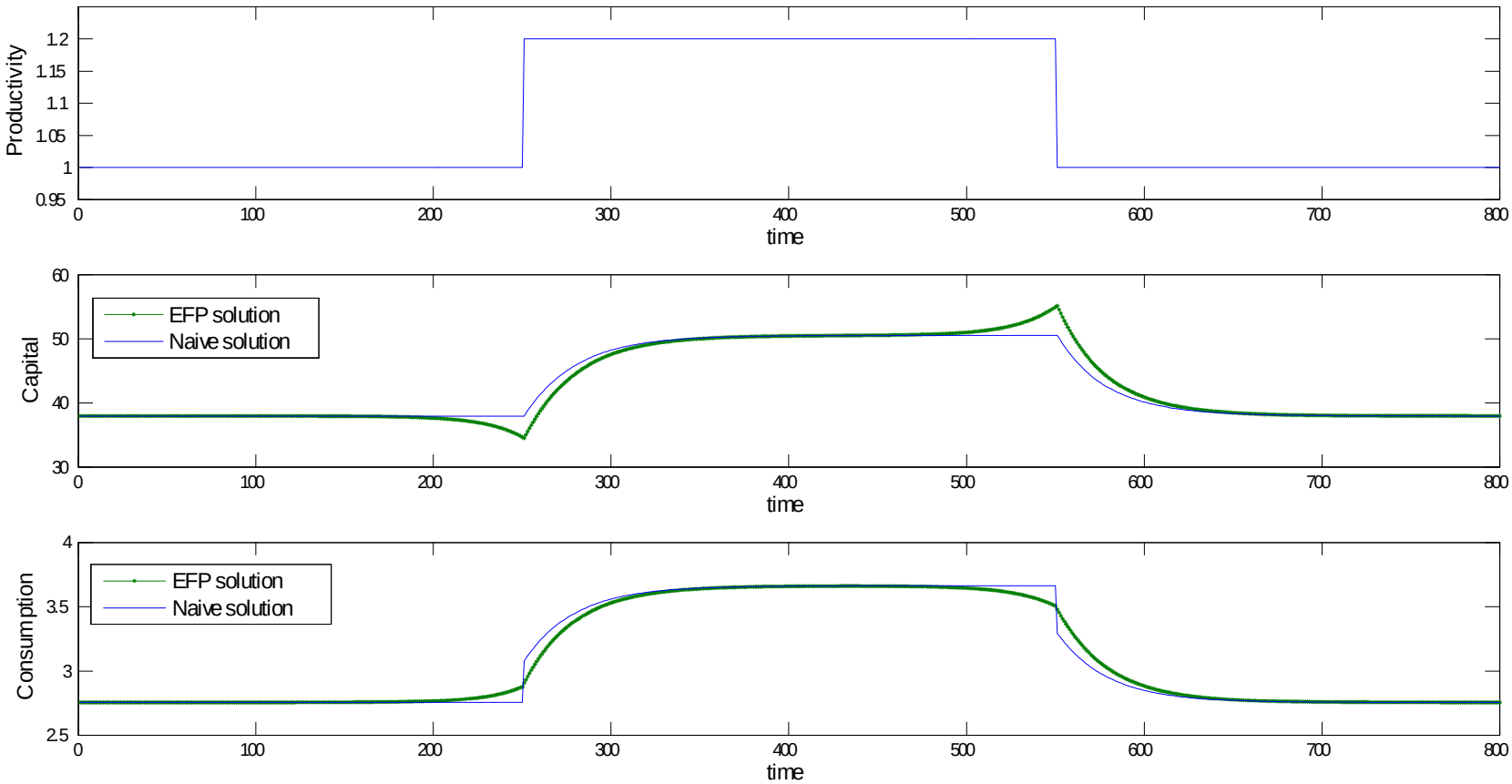
- However, there are real-world situations when regime shifts are nonrecurring and anticipated by agents in advance,
  - e.g., seasonal changes, presidential elections with anticipated outcome, forward-looking policy announcements, anticipated technological advances, etc.
- Recently, it is advocated in, e.g., Cochrane (1994), Beaudry and Portier (2006), Schmitt-Grohé and Uribe (2012).
- Anticipated accession of new members to the European Union; see Garmel, Maliar and Maliar (2008).

Application 2: A nonstationary model with a parameter shift (cont.)

Recurrent anticipated regime changes

- *Schmitt-Grohé and Uribe (2012)* study anticipated parameter shifts of fixed time horizons in the context of stationary Markov models;
  - the parameter shifts systematically occur, for example, each fourth or each eighth periods.
- However, the anticipated parameter shifts may be either nonrecurring and do not have fixed anticipation horizons.
- A distinctive feature of the EFP analysis is that we can solve a model with **any given sequence of anticipated nonrecurrent technology shocks.**

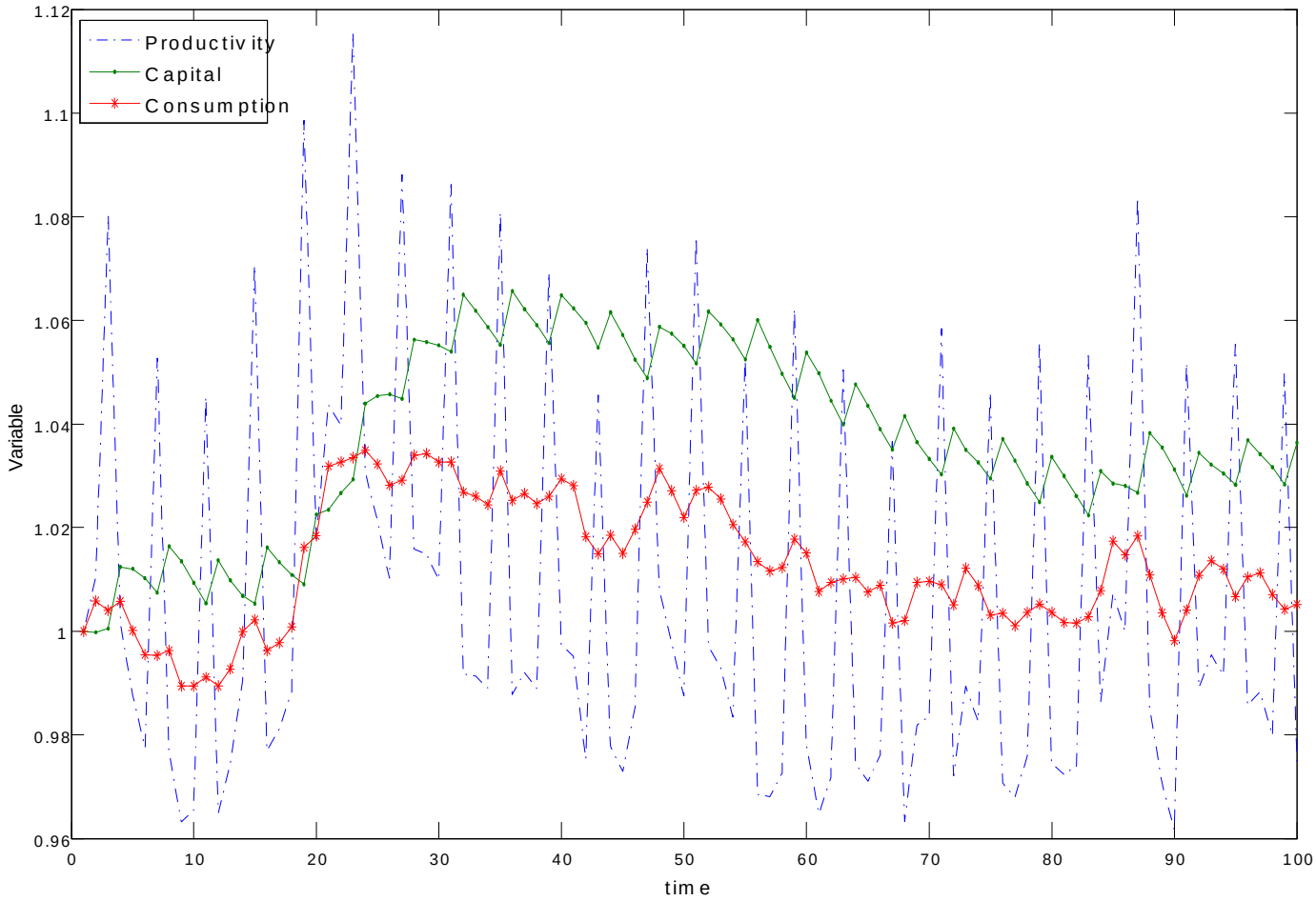
Application 2: Anticipated technology shocks



Application 2: A model with seasonal changes

- Seasonal adjustments are a special case of anticipated regime switches;
  - see Barsky and Miron (1989) for well documented evidence on the importance of seasonal changes for the business cycle.
- Hansen and Sargent (1993) and Christiano and Todd (2002) provide examples in which using seasonally adjusted data does not distort the business cycle analysis.
- Saijo (2013) argues that inadequate treatment of seasonal changes may lead to a significant bias in the parameter estimates.
- To investigate how seasonality interacts with other endogenous variables, it is important for macroeconomics to model the seasonal changes explicitly.

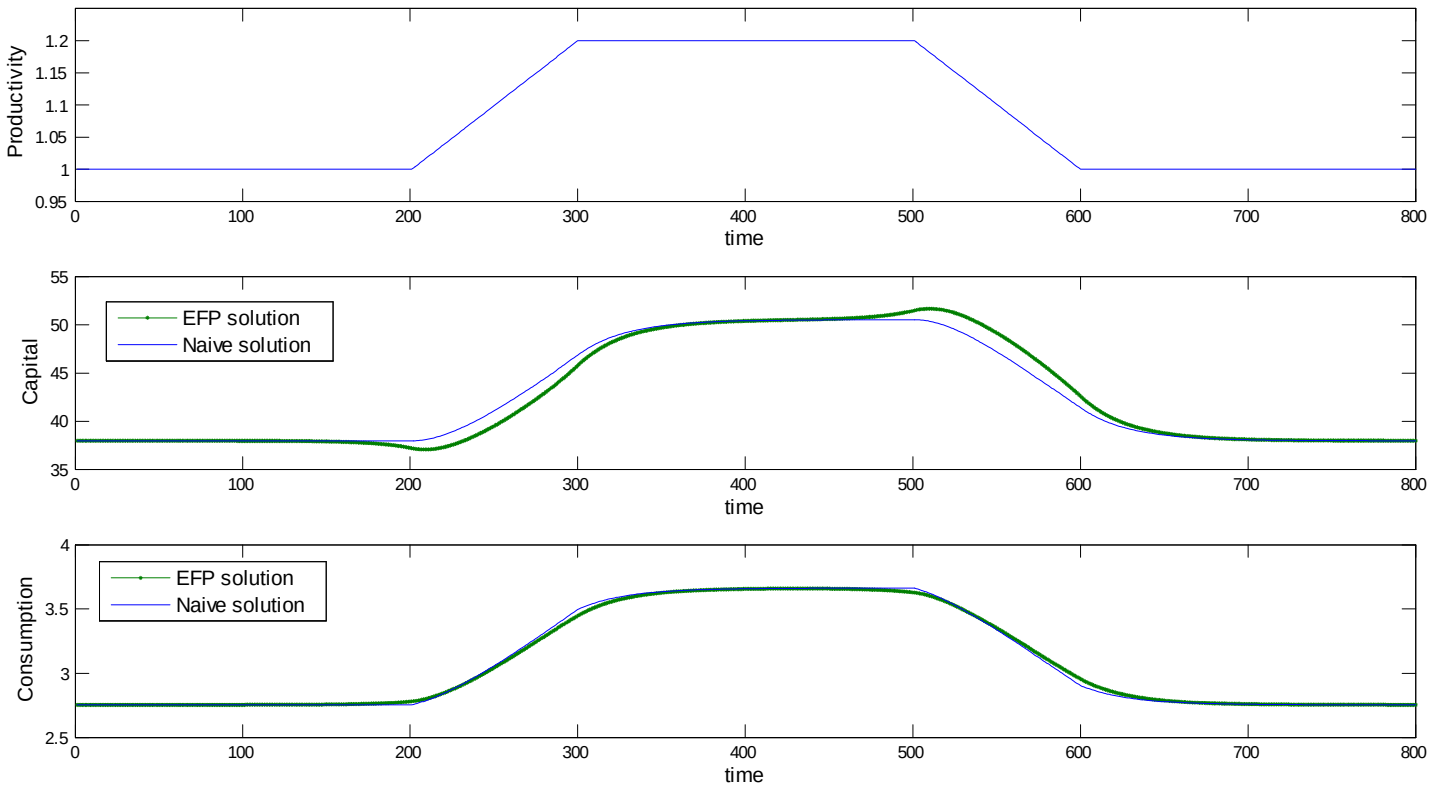
Application 2: A model with seasonal changes (cont.)



