Merging Simulation and Projection Approaches to Solve High-Dimensional Problems

Kenneth L. Judd, Lilia Maliar and Serguei Maliar

Stanford, Summer Workshop

Epsilon-distinguishable set (EDS) algorithm

- **A novel accurate method for solving dynamic economic models:** works for problems with high dimensionality, intractable for earlier solution methods:
  - *we accurately solve models with 20-50 state variables using a laptop.*
- **Related literature focuses on much lower dimensionality:** a special JEDC 2011's issue compares solution methods using models with 12-20 state variables.
- **Examples of potential applications of the EDS algorithm:**
  - macroeconomics (many heterogeneous consumers);
  - international economics (many countries);
  - industrial organization (many firms);
  - finance (many assets);
  - climate change (many sectors and countries); etc.
- **EDS algorithm is a global method:** can handle strong non-linearities and inequality constraints.
  - *we solve a new Keynesian model with the zero lower bound.*

Epsilon-distinguishable set (EDS) algorithm

- **EDS algorithm merges stochastic simulation and projection approaches:**
  - we use simulation to approximate the ergodic measure of the solution;
  - we construct a fixed grid covering the support of the constructed ergodic measure;
  - we use projection techniques to accurately solve the model on that grid.

- **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.*

Figure 1a. A set of simulated points

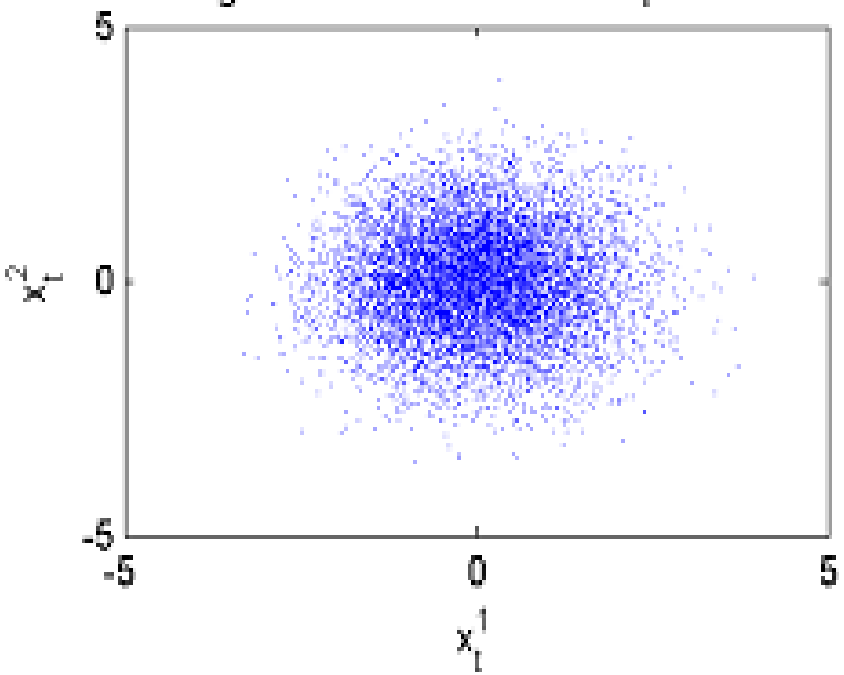
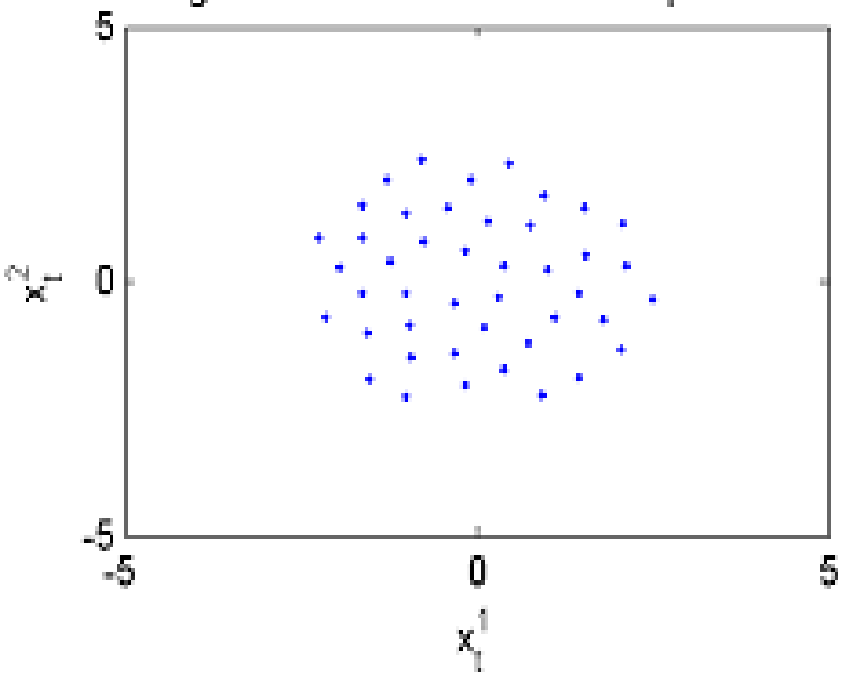


Figure 1b. A subset of simulated points



Ingredients of the EDS algorithm used for high-dimensional models

- **Endogenous solution domain:** our EDS grid is constructed by approximating the ergodic set - we avoid costs of finding a solution in the areas of state space that are never visited in equilibrium.
- **Low-cost integration:** non-product monomial and one-point quadrature integration rules.
- **Derivative-free solvers for finding the polynomial coefficients:** fixed-point iteration.
- **Efficient vectorized approaches for finding the control variables:** precomputation and iteration-on-allocation by Maliar, Maliar and Judd (2011).

$$\Downarrow$$

- **Taken together, these ingredients allow us to meet challenges of high-dimensional problems.**

Remarks

- **Codes**
  - Not yet available for the EDS method.
  - But a simple and well-documented MATLAB code is available for generalized stochastic simulation method (GSSA).
  - GSSA is less efficient but can also solve models with 20-50 state variables.

- **Our class of problems differs from Krusell and Smith (1998)**
  - We can have any heterogeneity of agents but the number of heterogeneous agents is not too large (like 20-50) and we work with the true state space.
  - Krusell and Smith (1998) have a continuum of ex ante identical agents, and they describe aggregate behavior with a reduced state space (moments of aggregate variables).

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.

Conventional projection methods, Judd (1992)

Characteristic features

- Solve a model on a prespecified grid of points.
- Use quadrature integration for approximating conditional expectations.
- Compute polynomial coefficients of policy functions using Newton's type solver.

A projection method for the optimal growth model

- Choose a grid of M points in the state space $\{k_m, \theta_m\}_{m=1}^{M}$.
- Choose nodes, ϵ_j, and weights, ω_j, $j = 1, ..., J$, for approximating integrals.
- Parameterize capital policy function $K(\cdot) \approx \widehat{K}(\cdot; b)$ by a polynomial

$$\widehat{K}(k_m, \theta_m; b) = b_0 + b_1 k_m + b_2 \theta_m + b_3 k_m^2 + b_4 k_m \theta_m + ... + b_n \theta_m^L$$

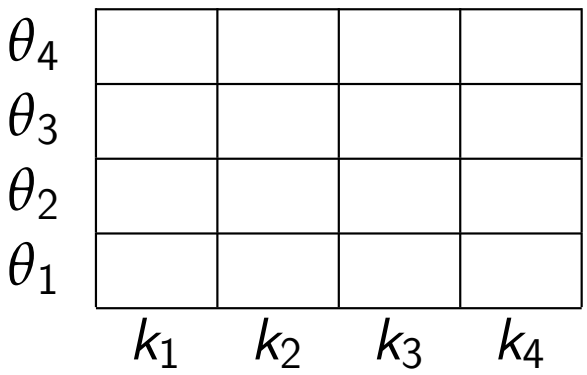
- Solve for $b \equiv (b_0, b_1, ..., b_n)$ that satisfies the Euler equation

$$\min_b \left\| u'(c_m(b)) - \beta \sum_{j=1}^{J} \omega_j \left[u'\left(c'_{m,j}(b)\right) \left(\left[1 - \delta + \theta'_{m,j} f'\left(k'_m(b)\right) \right] \right) \right] \right\|$$

$$\begin{aligned}
\theta'_{m,j} &= \theta_m^{\rho} \exp(\epsilon_j), \\
k'_m(b) &= \widehat{K}(k_m, \theta_m; b) \\
k''_{m,j}(b) &= \widehat{K}\left(k'_m, \theta'_{m,j}; b\right) \\
c_m(b) &= (1-\delta) k_m + \theta_m f(k_m) - k'_m(b) \\
c'_{m,j}(b) &= (1-\delta) k'_m(b) + \theta'_{m,j} f\left(k'_m(b)\right) - k''_{m,j}(b)
\end{aligned}$$

Conventional projection methods: curse of dimensionality

- Very accurate and fast with few state variables but cost grows exponentially with dimensionality!
  (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!



- 2 state variables with 4 grid points $\Rightarrow 4 \times 4 = 4^2 = 16$
- 3 state variables with 4 grid points $\Rightarrow 4^3 = 64$
- 10 state variables with 4 grid points $\Rightarrow 4^{10} = 1,048,576$
(With 100 grid points $\Rightarrow 100^{10} = 10^{20}$).

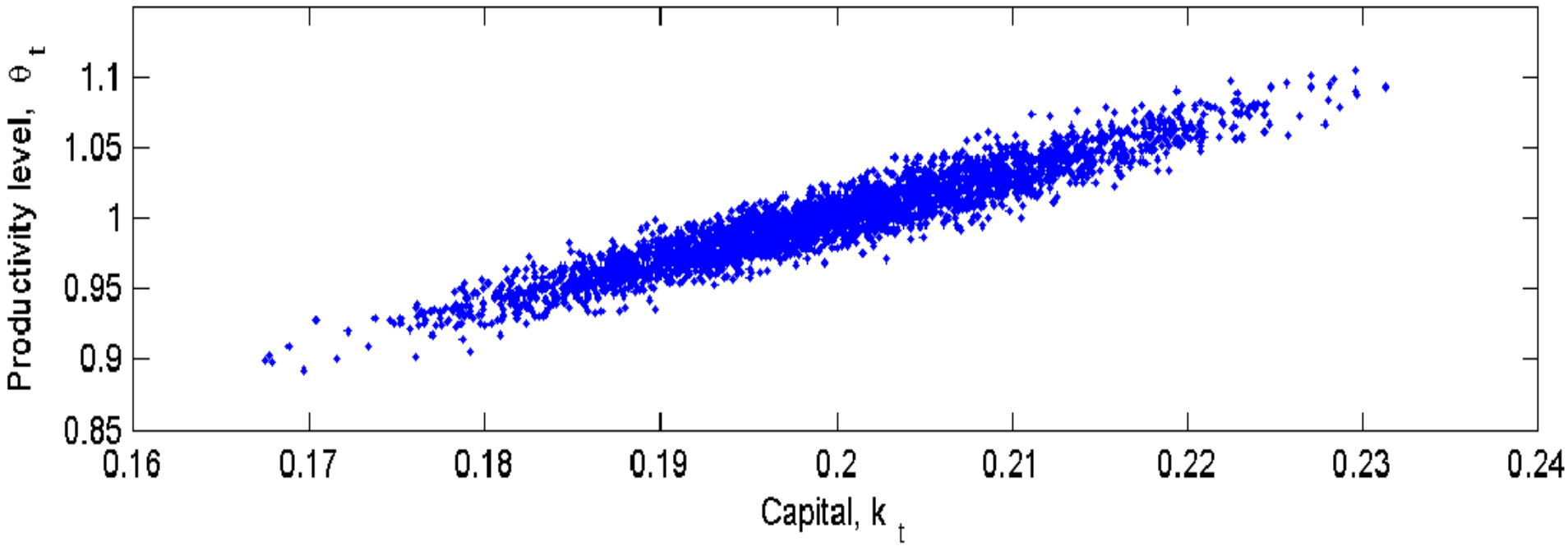
- *Krueger and Kubler (2004):* Smolyak's sparse grid - reduces the number of points within the multidimensional hypercube domain but not the size of the hypercube domain itself.