Application 3: A nonstationary model with a parameter drift

- Evidence in favor of parameter drifting;
  - e.g., Clarida, Galí and Gertler (2000), Lubik and Schorfheide (2004), Cogley and Sargent (2005), Galí (2006), Goodfriend and King (2009), Canova (2009).
- The literature assumes that parameters follow a stationary autoregressive process;
  - e.g., Fernández-Villaverde and Rubio-Ramírez (2007), Fernández-Villaverde, Guerrón-Quintana and Rubio-Ramírez (2010).
- However, if the parameters follow a **time-trend**, the equilibrium decision rules change each period and there is no stationary solution.

Application 3: A nonstationary model with a parameter drift (cont.)



Application 3: Diminishing volatility

- A large body of recent literature documents the importance of degree of uncertainty for the business cycle.
- This literature argues that volatility changes over time and models volatility (e.g., standard deviation of the productivity level) as a stochastic process or as a regime switch;
  - e.g., Bloom (2009), Fernández-Villaverde and Rubio-Ramírez (2010), Fernández-Villaverde, Guerrón-Quintana and Rubio-Ramírez (2010).
- The literature normally assumes that the standard deviation of exogenous shocks either follows a Markov process or experiences recurring Markov regime switches.
- In the latter case, volatility can be treated as an additional state variable, and in the former case, the regime is an additional state variable; in both cases, it is possible to cast the model with changing volatility into the conventional stationary framework.

Application 3: Diminishing volatility (cont.)

However, there is evidence that the volatility has a well pronounced time trend.

- Mc Connel and Pérez-Quiros (2000) document a *monotone* structural decline in the volatility of real GDP growth in the U.S. economy.
- Blanchard and Simon (2001) find a *nonmonotone* pattern of the decline in the U.S. GDP volatility: there was a steady decline in the volatility from the 1950s to 1970, then there was a stationary pattern and finally, there was another decline in the late 1980s and the 1990s.
- Stock and Watson (2003) find a *sharp reduction* in volatility of U.S. GDP growth in the first quarter of 1984.
- This kind of evidence cannot be reconciled in a model in which stochastic volatility follows a standard AR(1) process with stationary transitions.

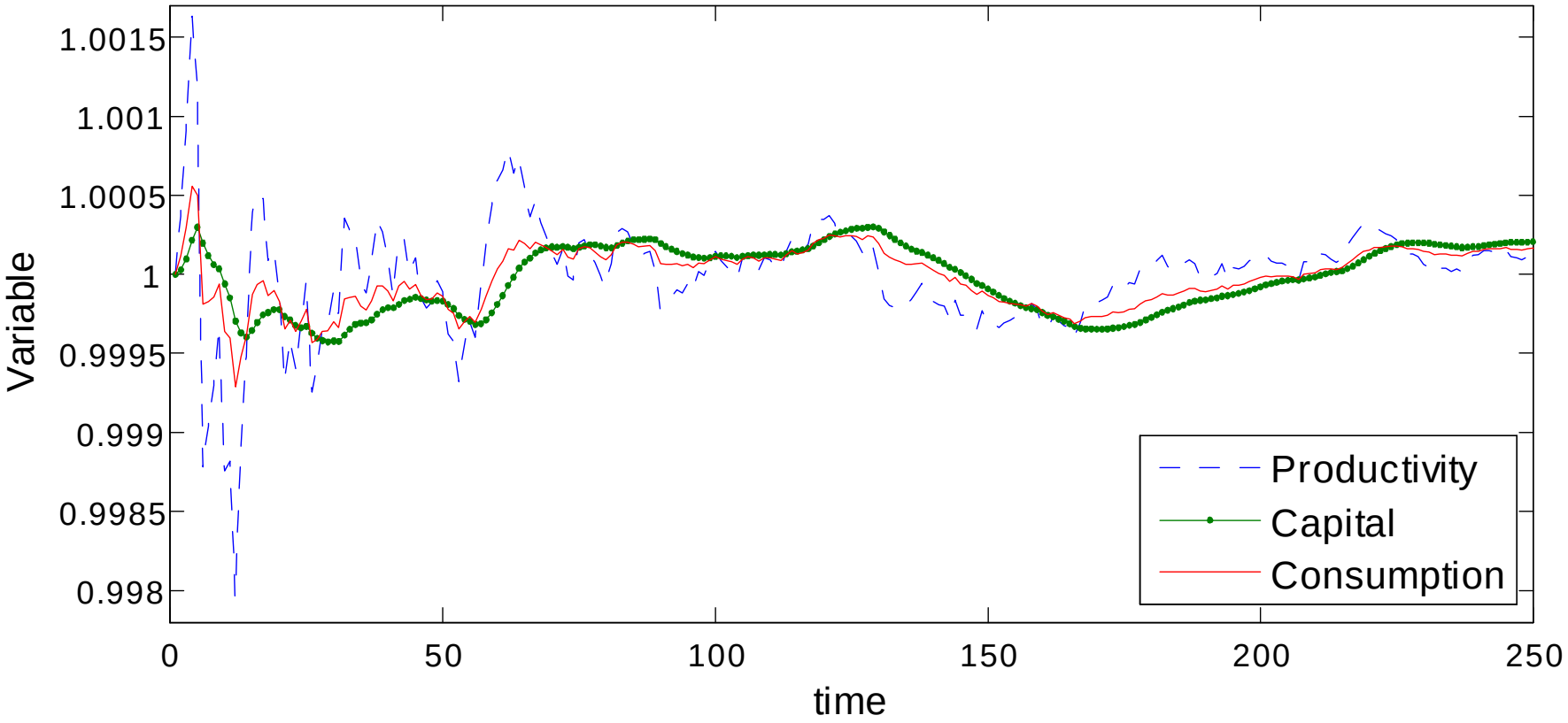
Application 3: Diminishing volatility (cont.)

- We consider a model in which the volatility has both a stochastic and deterministic components.
- We modify the standard neoclassical stochastic growth model to include a diminishing volatility of the productivity shock:

$$\ln z_t = \rho \ln z_{t-1} + \sigma_t \varepsilon_t, \qquad \sigma_t = \frac{B}{t^{\rho_\sigma}}, \quad \varepsilon_t \sim \mathcal{N}(0,1),$$

  - B = a scaling parameter;
  - ρ_σ = a parameter that governs the volatility of z_t.
- The standard deviation of the productivity shock $B\sigma / t^{\rho_\sigma}$ decreases over time, reaching zero in the limit, $\lim_{t\to\infty} \frac{B\sigma}{t^{\rho_\sigma}} = 0$.

Application 3: Diminishing volatility (cont.)



Application 4: Calibrating a growth model with a parameter drift to unbalanced U.S. data

- There is a large group of econometric methods that estimate and calibrate economic models by constructing numerical solutions explicitly, including
  - simulated method of moments (e.g., Canova (2007));
  - Bayesian estimation method (e.g., Smets and Wouters (2003), and Del Negro, Schorfheide, Smets and Wouters (2007));
  - maximum likelihood method (e.g., Fernández-Villaverde and Rubio-Ramírez (2007)).
- Normally, the related literature
  - imposes restrictions on the model that lead to a balanced growth path,
  - converts the model into stationary model,
  - solves it for stationary Markov equilibrium by using conventional methods.

Shortcomings of balanced growth models

However, there are two potential problems with this approach.

- First, the restrictions that are necessary to impose for balanced growth might not be the empirically-relevant ones.
  - For example, we might want to analyze a model with nonhomothetic utility and production functions, several kinds of technical progress and parameter shifts and drifts.
  - However, any deviation from the restrictions in King, Plosser and Rebelo (1988) destroys the property of balanced growth and hence, destroys the conventional Markov stationary equilibria.
- Second, the real world data are not always consistent with the assumption of balanced growth, in particular, different variables might grow at different and possibly time-varying rates.

We illustrate how EFP can be used to calibrate and estimate parameters in an unbalanced growth model by using the data on U.S. economy.

Time series to match

- We took macroeconomic data on the U.S. economy from the webpages of the Bureau of Economic Analysis and the Federal Reserve Bank of St. Louis.
- The sample spans over the period 1964:Q1 - 2011:Q4.
- Investment is defined as nonresidential and residential private fixed investment.
- Consumption is defined as a sum of nondurables and services.
- Capital is given by a sum of fixed assets and durables;
  - capital series are annual (in contrast to the other series which are quarterly);
  - we interpolate annual series of capital to get quarterly series using spline interpolation.
- Output is obtained as a sum of consumption and investment.
- We deflate the constructed variables with the corresponding implicit price deflator and we convert them in per capita terms.

The model with a depreciation rate drift

- While the constructed data are grossly consistent with Kaldor's (1961) facts, we still observe visible differences in growth rates across variables.
- We do not test whether or not such differences in growth rates are statistically significant but formulate and estimate an unbalanced growth model in which different variables can grow at differing rates.
- We extend the benchmark model to include time-varying depreciation rate of capital,

$$\max_{\{c_t,k_{t+1}\}_{t=0,\ldots,\infty}} E_0 \sum_{t=0}^{\infty} \beta^t u(c_t)$$

$$\text{s.t. } c_t + k_{t+1} = A_t z_t k_t^{\alpha} + (1 - d_t\delta_t)\, k_t,$$

$$\ln \delta_t = \rho_\delta \ln \delta_{t-1} + \varepsilon_{\delta,t}, \qquad \varepsilon_{\delta,t} \sim \mathcal{N}\left(0, \sigma^2_{\varepsilon_d}\right),$$

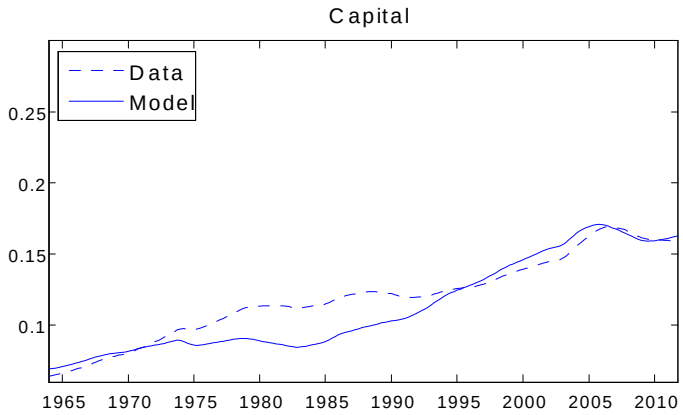
$$\ln z_t = \rho_z \ln z_{t-1} + \varepsilon_{z,t}, \qquad \varepsilon_{z,t} \sim \mathcal{N}\left(0, \sigma^2_{\varepsilon_z}\right),$$

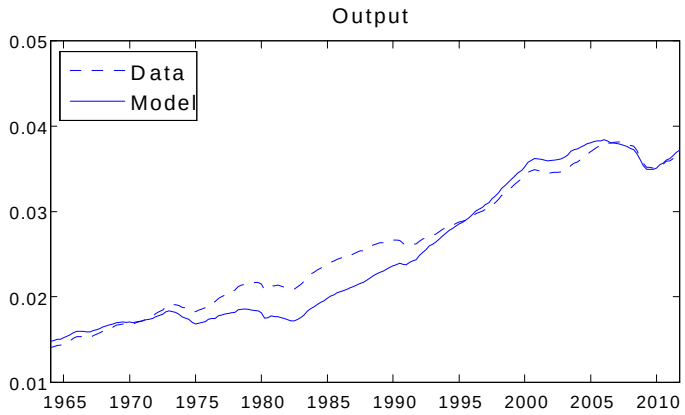
$d_t\delta_t$ = a time-varying depreciation rate; d_t = a trend component of depreciation, $d_t = d_0 g_d^t$; δ_t = a stochastic shock to depreciation.