Stochastic simulation methods, e.g., Marcet (1988)

- Compute solution on simulated series.
  - Draw shocks $\{\epsilon_t\}_{t=1}^T$ Compute and fix productivity levels $\{\theta_t\}_{t=1}^T$.
  - Guess a decision function $\widehat{K}(k, \theta; b)$.
  - Simulate time series $\{c_t, k_{t+1}\}_{t=0}^T$.
  - Check equilibrium conditions and recompute $\widehat{b}$.
  - Iterate on b until convergence.
- Use Monte Carlo integration for approximating conditional expectations.
- Use learning techniques for solving for parameters of decision functions.

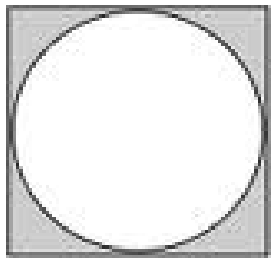
Advantage of stochastic simulation method

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

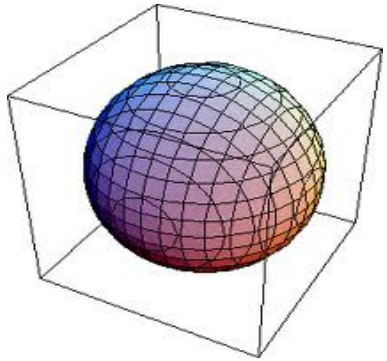
- How much can we save on cost using the ergodic-set domain comparatively to the hypercube domain?
- Suppose the ergodic set is a circle (it was an ellipse in the figure).
- In the 2-dimensional case, a circle inscribed within a square occupies about 79% of the area of the square.
- The reduction in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

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

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

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

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

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

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!

Merging projection and stochastic simulation

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 (EDS) 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*.

Mathematical foundations

We provide mathematical foundations for the EDS grid

- We establish computational complexity, dispersion, cardinality and degree of uniformity of the EDS grid constructed on simulated series.
- We perform the typical and the worst-case analysis for the discrepancy of the EDS grid.
- We relate our results to recent mathematical literature on
  - covering problems (e.g., measuring entropy); see, Temlyakov (2011).
  - random sequential packing problems, (e.g., germ contagion); see, Baryshnikov et al. (2008).

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.

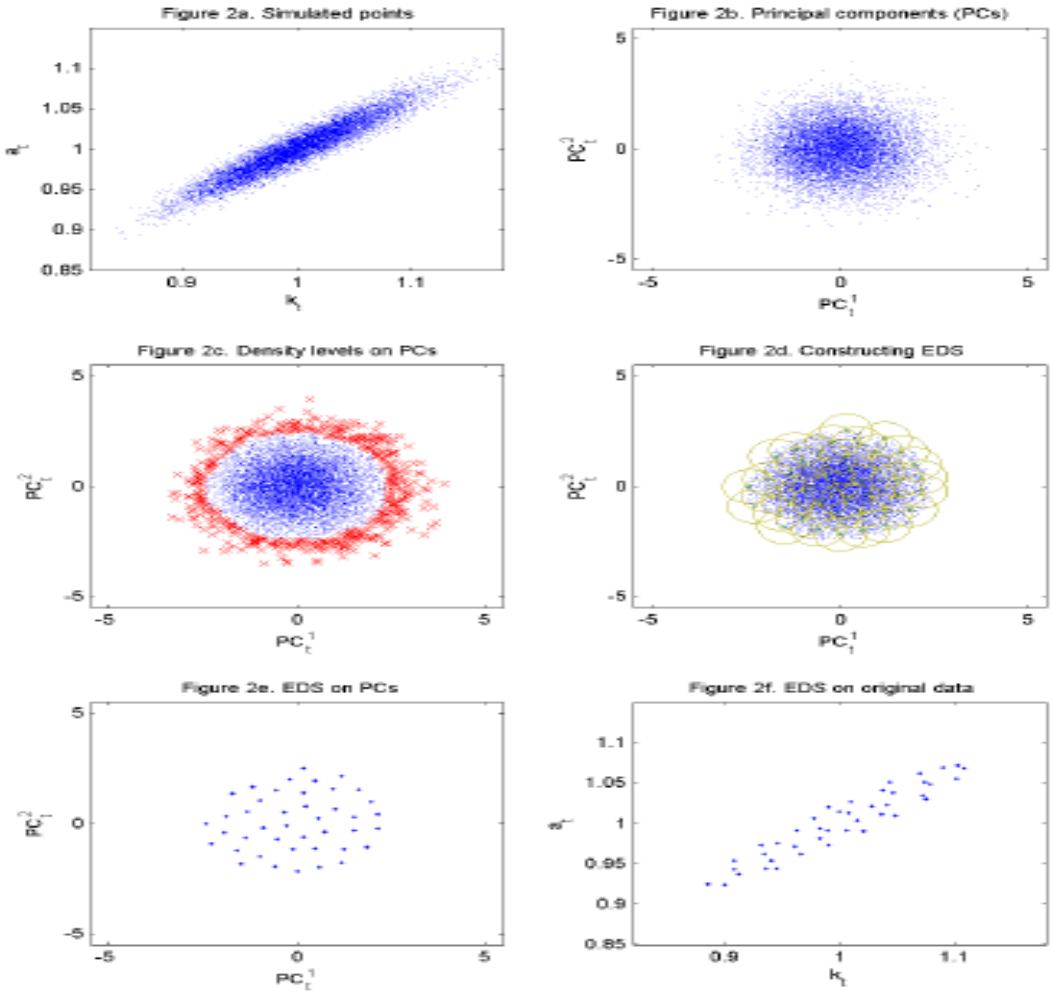
Principal component transformation

- 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 $\mathrm{PC} \equiv XV$.
- $\mathrm{PC} = \left(\mathrm{PC}^1, ..., \mathrm{PC}^d\right) \in \mathbb{R}^{n \times d}$ are *principal components* (PCs) of X, and are orthogonal (uncorrelated), i.e., $\left(\mathrm{PC}^{\ell'}\right)^{\top} \mathrm{PC}^{\ell} = 0$ for any $\ell' \neq \ell$.
- Distance between two observations x_i and x_j is the Euclidean distance between their PCs

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

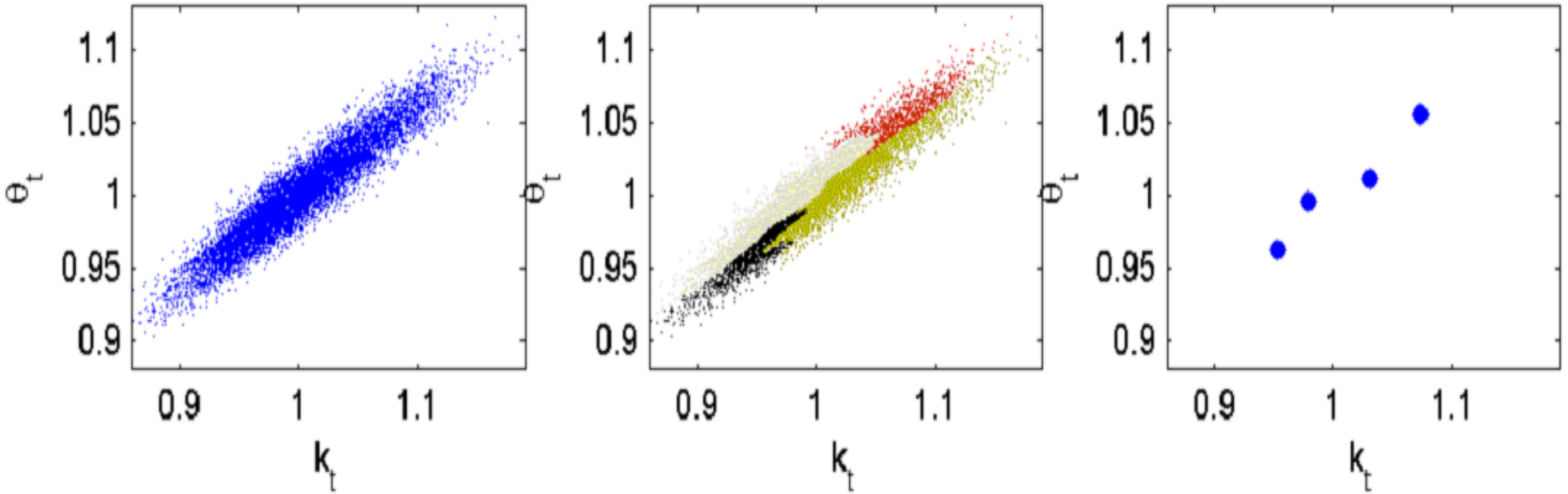
where PC^1, ...,PC^d are normalized to unit variance.

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.



Dispersion of points in the EDS

Def. *Let P be a set consisting of points $x_1, ..., x_n \in X \subseteq \mathbb{R}^d$, and let (X, D) be a bounded metric space. The dispersion of P in X is given by*

$$d_n(P; X) = \sup_{x \in X} \inf_{1 \leq i \leq n} D(x, x_i), \tag{1}$$

where D is a (Euclidean) metric on X.

Def. *Let S be a sequence of elements on X, and let $x_1, ..., x_n \in X \subseteq \mathbb{R}^d$ be the first n terms of S. The sequence S is called low-dispersion if $\lim_{n \to \infty} d_n(S; X) = 0$.*

Proposition. Let P be any set of n points $x_1, ..., x_n \in X \subseteq \mathbb{R}^d$ with a dispersion $d_n(P; X) < \varepsilon$. Let (X, D) be a bounded metric space, and let P^ε be an EDS $x_1^\varepsilon, ..., x_M^\varepsilon$ constructed by Algorithm P^ε. Then, the dispersion of P^ε is bounded by $\varepsilon < d_M(P^\varepsilon; X) < 2\varepsilon$.

Number of points in the EDS

Proposition. Let P be any set of n points $x_1, \ldots, x_n \in B(0, r) \subseteq \mathbb{R}^d$ with a dispersion $d_n(P; X) < \varepsilon$. Then, the number of points in P^ε constructed by Algorithm P^ε is bounded by $\left(\frac{r}{2\varepsilon}\right)^d \leq M \leq \left(1 + \frac{r}{\varepsilon}\right)^d$.

To construct an EDS with a given target number of points $\overline{M}$, we use a simple *bisection method*:

- fix ε_1 and ε_2 such that $M(\varepsilon_1) \leq \overline{M} \leq M(\varepsilon_2)$,
- take $\varepsilon = \frac{\varepsilon_1 + \varepsilon_2}{2}$, construct an EDS and find $M(\varepsilon)$;
- if $M(\varepsilon) > \overline{M}$, set $\varepsilon_1 = \varepsilon$ and otherwise, set $\varepsilon_2 = \varepsilon$, and proceed iteratively until $M(\varepsilon)$ converges to some limit.
- To find the initial values of ε_1 and ε_2, we use the bounds established in the above proposition.

Discrepancy

The degree of uniformity of EDSs. Standard notion of uniformity in the literature – discrepancy from the uniform distribution.

Def. *Let P be a set consisting of points $x_1, ..., x_n \in X \subseteq \mathbb{R}^d$, and let $\mathcal{J}$ be a family of Lebesgue-measurable subsets of X. The discrepancy of P under $\mathcal{J}$ is given by $\mathcal{D}_n(P; \mathcal{J}) = \sup_{J \in \mathcal{J}} \left| \frac{C(P;J)}{n} - \lambda(J) \right|$, where $C(P; J)$ counts the number of points from P in J, and $\lambda(J)$ is a Lebesgue measure of J.*

Proposition. Let P be any set of n points $x_1, ..., x_n \in B(0; 1) \subseteq \mathbb{R}^d$ with a dispersion $d_n(P; X) < \varepsilon$. Then, the discrepancy of an EDS constructed by Algorithm P^ε under B is bounded by $D_M(P^\varepsilon; \mathcal{B}) \leq \frac{\sqrt{2^d}-1}{\sqrt{2^d}+1}$.