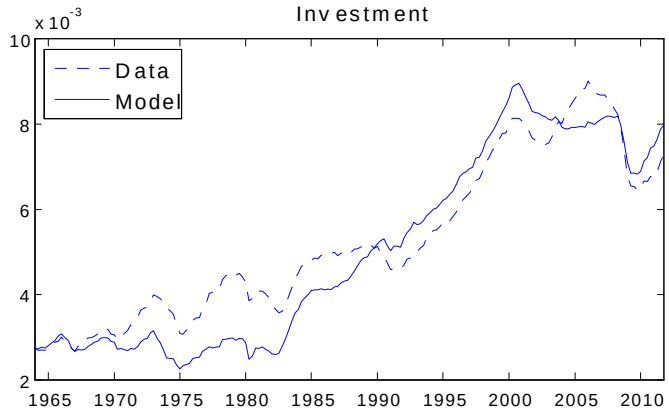
The model with a depreciation-rate drift

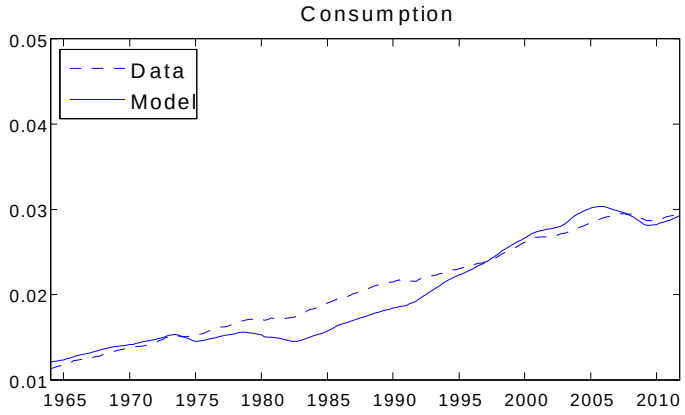
- Our assumption of a time trend in depreciation rate is consistent with the data of the Bureau of Economic Analysis.
- The aggregate depreciation rate changes over time because the composition of aggregate capital changes over time even if depreciation rates of each type of capital are constant; see Karabarbounis and Brent (2014).
- In turn, shocks to depreciation rate can result from the economic obsolescence of capital and are studied in, e.g., Liu, Waggoner and Zha (2011) and Gourio (2012).
- Gourio (2012) argues that a shock to the capital depreciation rate plays an important role in accounting for the business cycle fluctuations.

Fitted time series









Remarks on the estimation method

- The main goal of this application is not to advocate the role of time varying depreciation rate or some specific estimation and calibration technique.
- Rather, we would like to illustrate how estimation and calibration of the parameters can be carried out in the context of a nested fixed-point problem without assuming stationarity and balanced growth.
- Similar to the depreciation rate, we could have made all other parameters time dependent, including the discount factor β, the share of capital in production α and the parameters of the process for the productivity level.
- Furthermore, our simple estimation-calibration technique can be replaced by more sophisticated econometric techniques such as maximum likelihood, simulated method of moments, etc.

New Keynesian Model

Related literature

The questions of forward guidance and policy normalization are of interest to both practitioners and academia.

- Central banks and government agencies: *Williamson (2015), Orphanides (2015), Mendez and Murchinson (2015), Janet L. Yellen (2015), Kryvtsov and Mendez (2015), Engen, Laubach and Reifschneider (2015), etc.*
- Academic articles: *Carlstrom, Fuerst, Paustian (2012), Campbell, Evans, Fisher, Justiniano (2012), Caballero and Fahri (2014), Christiano, Eichenbaum and Trabandt (2014, 2015), del Negro, Giannoni, Patterson (2015), Di Maggio, Kermani and Palmer (2015), McKay, Nakamura and Steinsson (2015), Hills, Nakata, Schmidt (2016), Kaplan, Moll and Violante (2016), etc.*
- **Forward guidance puzzle**: *the effect of forward guidance is unrealistically large.*

A new Keynesian (NK) model

A stylized new Keynesian model with Calvo-type price frictions and a Taylor (1993) rule with the ZLB

- *Households* choose consumption and labor.
- Perfectly competitive *final-good firms* produce goods using intermediate goods.
- Monopolistic *intermediate-good firms* produce goods using labor and are subject to sticky price (á la Calvo, 1983).
- *Monetary authority* obeys a Taylor rule with zero lower bound (ZLB).
- *Government* finances a stochastic stream of public consumption by levying lump-sum taxes and by issuing nominal debt.
- *6 exogenous shocks and 8 state variables* $\Longrightarrow$ The model is large scale (it is expensive to solve or even intractable under conventional global solution methods that rely on product rules).
- This model is studied in "Merging Simulation and Projection Approaches to Solve High-Dimensional Problems with an Application to a New Keynesian Model" by Maliar and Maliar (QE, forthcoming).

The representative household

The utility-maximization problem:

$$\max_{\{C_t, L_t, B_t\}_{t=0,\ldots,\infty}} E_0 \sum_{t=0}^{\infty} \beta^t \exp\left(\eta_{u,t}\right) \left[\frac{C_t^{1-\gamma} - 1}{1-\gamma} - \exp\left(\eta_{L,t}\right) \frac{L_t^{1+\vartheta} - 1}{1+\vartheta} \right]$$

$$\text{s.t. } P_t C_t + \frac{B_t}{\exp\left(\eta_{B,t}\right) R_t} + T_t = B_{t-1} + W_t L_t + \Pi_t$$

where $\left(B_0, \eta_{u,0}, \eta_{L,0}, \eta_{B,0}\right)$ is given.
– C_t, L_t, and B_t = consumption, labor and nominal bond holdings, resp.;
– P_t, W_t and R_t = the commodity price, nominal wage and (gross) nominal interest rate, respectively;
– T_t = lump-sum taxes;
– Π_t = the profit of intermediate-good firms;
– β = discount factor; $\gamma > 0$ and $\vartheta > 0$.

The representative household

Stochastic processes for shocks

- $\eta_{u,t}$ and $\eta_{L,t}$ = exogenous preference shocks;
- $\eta_{B,t}$ = exogenous premium in the return to bonds;

$$\eta_{u,t+1} = \rho_u \eta_{u,t} + \epsilon_{u,t+1}, \qquad \epsilon_{u,t+1} \sim \mathcal{N}\left(0, \sigma_u^2\right)$$
$$\eta_{L,t+1} = \rho_L \eta_{L,t} + \epsilon_{L,t+1}, \qquad \epsilon_{L,t+1} \sim \mathcal{N}\left(0, \sigma_L^2\right)$$
$$\eta_{B,t+1} = \rho_B \eta_{B,t} + \epsilon_{B,t+1}, \qquad \epsilon_{B,t+1} \sim \mathcal{N}\left(0, \sigma_B^2\right)$$

Final-good producers

The profit-maximization problem:

- Perfectly competitive producers
- Use intermediate goods $i \in [0,1]$ as inputs

$$\max_{Y_t(i)} \quad P_t Y_t - \int_0^1 P_t(i) Y_t(i) di$$

$$\text{s.t. } Y_t = \left(\int_0^1 Y_t(i)^{\frac{\varepsilon-1}{\varepsilon}} di \right)^{\frac{\varepsilon}{\varepsilon-1}}, \qquad \varepsilon \geq 1 \tag{4}$$

– $Y_t(i)$ and $P_t(i)$ = quantity and price of an intermediate good i, resp.;
– Y_t and P_t = quantity and price of the final good, resp.;
– Eq (4) = production function (Dixit-Stiglitz aggregator function).

Result 1: Demand for the intermediate good i: $Y_t(i) = Y_t \left(\frac{P_t(i)}{P_t} \right)^{-\varepsilon}$.

Result 2: Aggregate price index $P_t = \left(\int_0^1 P_t(i)^{1-\varepsilon} di \right)^{\frac{1}{1-\varepsilon}}$.

Intermediate-good producers

The cost-minimization problem:

- Are monopolistically competitive
- Use labor as an input
- Are hit by a productiviy shock
- Are subject to sticky prices

$$\min_{L_t(i)} \quad \text{TC}\left(Y_t\left(i\right)\right) = \left(1-v\right) W_t L_t\left(i\right)$$

$$\text{s.t.} \quad Y_t\left(i\right) = \exp\left(\eta_{a,t}\right) L_t\left(i\right)$$

$$\eta_{a,t+1} = \rho_a \eta_{a,t} + \epsilon_{a,t+1}, \qquad \epsilon_{a,t+1} \sim \mathcal{N}\left(0, \sigma_a^2\right)$$

– TC = nominal total cost (net of government subsidy v);
– $L_t\left(i\right)$ = labor input;
– $\exp\left(\eta_{a,t}\right)$ is the productivity level.

Intermediate-good producers (price decisions)

Calvo-type price setting:
$1-\theta$ of the firms sets prices optimally, $P_t(i) = \widetilde{P}_t$, for $i \in [0,1]$;
θ is not allowed to change the price, $P_t(i) = P_{t-1}(i)$, for $i \in [0,1]$.

The profit-maximization problem of a reoptimizing firm i:

$$\max_{\widetilde{P}_t} \quad \sum_{j=0}^{\infty} \beta^j \theta^j E_t \left\{ \Lambda_{t+j} \left[\widetilde{P}_t Y_{t+j}(i) - P_{t+j} \mathrm{mc}_{t+j} Y_{t+j}(i) \right] \right\}$$

$$\text{s.t.} \quad Y_t(i) = Y_t \left(\frac{P_t(i)}{P_t} \right)^{-\varepsilon} \tag{5}$$

– Eq (5) is the demand for an intermediate good i;
– Λ_{t+j} is the Lagrange multiplier on the household's budget constraint;
– mc_{t+j} is the real marginal cost of output at time $t+j$.