Existence results for a covering-number problem

- Temlyakov (2011) studies the problem of finding a covering number – a minimum number of balls of radius ε which cover a given compact set (such as a d-dimensional hypercube or hypersphere).
- He shows that there exists an EDS P^{ε} on a unit hypercube $[0,1]^d$ whose discrepancy converges to 0 as $M \longrightarrow \infty$ (i.e., $\varepsilon \longrightarrow 0$).
- However, constructing such an EDS is operationally difficult and costly.
- Also, Temlyakov (2011) selects points from a compact subset of $\mathbb{R}^d$, and his analysis cannot be directly applied to our problem of finding an ε-distinguishable subset of a given finite set of points.

Probabilistic results: random sequential packing problems

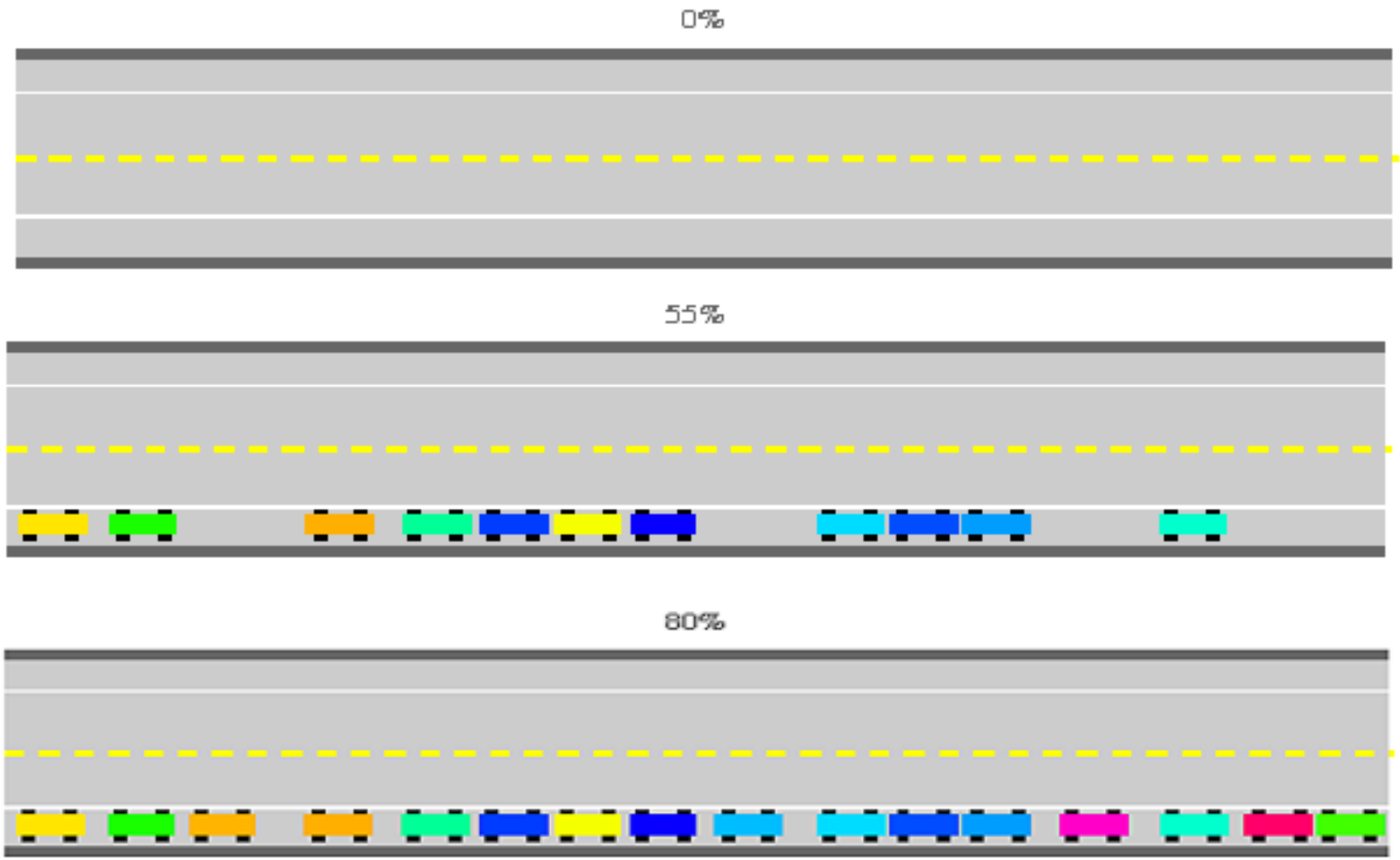
Probabilistic analysis of an EDS is non-trivial as points are spatially dependent: once we place a point in an EDS, it affects the placement of all subsequent points.

A random sequential packing problem:

- consider a bounded set $X \subseteq \mathbb{R}^d$ and a sequence of d-dimensional balls whose centers are i.i.d. random vectors $x_1, ..., x_n \in X$ with a given density function g.
- A ball is packed if and only if it does not overlap with any ball which has already been packed. If not packed, the ball is discarded. At saturation, the centers of accepted balls constitute an EDS.

Probabilistic results: random sequential packing problems

Rényi's (1958) car parking model



Probabilistic results: random sequential packing problems

Rényi's (1958) car parking model

Probabilistic results for random sequential packing problems

- For a multidimensional case, Baryshnikov et al. (2008) show that the sequential packing measure, induced by the accepted balls centers, satisfies the LIL.
- Thus, the discrepancy of EDS converges to 0 asymptotically if the density of points in an EDS is uniform in the limit $\varepsilon \longrightarrow 0$. However, the density of points in an EDS depends on the density function g of the stochastic process used to produce the data.
- Hence, an EDS needs not be uniform in the limit even in the probabilistic sense (unless the density function is uniform).

Our best- and worst-case scenarios

Implications of our analysis for Rényi's (1958) car parking model.
The best- and worst-case scenarios: cars occupy between 50% and 100% of the roadside ($\frac{1}{2} \leq \lim_{\varepsilon \to 0} M\varepsilon \leq 1$).

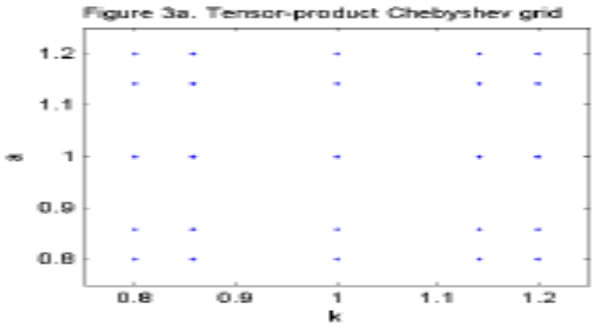
1. Distance ε between cars: evil drivers park their cars to leave as little parking space to other drivers as possible
2. Distance 0 between cars: a police officer directs the cars to park in a socially efficient way).

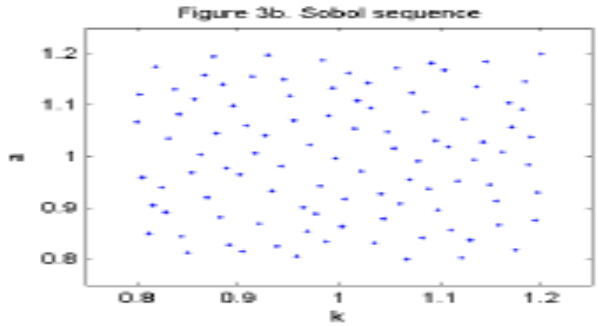
Our best- and worst-case scenarios

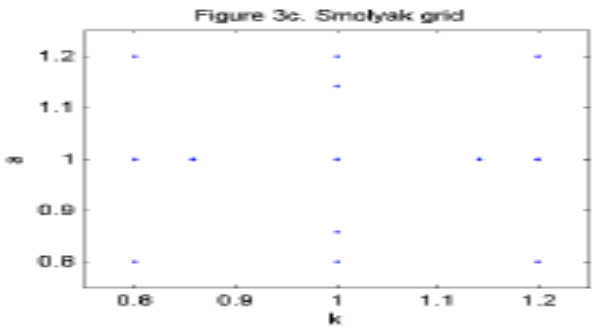
The worst-case scenario for discrepancy in Rényi's (1958) model, $\mathcal{D}_M^s\left(P^\varepsilon;\mathcal{B}\right) \leq \frac{\sqrt{2}-1}{\sqrt{2}+1} \approx 0.17$, which is obtained under $\lambda^* = \frac{\sqrt{2}}{\sqrt{2}+1}$.

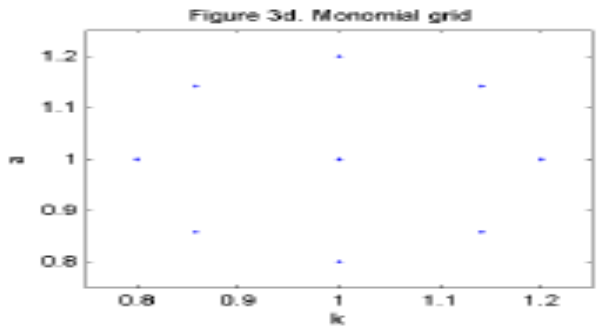
1. To attain this bound, consider an EDS on $[0,1]$ such that on the interval $[0,\lambda^*]$, all points are situated on a distance 2ε, and on $[\lambda^*,1]$, all points are situated on the distance ε.
2. In the first interval, we have $\frac{\lambda^*}{2\varepsilon} \leq M \leq \frac{\lambda^*}{2\varepsilon}+1$ points and in the second interval, we have $\frac{1-\lambda^*}{\varepsilon} \leq M \leq \frac{1-\lambda^*}{\varepsilon}+1$ points.
3. On the first interval, the limiting discrepancy is $\lim_{\varepsilon\to 0}\left[\lambda^* - \frac{\frac{\lambda^*}{2\varepsilon}}{\frac{\lambda^*}{2\varepsilon}+\frac{1-\lambda^*}{\varepsilon}}\right] = \frac{\sqrt{2}-1}{\sqrt{2}+1} \approx 0.17$, the same value as implied by our propositions.

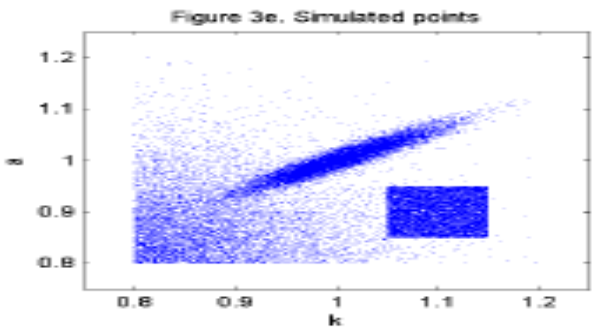
Comparison of EDS grid with other grids in the literature

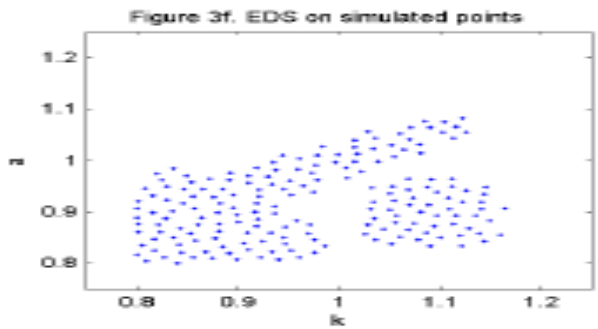












Deterministic integration methods

Our EDS relies on accurate Gauss-Hermite quadrature integration

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

where $\{\epsilon_j\}_{j=1}^{J}$ = 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, $\omega_1 = 1$.

Multidimensional Gauss Hermite product rules

In multidimensional problem, we can use Gauss Hermite product rules.