Government

The government budget constraint:

$$T_t + \frac{B_t}{\exp\left(\eta_{B,t}\right) R_t} = P_t \frac{\overline{G} Y_t}{\exp\left(\eta_{G,t}\right)} + B_{t-1} + vW_t L_t$$

- $\frac{\overline{G} Y_t}{\exp\left(\eta_{G,t}\right)} = G_t$ is government spending;
- $vW_t L_t$ is the subsidy to the intermediate-good firms;
- $\eta_{G,t}$ is a government-spending shock,

$$\eta_{G,t+1} = \rho_G \eta_{G,t} + \epsilon_{G,t+1}, \qquad \epsilon_{G,t+1} \sim \mathcal{N}\left(0, \sigma_G^2\right)$$

Natural equilibrium

"Natural equilibrium" - the model in which the potential inefficiencies have been eliminated:

- Natural level of output $Y_{N,t}$ in the Taylor rule is a solution to **a planner's problem**

$$\max_{\{C_t, L_t\}_{t=0,\ldots,\infty}} E_0 \sum_{t=0}^{\infty} \beta^t \exp\left(\eta_{u,t}\right) \left[\frac{C_t^{1-\gamma} - 1}{1-\gamma} - \exp\left(\eta_{L,t}\right) \frac{L_t^{1+\vartheta} - 1}{1+\vartheta} \right]$$

$$\text{s.t. } C_t = \exp\left(\eta_{a,t}\right) L_t - G_t$$

where G_t is given. This implies

$$Y_{N,t} = \left[\frac{\exp\left(\eta_{a,t}\right)^{1+\vartheta}}{\left[\exp\left(\eta_{G,t}\right)\right]^{-\gamma} \exp\left(\eta_{L,t}\right)} \right]^{\frac{1}{\vartheta+\gamma}}$$

Summary of equilibrium conditions

- **FOCs of the intermediate-good firms**

$$S_t = \frac{1}{\exp\left(\eta_{a,t}\right)} \cdot \exp\left(\eta_{u,t} + \eta_{L,t}\right) L_t^{\varphi} Y_t + \beta\theta E_t\left\{\pi_{t+1}^{\varepsilon} S_{t+1}\right\}$$

$$F_t = C_t^{-\gamma} Y_t + \beta\theta E_t\left\{\pi_{t+1}^{\varepsilon-1} F_{t+1}\right\}$$

$$\frac{S_t}{F_t} = \left[\frac{1-\theta\pi_t^{\varepsilon-1}}{1-\theta}\right]^{\frac{1}{1-\varepsilon}}$$

- **Euler equation of the household's problem**

$$\exp\left(\eta_{u,t}\right) C_t^{-\gamma} = \beta \exp\left(\eta_{B,t}\right) R_t E_t\left[\frac{\exp\left(\eta_{u,t+1}\right) C_{t+1}^{-\gamma}}{\pi_{t+1}}\right]$$

- **Law of motion for the price distortion Δ_t**

$$\Delta_t = \left[(1-\theta)\left[\frac{1-\theta\pi_t^{\varepsilon-1}}{1-\theta}\right]^{\frac{\varepsilon}{\varepsilon-1}} + \theta\frac{\pi_t^{\varepsilon}}{\Delta_{t-1}}\right]^{-1}$$

Summary of equilibrium conditions

We have

- Stochastic processes for 6 exogenous shocks $\left\{\eta_{u,t}, \eta_{L,t}, \eta_{B,t}, \eta_{a,t}, \eta_{G,t}, \eta_{R,t}\right\}$.
- 7 endogenous equilibrium equations (5 above equations, $C_t = \left(1 - \frac{\overline{G}}{\exp\left(\eta_{G,t}\right)}\right) Y_t$, and $Y_t = \exp\left(\eta_{a,t}\right) L_t \Delta_t$.
- 8 unknowns $\left\{C_t, Y_t, R_t, L_t, \Delta_t, \pi_t, F_t, S_t\right\}$.
- 2 endogenous state variables $\left\{\Delta_{t-1}, R_{t-1}\right\}$.
- Thus, there are 8 (endogenous plus exogenous) state variables.

Summary of equilibrium conditions

- 7 equations and 8 unknowns:

$$\{C_t, L_t, Y_t, \pi_t, \Delta_t, R_t, S_t, F_t\}$$

- Number of unknown endogenous variables $>$ Number of equations $\Longrightarrow$ The model is not closed.
- To close the model, we need exogenous monetary policy.

Monetary authority

Taylor rule with ZLB on the net nominal interest rate:

$$R_t = \max\left\{1, \quad R_* \left(\frac{R_{t-1}}{R_*}\right)^{\mu} \left[\left(\frac{\pi_t}{\pi_*}\right)^{\phi_\pi} \left(\frac{Y_t}{Y_{N,t}}\right)^{\phi_y}\right]^{1-\mu} \exp\left(\eta_{R,t}\right)\right\}$$

- R_* = long-run gross nominal interest rate;
- π_t = gross inflation rate between $t-1$ and t;
- π_* = inflation target;
- $Y_{N,t}$ = natural level of output;
- $\eta_{R,t}$ = monetary shock

$$\eta_{R,t+1} = \rho_R \eta_{R,t} + \epsilon_{R,t+1}, \qquad \epsilon_{R,t+1} \sim \mathcal{N}\left(0, \sigma_R^2\right)$$

Monetary policy normalization

- **How to normalize the monetary policy after the end of the crisis?**
- The Great recession: ZLB and unconventional monetary policies (forward guidance and quantitative easing).
- Normalizing = switching back to Taylor rule.

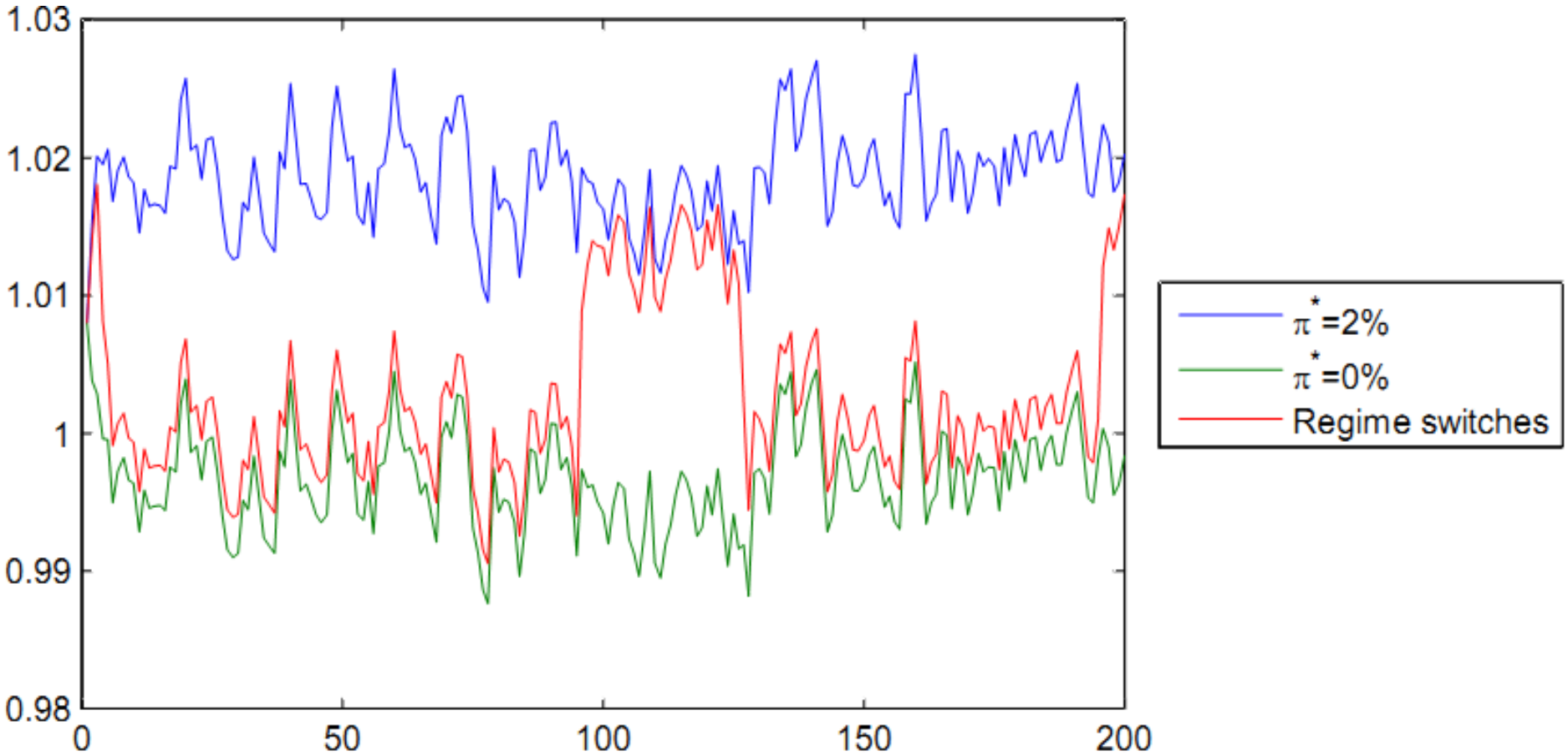
$$R_t = \max\left\{1, \quad R_* \left(\frac{R_{t-1}}{R_*}\right)^{\mu} \left[\left(\frac{\pi_t}{\pi_*}\right)^{\phi_\pi} \left(\frac{Y_t}{Y_{N,t}}\right)^{\phi_y}\right]^{1-\mu} \exp\left(\eta_{R,t}\right)\right\}$$

Questions:

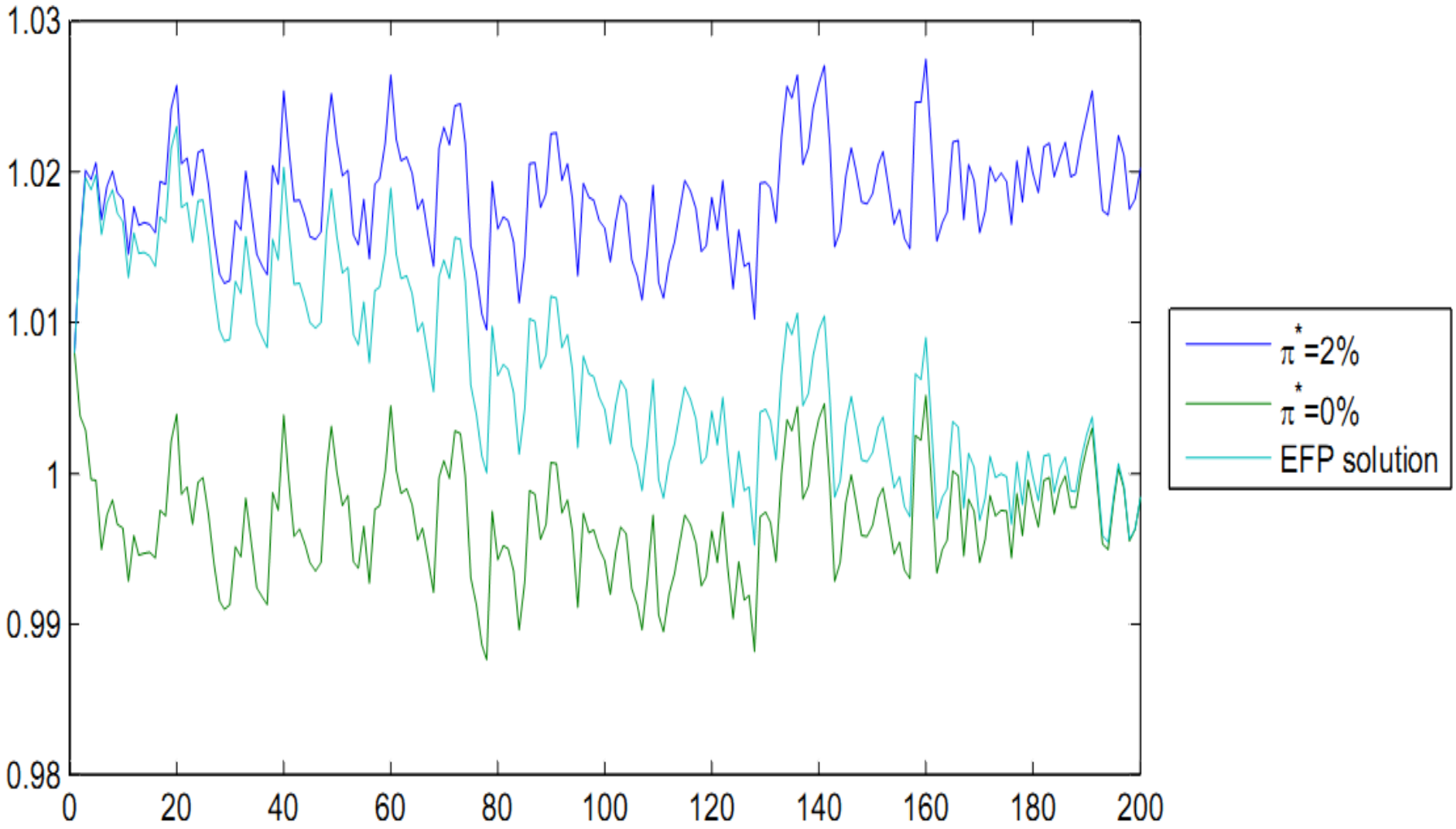
- Should the Fed normalize policy **now** or **later**?
- Should the Fed normalize policy **gradually** or **all at once**?
- Should the regime shift be **announced in advance**?
- Should the policy normalization be **time** or **state dependent**?:

$\Rightarrow$ We need to analyze and compare different transitions out of the ZLB.