Example

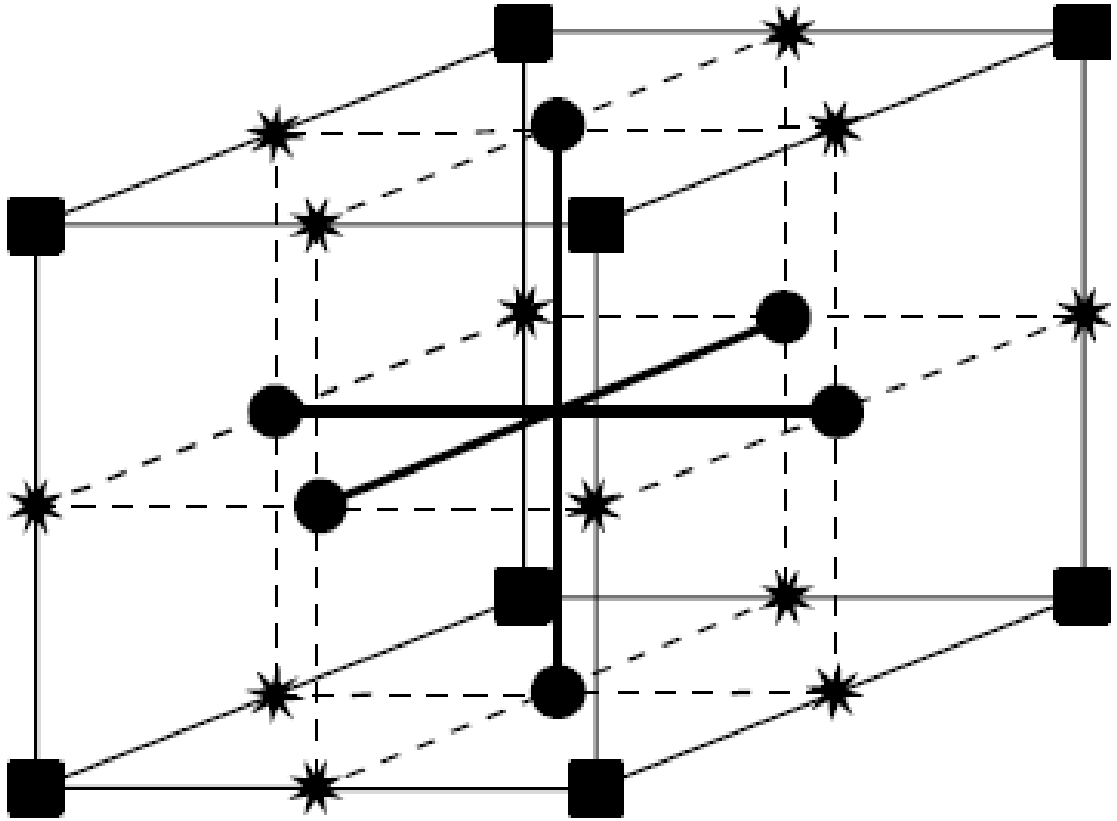
Let $\epsilon_{t+1}^{h} \sim \mathcal{N}\left(0, \sigma^2\right)$, $h = 1, 2, 3$ be uncorrelated random variables. A two-node Gauss-Hermite product rule, $Q\left(2\right)$, (obtained from the two-node Gauss-Hermite rule) has 2^3 nodes, which are as follows:

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

Non-product 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.

Solving for polynomial coefficients: fixed-point iteration

Derivative-free solvers

- The cost of Newton's type method grows quickly with dimensionality because of the growing number of terms in Jacobian and Hessian.
- A simple and efficient alternative is fixed-point iteration

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

  where $\xi \in (0, 1)$ is damping parameter.
- Cost of fixed-point iteration grows little with dimensionality.
- Fixed-point iteration works for very high dimensions, like 400 state variables!