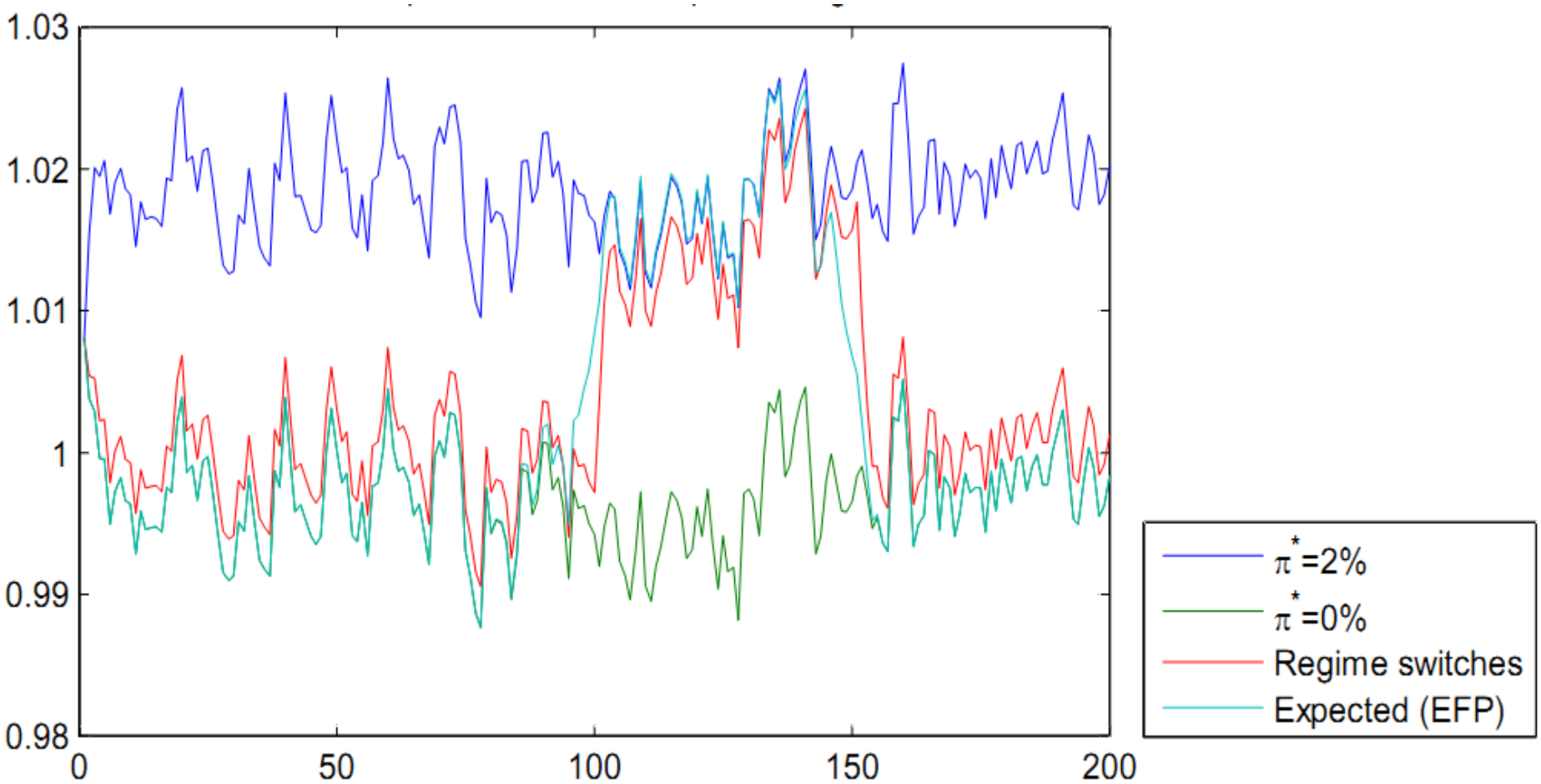
Model with regimen switches: Davig and Leeper (2008)



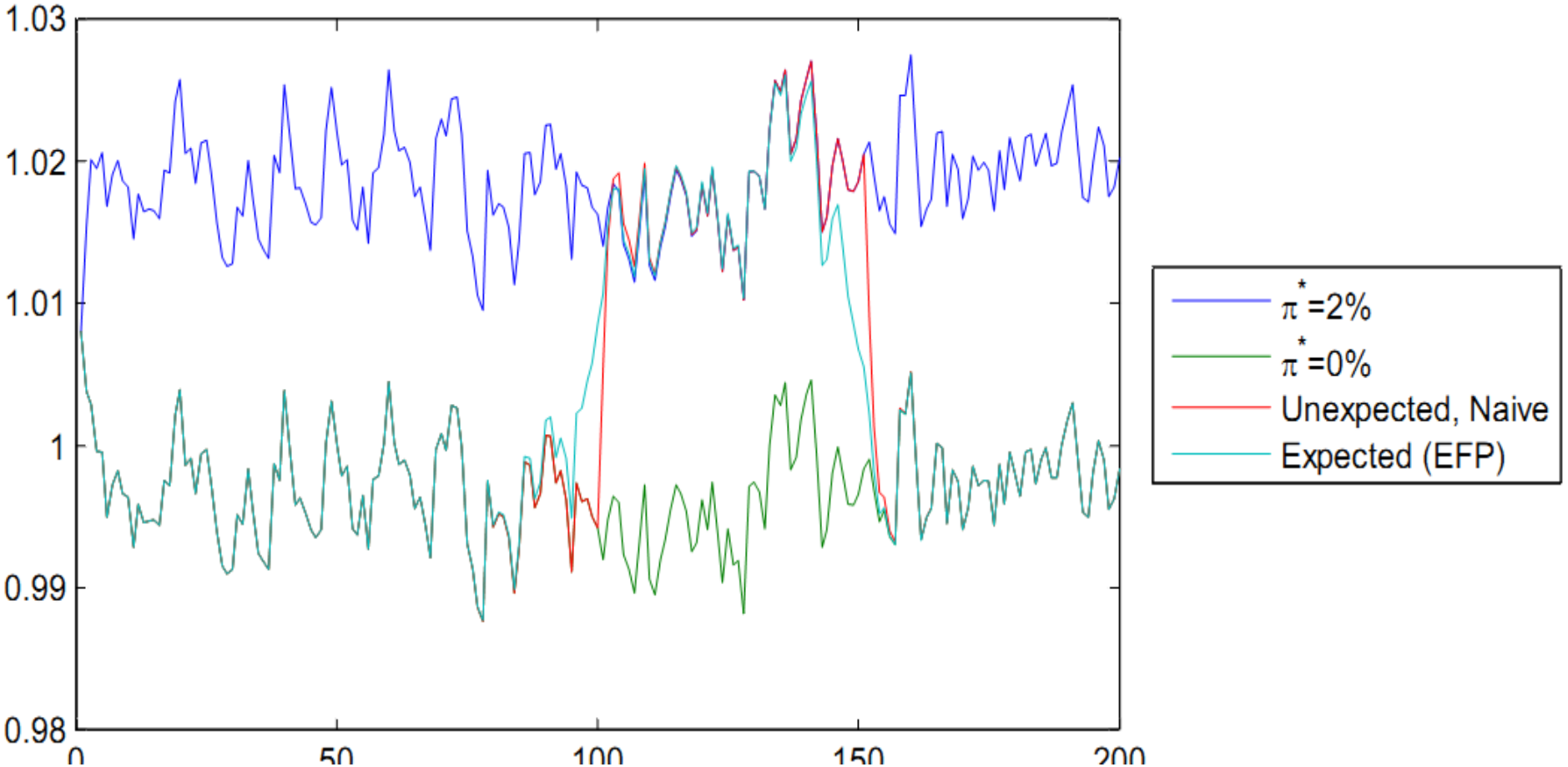
Model with parameter drift: EFP, Maliar et al. (2015)



Model with parameter shift: EFP versus regime switches



Model with parameter shift: EFP versus naive solution



Forward guidance puzzle

- "*Forward guidance puzzle*": an observation that output and consumption are excessively sensitive to central bank announcements about future interest rates in new Keynesian models.
  – Del Negro, Giannoni and Patterson (2015).
- McKay, Nakamura and Steinsson (2016):

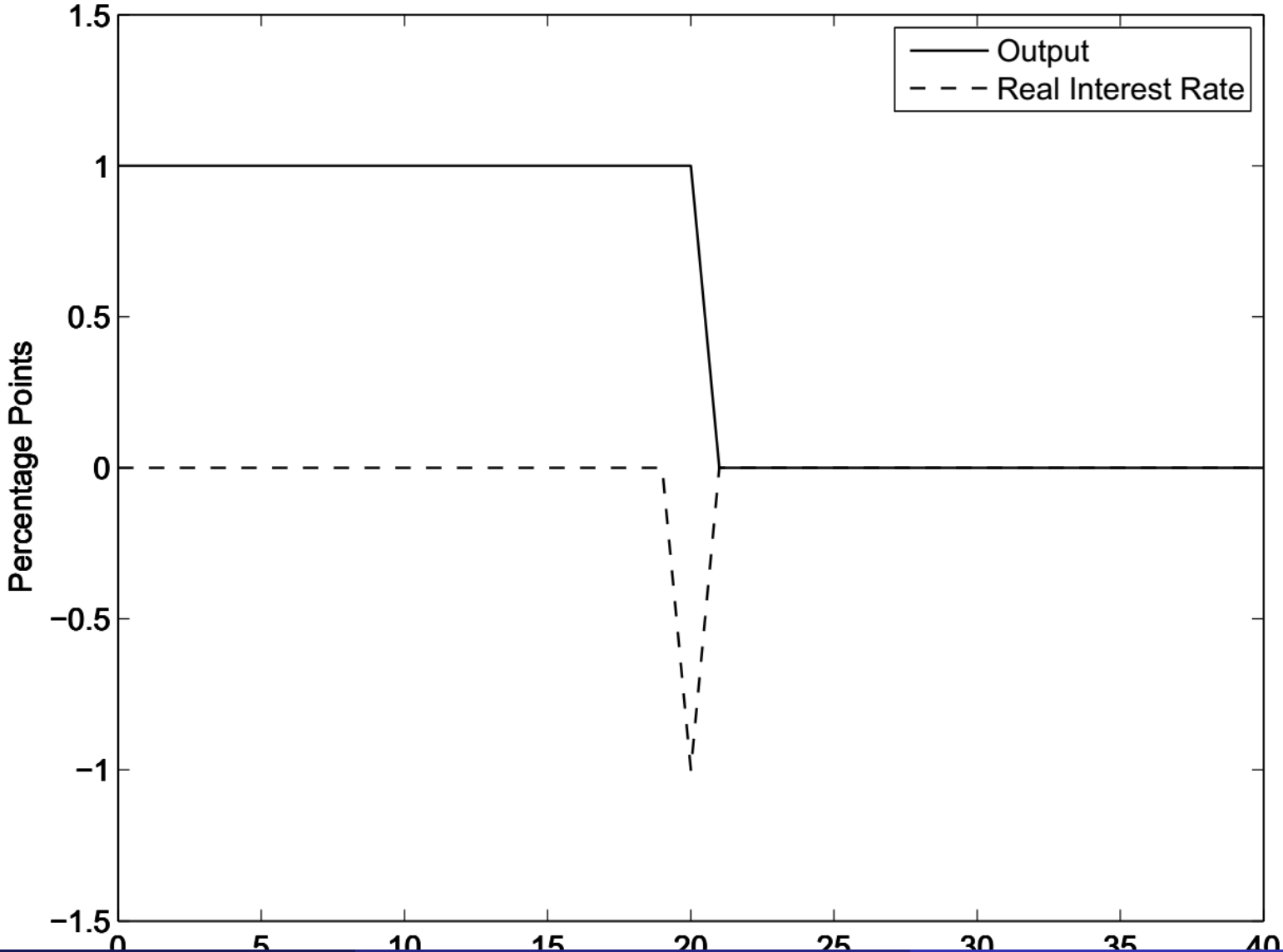
$$x_t = E_t\left[x_{t+1}\right] - \left[\log R_t - E_t \log \pi_{t+1} - \log R_*\right],$$
$$\log \pi_t = \beta E_t\left[\log \pi_{t+1}\right] + \kappa x_t,$$
$$\log R_{t+j} = E_{t+j}\left[\log \pi_{t+j+1}\right] + \log R_* + \varepsilon_{t,t+j}.$$

  $x_t \equiv \log Y_t - \log Y_{N,t}$;
  $\varepsilon_{t,t+j}$ = a $t+j$-period shock to the interest rate that is announced in period t. This yields

$$x_t = -\sum_{j=0}^{\infty} \varepsilon_{t,t+j}.$$

- *Today's shock to the interest rate has the same effect as a shock that happens a million years from now!*

Forward guidance puzzle



Why forward guidance puzzle was not observed in our experiments

- In some experiments, we fix interest rate for certain number of periods but it does not affect the present decisions (recall the last graph).
- Thus, the "forward guidance puzzle" was not observed in our experiments
- Why?
- This question is addressed in Maliar (2016) "Forward guidance puzzle and turnpike theorem".

Forward guidance and turnpike theorem

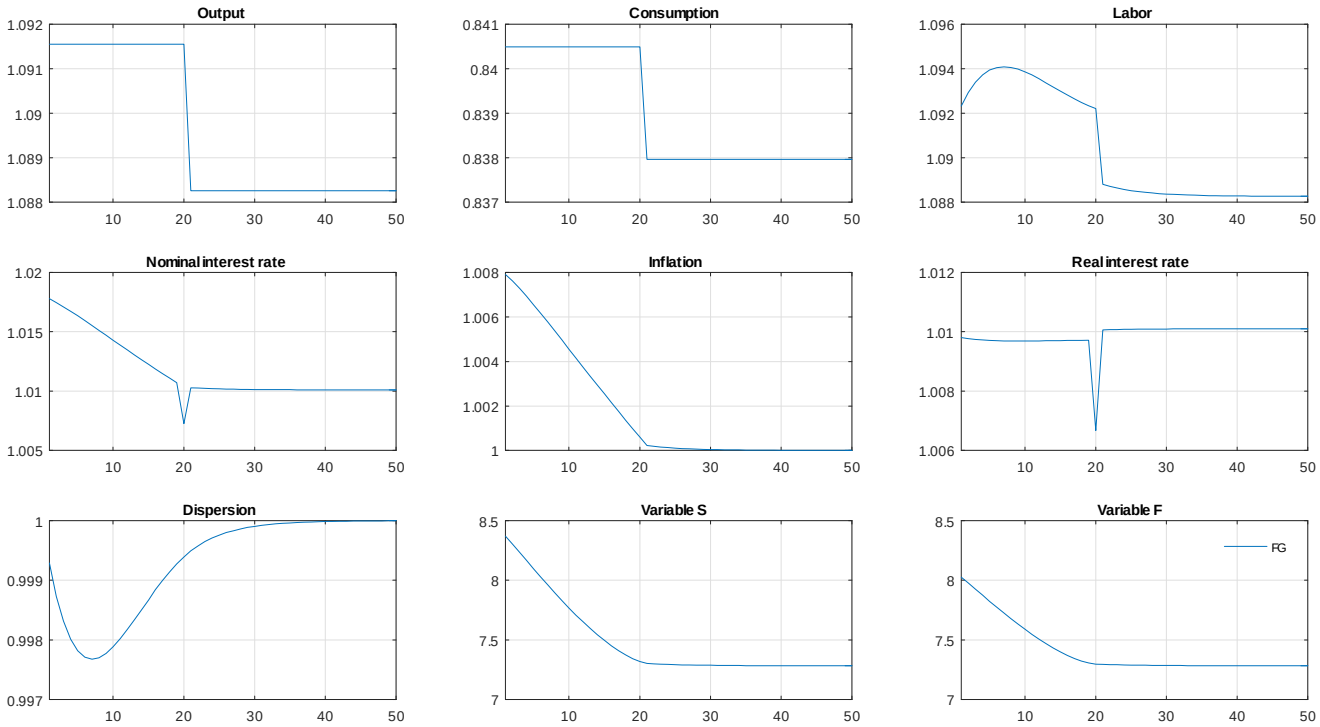
Maliar (2016) shows the following results:

- The neoclassical growth model satisfies the turnpike theorem under any parameterization, and future events have negligible effects on the present. $\Longrightarrow$ Forward guidance will have no effect!
- A new Keynesian model may or may not satisfy the turnpike theorem depending on parameterization. $\Longrightarrow$ This determines whether or not forward guidance puzzle is observed.
- We find that the forward guidance puzzle holds under very special parameterization (empirically implausible).
- Generally, the new Keynesian economy satisfies the turnpike theorem, and the forward guidance puzzle is not observed.
- Also, we find that the effect of forward guidance on output can be detrimental depending on specific parameterization.
- These findings hold in both linear and nonlinear models and are robust to the introduction of uncertainty.

Forward guidance and turnpike theorem (cont.)

- **Example 1.** Taylor rule:

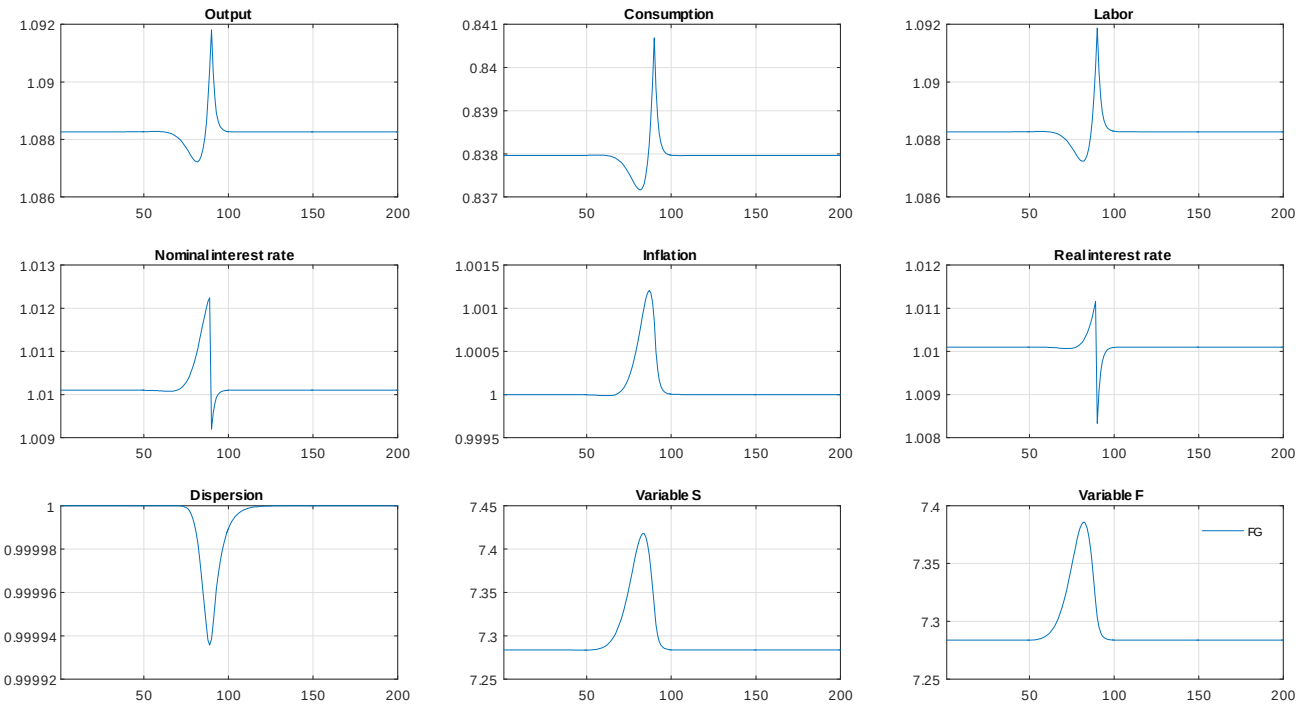
$$\log R_{t+j} = \phi_{E\pi} \cdot E_{t+j}\left[\log \pi_{t+j+1}\right] + \log R_* + \varepsilon_{t,t+j}, \quad \phi_{E\pi} = 1.000001.$$



- Forward guidance puzzle is observed and the turnpike theorem is not satisfied

Forward guidance and turnpike theorem (cont.)

Example 2. Taylor rule ($\phi_\pi = 2.21$, $\phi_{E\pi} = 0$, $\mu = 0.82$, $\phi_y = 0.07$):
$\log R_t - \log R_* = \mu \log R_{t-1} - \mu \log R_* + (1-\mu)\left[\phi_{E\pi}(E_t \log \pi_{t+1} - \log \pi_{tar}) + \phi_\pi(\log \pi_t - \log \pi_{tar}) + \phi_y(\log Y_t - \log Y_t^N)\right] + \eta_t^R$.



- Forward guidance puzzle is not observed and the turnpike theorem is satisfied

Conclusion

- Stationary Markov class of models is a dominant framework in recent economic literature.
- A shortcoming of this framework is that it generally restricts the parameters of economic models to be constant, and it restricts the behavior patterns to be time invariant.
- In this paper, we construct a more flexible class of nonstationary Markov models that allows for time-varying structural parameters and decision functions.
- We propose EFP framework for solving, calibrating, simulating and estimating of parameters in such models.
- We show how extended function path (EFP) can be used for analyzing nonstationary and nonrecurrent transitions from one policy to another.
- Literally, EFP makes it possible to construct a unique historical path of real-world economies.

Computer codes

"*EFP_MMTT_2015.zip*" - *Extended Function Path (EFP) method for time-dependent models.*

- Computes an accurate solution to a test-model with labor augmenting technological progress and balanced growth using a transformation to stationary model.
- Computes an EFP solution to a nonstationary test model directly, without using the property of balanced growth.

Stochastic environment formally

Time is discrete and infinite, $t = 0, 1, ...$ Let $(\Omega, \mathcal{F}, P)$ be a probability space:

a) $\Omega = \Pi_{t=0}^{\infty} \Omega_t$ is a space of sequences $\{\varepsilon_0, \varepsilon_1 ...\}$ such that $\varepsilon_t \in \Omega_t$ for all t, where Ω_t is a compact metric space endowed with the Borel σ–field $\mathcal{E}_t$. Here, Ω_t is the set of all possible states of the environment at t and $\varepsilon_t \in \Omega_t$ is the state of the environment at t.

b) $\mathcal{F}$ is the σ–algebra on Ω generated by cylinder sets of the form $\Pi_{\tau=0}^{\infty} A_\tau$, where $A_\tau \in \mathcal{E}_\tau$ for all τ and $A_\tau = \Omega_\tau$ for all but finitely many τ.

c) P is the probability measure on $(\Omega, \mathcal{F})$.

We denote by $\{\mathcal{F}_t\}$ a filtration on Ω, where $\mathcal{F}_t$ is a sub σ–field of $\mathcal{F}$ induced by a partial history up of environment $h_t = (\varepsilon_0, ..., \varepsilon_t) \in \Pi_{\tau=0}^{t} \Omega_\tau$ up to period t, i.e., $\mathcal{F}_t$ is generated by cylinder sets of the form $\Pi_{\tau=0}^{t} A_\tau$, where $A_\tau \in \mathcal{E}_\tau$ for all $\tau \leq t$ and $A_\tau = \Omega_\tau$ for $\tau > t$. In particular, we have that $\mathcal{F}_0$ is the course σ–field $\{0, \Omega\}$, and that $\mathcal{F}_\infty = \mathcal{F}$.

Stochastic process

We now provide definitions of stationary and Markov stochastic processes; these definitions are standard and closely follow Stokey and Lucas with Prescott (1989).

Definition 1. *(Stochastic process).* A stochastic process on $(\Omega, \mathcal{F}, P)$ is an increasing sequence of σ–algebras $\mathcal{F}_1 \subseteq \mathcal{F}_2 \subseteq ... \subseteq \mathcal{F}$; a measurable space $(Z, \mathcal{Z})$; and a sequence of functions $s_t : \Omega \to Z$ for $t \geq 0$ such that each s_t is $\mathcal{F}_t$ measurable.

Stationary stochastic process

Related economic literature commonly imposes a restriction of stationarity on stochastic processes.

Definition 2. *(Stationary stochastic process).* A stochastic process s on $(\Omega, \mathcal{F}, P)$ is called stationary if the unconditional probability measure

$$P_{t+1,...,t+n}(S) = P\left(\left\{\varepsilon \in \Omega : \left[s_{t+1}(\varepsilon), ..., s_{t+n}(\varepsilon)\right] \in S\right\}\right) \tag{6}$$

is independent of t for all $S \in \mathcal{Z}^n$, $t \geq 0$ and $n \geq 1$.

Stochastic process with stationary transitions

A related notion of stationarity is the property of stationary transitions. We denote by $P_{t+1,...,t+n}\left(S|s_t = \overline{s}_t, ..., s_0 = \overline{s}_0\right)$ the probability of the event $\left\{\varepsilon \in \Omega : \left[s_{t+1}\left(\varepsilon\right), ..., s_{t+n}\left(\varepsilon\right)\right] \in S\right\}$, given that the event $\left\{\varepsilon \in \Omega : \overline{s}_t = s_t\left(\varepsilon\right), ..., \overline{s}_0 = s_0\left(\varepsilon\right)\right\}$ occurs.

Definition 3. *(Stochastic process with stationary transitions).* A stochastic process s on $(\Omega, \mathcal{F}, P)$ has stationary transitions if the conditional probabilities

$$P_{t+1,...,t+n}\left(S|s_t = \overline{s}_t, ..., s_0 = \overline{s}_0\right) \tag{7}$$

are independent of t for all $S \in \mathcal{Z}^n$, $\varepsilon \in \Omega$, $t \geq 0$ and $n \geq 1$.

The assumption of stationary transition probabilities (7) implies the property of stationarity (6) provided that the corresponding unconditional probability measures exist.

Markov process

In general, $P_{t+1,...,t+n}(S)$ and $P_{t+1,...,t+n}(S|\cdot)$ depend on the entire history of the events up to t (i.e., the stochastic process s_t is measurable with respect to the sub σ–field $\mathcal{F}_t$). However, history-dependent processes are difficult to analyze in a general case. It is of interest to distinguish special cases in which the dependence on history has relatively simple and tractable form. A well-known case is a class of Markov processes.

Definition 4. *(Markov process).* A stochastic process s on $(\Omega, \mathcal{F}, P)$ is (first-order) Markov if

$$P_{t+1,...,t+n}(S|s_t = \overline{s}_t, ..., s_0 = \overline{s}_0) = P_{t+1,...,t+n}(S|s_t = \overline{s}_t), \quad (8)$$

for all $S \in \mathcal{Z}^n$, $t \geq 0$ and $n \geq 1$.

The key property of a Markov process is that it is memoryless, namely, all past history $(s_t, ..., s_0)$ is irrelevant for determining the future realizations except of the most recent past s_t.

Finite horizon version of the economy and its feasible program

Consider a finite horizon version of the economy (1)–(3) with a given terminal condition for capital k_T:

$$\max_{\{c_t, k_{t+1}\}_{t=0}^{T}} E_0 \left[\sum_{t=0}^{T} \beta^t u_t (c_t) \right] \tag{9}$$

$$\text{s.t. } c_t + k_{t+1} = (1-\delta) k_t + f_t (k_t, z_t) \tag{10}$$

$$z_{t+1} = \varphi_t (z_t, \varepsilon_{t+1}), \tag{11}$$

initial condition (k_0, z_0) and terminal condition k_T are given.

Definition A1 *(Feasible programs in the finite horizon economy).* A feasible program in the finite horizon economy is a pair of adapted (i.e., $\mathcal{F}_t$ measurable for all t) processes $\{c_t, k_t\}_{t=0}^{T}$ such that given initial condition k_0 and a partial history $h_T = (\varepsilon_0, ..., \varepsilon_T)$, these processes reach a given terminal condition k_T at T and satisfy $c_t \geq 0$, $k_t \geq 0$ and (10), (11) for all $t = 1, ...T$.

Optimal program in the finite horizon economy with a zero terminal condition

Let $\Im^{T,0}$ denote a set of all finite horizon feasible programs from given initial capital k_0 and given partial history $h_T \equiv (\varepsilon_0, ..., \varepsilon_T)$ that attain given $k_T = 0$ at T.

Definition A2 *(Optimal program in the finite horizon model).* A feasible finite horizon program $\left\{c_t^{T,0}, k_t^{T,0}\right\}_{t=0}^{T} \in \Im^{T,0}$ is called optimal if

$$E_0\left[\sum_{t=0}^{T} \beta^t \left\{u_t(c_t^{T,0}) - u_t\left(c_t\right)\right\}\right] \geq 0 \qquad \text{(A1)}$$

for every feasible process $\left\{c_t, k_t\right\}_{t=0}^{T} \in \Im^{T,0}$.

The existence result for the finite horizon economy with a zero terminal condition is established using Bellman equation approach (see Mitra and Nyarko (1991), Theorem 3.1) and Euler equation approach (see Majumdar and Zilcha (1987), Theorems 1 and 2).

Lemma 1

We next show that under terminal condition $k_T^{T,0} = k_T = 0$, the optimal program in the finite horizon economy has a well-defined limit.

Lemma 1. *A finite horizon optimal program* $\left\{c_t^{T,0}, k_t^{T,0}\right\}_{t=0}^{T} \in \Im^{T,0}$ *with a zero terminal condition* $k_T^{T,0} = 0$ *converges to a limit program* $\left\{c_t^{\lim}, k_t^{\lim}\right\}_{t=0}^{\infty}$ *when* $T \longrightarrow \infty$, *i.e.,*

$$k_t^{\lim} \equiv \lim_{T\to\infty} k_t^{T,0} \quad \text{and} \quad c_t^{\lim} \equiv \lim_{T\to\infty} c_t^{T,0}, \quad \text{for } t = 0, 1, ... \qquad \text{(A2)}$$

Proof of Lemma 1 (I)

i) Extending time horizon from T to $T+1$ increases T-period capital of the finite horizon optimal program, i.e., $k_T^{T+1,0} > k_T^{T,0}$.
– To see this, note that the model with time horizon T has zero (terminal) capital $k_T^{T,0} = 0$ at T.
– When time horizon is extended from T to $T+1$, the model has zero (terminal) capital $k_{T+1}^{T+1,0} = 0$ at $T+1$ but it has strictly positive capital $k_T^{T+1,0} > 0$ at T; this follows by the Inada conditions–Assumption 4.

Proof of Lemma 1 (II)

ii) The optimal program for the finite horizon economy has the following property of monotonicity with respect to the terminal condition:
– If $\{c_t', k_t'\}_{t=0}^{T}$ and $\{c_t'', k_t''\}_{t=0}^{T}$ are two optimal programs for the finite horizon economy with terminal conditions $\kappa' < \kappa''$, then the respective optimal capital choices have the same ranking in each period, i.e., $k_t' \leq k_t''$ for all $t = 1, ...T$.
– This follows by either Bellman equation (Mitra and Nyarko (1991)) or Euler equation (Majumdar and Zilcha (1987)) or lattice (Hopenhayn and Prescott (1992)) programming techniques.
– Hence, $\left\{k_t^{T,0}\right\}_{t=0}^{T}$ shifts up (weakly) in a pointwise manner when T increases to $T+1$, i.e., $k_t^{T+1,0} \geq k_t^{T,0}$ for $t \geq 0$.

Proof of Lemma 1 (III)

iii) By construction, capital from the optimal program $\left\{c_t^{T,0}, k_t^{T,0}\right\}_{t=0}^{T}$ is bounded from above by the capital accumulation process $\{0, k_t^{\max}\}_{t=0}^{T}$, i.e., $k_t^{T,0} \leq k_t^{\max} < \infty$ for $t \geq 0$.
– A sequence that is bounded and monotone is known to have a well-defined limit. ■

Lemma 2

- We show that the optimal program of the T-period stationary economy converges to the same limit program as the optimal program of the finite horizon economy with a zero terminal condition.
- Let $\Im^T$ denote a set of all feasible finite horizon programs that attains a terminal condition of the T-period stationary economy.
- We assume the same initial capital (k_0, z_0) and the same partial history $h_T \equiv (\varepsilon_0, ..., \varepsilon_T)$ as those fixed for the finite horizon economy.

Lemma 2. *The optimal program of the T-period stationary economy $\left\{c_t^T, k_t^T\right\}_{t=0}^T \in \Im^T$ converges to a unique limit program $\left\{c_t^{\lim}, k_t^{\lim}\right\}_{t=0}^{\infty} \in \Im^{\infty}$ defined in (A2) as $T \to \infty$ i.e., for all $t \geq 0$*

$$k_t^{\lim} \equiv \lim_{T \to \infty} k_t^T \quad \text{and} \quad c_t^{\lim} \equiv \lim_{T \to \infty} c_t^T. \tag{A3}$$

Proof of Lemma 2 (I)

i) Observe that:
– The optimal program of the T-period stationary economy has a positive capital stock $k_T^T > 0$ at T;
– (This is because the terminal capital is generated by the capital decision function of a stationary version of the model);
– But for the optimal program $\left\{c_t^{T,0}, k_t^{T,0}\right\}_{t=0}^{T} \in \Im^{T,0}$ of the finite horizon economy, it is zero by definition, $k_T^{T,0} = 0$.

ii) The property of monotonicity with respect to terminal condition implies that if $k_T^T > k_T^{T,0}$, then $k_t^T \geq k_t^{T,0}$ for all $t = 1, ..., T$; see ii) of the proof to Lemma 1.

iii) Fix some $\tau \in \{1, ..., T\}$. We show that up to τ, $\left\{c_t^T, k_t^T\right\}_{t=0}^{\tau}$ does not give higher expected utility than $\left\{c_t^{T,0}, k_t^{T,0}\right\}_{t=0}^{\tau}$, i.e.,

$$E_0\left[\sum_{t=0}^{\tau} \beta^t \left\{u_t\left(c_t^T\right) - u_t(c_t^{T,0})\right\}\right] \leq 0. \qquad \text{(A4)}$$

Proof of Lemma 2 (II)

– Toward contradiction, assume that it does, i.e.,

$$E_0\left[\sum_{t=0}^{\tau}\beta^t\left\{u_t\left(c_t^T\right)-u_t(c_t^{T,0})\right\}\right]>0. \tag{A5}$$

– Consider a new process

$$\left\{c_t',k_t'\right\}_{t=0}^{\tau}\equiv\left\{c_t^T,k_t^T\right\}_{t=0}^{\tau-1}\cup\left\{c_\tau^T+k_\tau^T-k_\tau^{T,0},k_\tau^{T,0}\right\}$$

– (it follows $\left\{c_t^T,k_t^T\right\}_{t=0}^{T}\in\Im^T$ up to period $\tau-1$ and that drops down at τ to match $k_\tau^{T,0}$ of the finite horizon program $\left\{c_t^T,k_t^T\right\}_{t=0}^{T}\in\Im^{T,0}$).

– By monotonicity in part ii), we have $k_\tau^T-k_\tau^{T,0}\geq 0$, so that

$$E_0\left[\sum_{t=0}^{\tau}\beta^t\left\{u_t\left(c_t'\right)-u_t\left(c_t^T\right)\right\}\right]$$
$$=E_0\left[\beta^\tau\left\{u_t\left(c_\tau^T+k_\tau^T-k_\tau^{T,0}\right)-u_t\left(c_\tau^T\right)\right\}\right]\geq 0, \tag{12}$$

where the last inequality follows by Assumption 2 of strictly increasing u_t.

Proof of Lemma 2 (III)

iv) By construction $\{c_t', k_t'\}_{t=0}^{\tau}$ and $\left\{c_t^{T,0}, k_t^{T,0}\right\}_{t=0}^{\tau}$ reach the same capital $k_\tau^{T,0}$ at τ.

– Let us extend the program $\{c_t', k_t'\}_{t=0}^{\tau}$ to T by assuming that it follows the process $\left\{c_t^{T,0}, k_t^{T,0}\right\}_{t=0}^{T}$ from the period $\tau+1$ up to T, i.e., $\{c_t', k_t'\}_{t=\tau+1}^{T} \equiv \left\{c_t^{T,0}, k_t^{T,0}\right\}_{t=\tau+1}^{T}$.

– Then, we have

$$E_0\left[\sum_{t=0}^{T}\beta^t\left\{u_t\left(c_t'\right)-u_t(c_t^{T,0})\right\}\right]=E_0\left[\sum_{t=0}^{\tau}\beta^t\left\{u_t\left(c_t'\right)-u_t(c_t^{T,0})\right\}\right]$$

$$\geq E_0\left[\sum_{t=0}^{\tau}\beta^t\left\{u_t\left(c_t^{T}\right)-u_t\left(c_t^{T,0}\right)\right\}\right]>0, \quad \text{(A7)}$$

where the last two inequalities follow by result (**??**) and assumption (A5), respectively.

Proof of Lemma 2 (IV)

– Thus, we obtain a contradiction: The constructed program $\{c_t', k_t'\}_{t=0}^{T} \in \Im^{T,0}$ is feasible in the finite horizon economy with a zero terminal condition, $k_T' = 0$, and it gives strictly higher expected utility than the optimal program $\left\{c_t^{T,0}, k_t^{T,0}\right\}_{t=0}^{T} \in \Im^{T,0}$ in that economy.

v) Holding τ fixed, we compute the limit of (A4) by letting T go to infinity:

$$\lim_{T\to\infty} E_0\left[\sum_{t=0}^{\tau}\beta^t\left\{u_t\left(c_t^T\right) - u_t(c_t^{T,0})\right\}\right] =$$

$$\lim_{T\to\infty} E_0\left[\sum_{t=0}^{\tau}\beta^t u_t\left(c_t^T\right)\right] - E_0\left[\sum_{t=0}^{\tau}\beta^t u_t\left(c_t^{\lim}\right)\right] \leq 0. \quad \text{(A8)}$$

Proof of Lemma 2 (V)

vi) The last inequality implies that for any $\tau \geq 1$, the limit program $\left\{c_t^{\lim}, k_t^{\lim}\right\}_{t=0}^{\infty} \in \Im^{\infty}$ of the finite horizon economy $\left\{c_t^{T,0}, k_t^{T,0}\right\}_{t=0}^{T} \in \Im^{T,0}$ with a zero terminal condition $k_T^{T,0} = 0$ gives at least as high expected utility as the optimal limit program $\left\{c_t^{T}, k_t^{T}\right\}_{t=0}^{T} \in \Im^{T}$ of the T-period stationary economy.

– Since Assumptions 1-8 imply that the optimal program is unique, we conclude that $\left\{c_t^{\lim}, k_t^{\lim}\right\}_{t=0}^{\infty} \in \Im^{\infty}$ is a unique limit of the optimal program $\left\{c_t^{T}, k_t^{T}\right\}_{t=0}^{T} \in \Im^{T}$ of the T-period stationary economy. ■

Lemma 3

- We now show a connection between the optimal programs of the finite horizon and infinite horizon economies.
- Namely, we show that the finite horizon economy with a zero terminal condition $k_T^{T,0} = 0$ converges to the nonstationary infinite horizon economy as $T \longrightarrow \infty$.

Lemma 3. *The limit program* $\left\{c_t^{\lim}, k_t^{\lim}\right\}_{t=0}^{\infty}$ *is a unique optimal program* $\left\{c_t^{\infty}, k_t^{\infty}\right\}_{t=0}^{\infty} \in \Im^{\infty}$ *in the infinite horizon nonstationary economy (1)–(3).*

Proof of Lemma 3 (I)

i) Toward contradiction, assume that $\left\{c_t^{\lim}, k_t^{\lim}\right\}_{t=0}^{\infty}$ is not an optimal program of the infinite horizon economy $\left\{c_t^{\infty}, k_t^{\infty}\right\}_{t=0}^{\infty} \in \Im^{\infty}$.
– By definition of limit, there exists a real number $\epsilon > 0$ and a subsequence of natural numbers $\left\{T_1, T_2, ...\right\} \subseteq \left\{0, 1, ...\right\}$ such that $\left\{c_t^{\infty}, k_t^{\infty}\right\}_{t=0}^{\infty} \in \Im^{\infty}$ gives strictly higher expected utility than the limit program of the finite horizon economy $\left\{c_t^{\lim}, k_t^{\lim}\right\}_{t=0}^{\infty}$, i.e.,

$$E_0\left[\sum_{t=0}^{T_n} \beta^t \left\{u_t\left(c_t^{\infty}\right) - u_t\left(c_t^{\lim}\right)\right\}\right] > \epsilon \text{ for all } T_n \in \left\{T_1, T_2, ...\right\}. \tag{A9}$$

ii) Let us fix some $T_n \in \left\{T_1, T_2, ...\right\}$ and consider any finite $T \geq T_n$.
– Assumptions 1-8 imply that $k_T^{\infty} > 0$, while $k_T^{T,0} = 0$ by definition of the finite horizon economy with a zero terminal condition.
– The monotonicity of the optimal program with respect to a terminal condition implies that if $k_T^{\infty} > k_T^{T,0}$, then $k_t^{\infty} \geq k_t^{T,0}$ for all $t = 1, ..., T$; see part ii) of the proof of Lemma 1.

Proof of Lemma 3 (II)

iii) Following the arguments in iii). of the proof of Lemma 2, we can show that up to period T_n, the optimal program $\left\{c_t^{\infty}, k_t^{\infty}\right\}_{t=0}^{T_n}$ does not give higher expected utility than $\left\{c_t^{T,0}, k_t^{T,0}\right\}_{t=0}^{T_n}$, i.e.,

$$E_0\left[\sum_{t=0}^{T_n}\beta^t\left\{u_t\left(c_t^{\infty}\right)-u_t\left(c_t^{T,0}\right)\right\}\right]\leq 0 \text{ for all } T_n. \tag{A10}$$

iv) Holding T_n fixed, we compute the limit of (A10) by letting T go to infinity:

$$\begin{aligned}\lim_{T\to\infty} E_0&\left[\sum_{t=0}^{T_n}\beta^t\left\{u_t\left(c_t^{\infty}\right)-u_t\left(c_t^{T,0}\right)\right\}\right]\\ &= E_0\left[\sum_{t=0}^{T_n}\beta^t u_t\left(c_t^{\infty}\right)-\beta^t u_t\left(c_t^{\lim}\right)\right]\leq 0 \text{ for all } T_n. \end{aligned}\tag{A11}$$

However, result (A11) contradicts to our assumption in (A9).

Proof of Lemma 3 (III)

v) We conclude that for any subsequence $\{T_1, T_2, ...\} \subseteq \{0, 1, ...\}$, we have

$$E_0 \left[\sum_{t=0}^{T_n} \beta^t \left\{ u_t \left(c_t^{\infty} \right) - u_t \left(c_t^{\lim} \right) \right\} \right] \leq 0 \text{ for all } T_n. \tag{A12}$$

However, under Assumptions 1-8, the optimal program $\{c_t^{\infty}, k_t^{\infty}\}_{t=0}^{\infty} \in \Im^{\infty}$ is unique, and hence, it must be that $\{c_t^{\infty}, k_t^{\infty}\}_{t=0}^{\infty}$ coincides with $\{c_t^{\lim}, k_t^{\lim}\}_{t=0}^{\infty}$ for all $t \geq 0$. ■

Theorem 3 (turnpike theorem)

- We now combine the results of Lemmas 1-3 together into a turnpike-style theorem.
- Lemma 1 shows that the optimal program of the finite horizon economy with a zero terminal condition $\left\{c_t^{T,0}, k_t^{T,0}\right\}_{t=0}^{T} \in \Im^{T,0}$ converges to the limit program $\left\{c_t^{\lim}, k_t^{\lim}\right\}_{t=0}^{\infty}$.
- Lemma 2 shows that the optimal program of the T-period stationary economy $\left\{c_t^{T}, k_t^{T}\right\}_{t=0}^{T}$ also converges to the same limit program $\left\{c_t^{\lim}, k_t^{\lim}\right\}_{t=0}^{\infty}$.
- Lemma 3 shows that the limit program of the finite horizon economies $\left\{c_t^{\lim}, k_t^{\lim}\right\}_{t=0}^{\infty}$ is optimal in the nonstationary infinite horizon economy.
- Then, it must be the case that the limit optimal program of the T-period stationary economy $\left\{c_t^{T}, k_t^{T}\right\}_{t=0}^{T}$ is optimal in the infinite horizon nonstationary economy. This argument is formalized below.

Proof of Theorem 2

Let us fix a real number $\varepsilon > 0$ and a natural number τ such that $1 \leq \tau < \infty$.

i) Lemma 1 shows that $\left\{c_t^{T,0}, k_t^{T,0}\right\}_{t=0}^{T} \in \Im^{T,0}$ converges $\left\{c_t^{\lim}, k_t^{\lim}\right\}_{t=0}^{\infty}$ as $T \to \infty$. Then, definition of limit implies that there exists $T_1 > 0$ such that $\left|k_t^{T,0} - k_t^{\lim}\right| < \frac{\varepsilon}{3}$ for $t = 0, ..., \tau$.

ii) Lemma 2 implies that $\left\{c_t^{T}, k_t^{T}\right\}_{t=0}^{T}$ also converges to limit program $\left\{c_t^{\lim}, k_t^{\lim}\right\}_{t=0}^{\infty}$ as $T \to \infty$. Then, there exists $T_2 > 0$ such that $\left|k_t^{\lim} - k_t^{T}\right| < \frac{\varepsilon}{3}$ for $t = 0, ..., \tau$.

iii) Lemma 3 implies the program $\left\{c_t^{T,0}, k_t^{T,0}\right\}_{t=0}^{T} \in \Im^{T,0}$ converges to the infinite horizon optimal program $\left\{c_t^{\infty}, k_t^{\infty}\right\}_{t=0}^{\infty}$ as $T \to \infty$. Then, there exists $T_3 > 0$ such that $\left|k_t^{T,0} - k_t^{\infty}\right| < \frac{\varepsilon}{3}$ for $t = 0, ..., \tau$.

Proof of Theorem 2 (cont.)

iv) Then, the triangular inequality implies

$$\begin{aligned}\left|k_t^T - k_t^\infty\right| &= \left|k_t^T - k_t^{\lim} + k_t^{\lim} - k_t^{T,0} + k_t^{T,0} - k_t^\infty\right| \\ &\leq \left|k_t^T - k_t^{\lim}\right| + \left|k_t^{\lim} - k_t^{T,0}\right| + \left|k_t^{T,0} - k_t^\infty\right| < \frac{\varepsilon}{3} + \frac{\varepsilon}{3} + \frac{\varepsilon}{3} = \varepsilon,\end{aligned}$$

for $T \geq T(\varepsilon, \tau) \equiv \max\{T_1, T_2, T_3\}$. ■