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

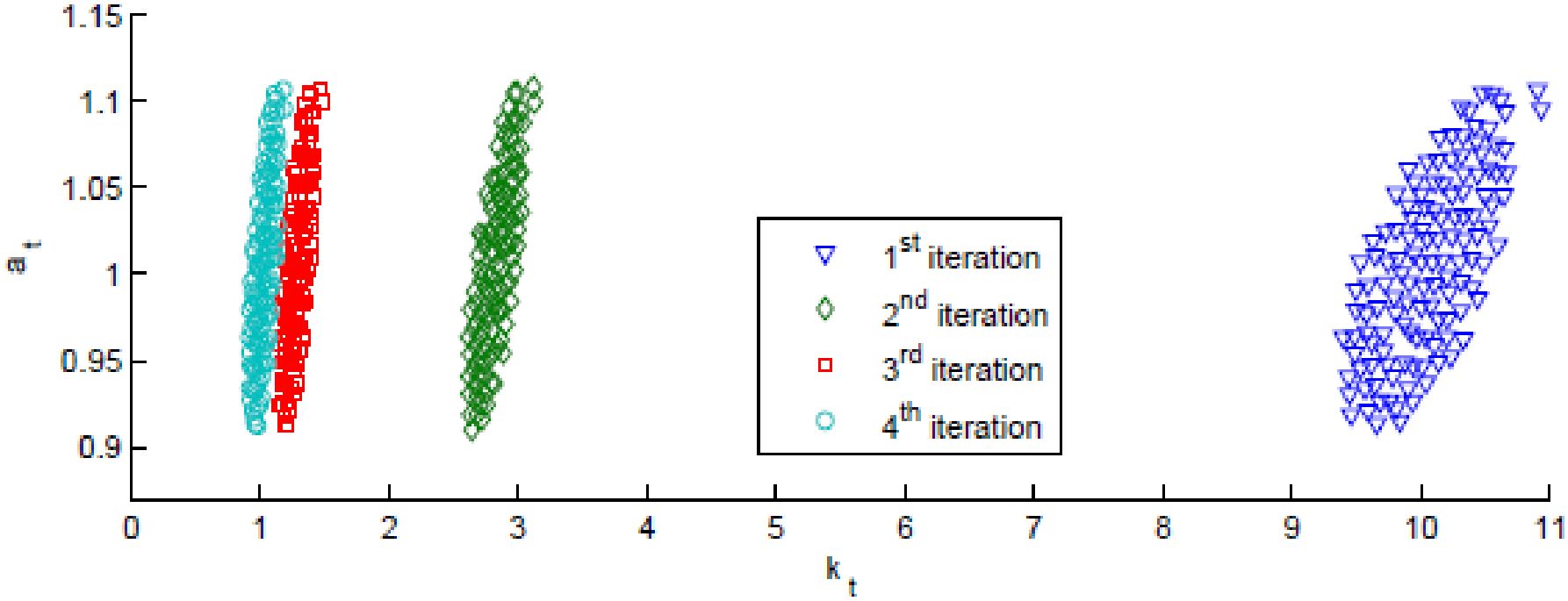


Table 2: Accuracy and speed in the one-agent model: Smolyak grid versus EDS grid

	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

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

- *Literature that estimates the models:*
  -Christiano, Eichenbaum and Evans (2005), Smets and Wouters (2003, 2007), Del Negro, Schorfheide, Smets and Wouters (2007).
- *Literature that finds numerical solutions:* mostly relies on local (perturbation) solution methods. Few papers apply global solution methods to low-dimensional problems.
- *Perturbation*:
  -most use linear approximations (Christiano, Eichenbaum&Rebelo, 2009);
  -some use quadratic approx. (Kollmann, 2002, Schmitt-Grohé&Uribe, 2007);
  -very few use cubic approximations (Rudebusch and Swanson, 2008).
- *Global solution methods*: at most 4 state variables and simplifying assumptions.
  -Adam and Billi (2006): all except one FOCs are linearized;
  -Adjemian and Juillard (2011): extended path method of Fair&Taylor (1984) $\Rightarrow$ perfect foresight.

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).

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

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

$$\text{s.t. } Y_t = \left(\int_0^1 Y_t(i)^{\frac{\varepsilon-1}{\varepsilon}} di \right)^{\frac{\varepsilon}{\varepsilon-1}}, \qquad \varepsilon \geq 1 \tag{2}$$

– $Y_t(i)$ 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 (2) = 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:

- Monopolisticly 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:

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

$$\text{s.t.} \quad Y_t(i) = Y_t \left(\frac{P_t(i)}{P_t} \right)^{-\varepsilon} \tag{3}$$

– Eq (3) 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,

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

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

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

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

- **Aggregate production**

$$Y_t = \exp\left(\eta_{a,t}\right) L_t \Delta_t$$

- **Aggregate resource constraint**

$$C_t + G_t = Y_t$$

- **Taylor rule with ZLB on the net nominal interest rate**

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

- **Natural level of output**

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

Summary of equilibrium conditions

We have

- Stochastic processes for 6 exogenous shocks $\{\eta_{u,t}, \eta_{L,t}, \eta_{B,t}, \eta_{a,t}, \eta_{G,t}, \eta_{R,t}\}$.
- 8 endogenous equilibrium equations & 8 unknowns $\{C_t, Y_t, R_t, L_t, \Delta_t, \pi_t, F_t, S_t\}$.
- 2 endogenous state variables $\{\Delta_{t-1}, R_{t-1}\}$.
- 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

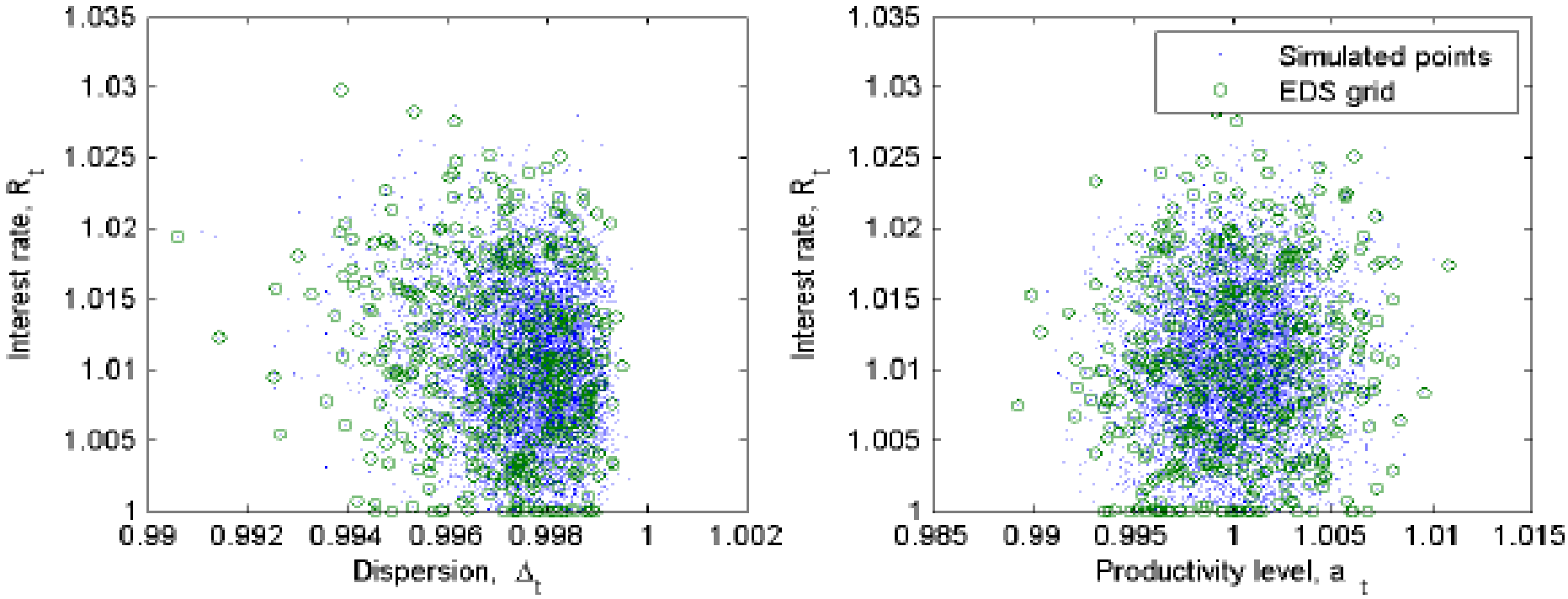


Table 4. Accuracy and speed in the NK model with 0% inflation target and 18.21% volatility of labor shock

	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

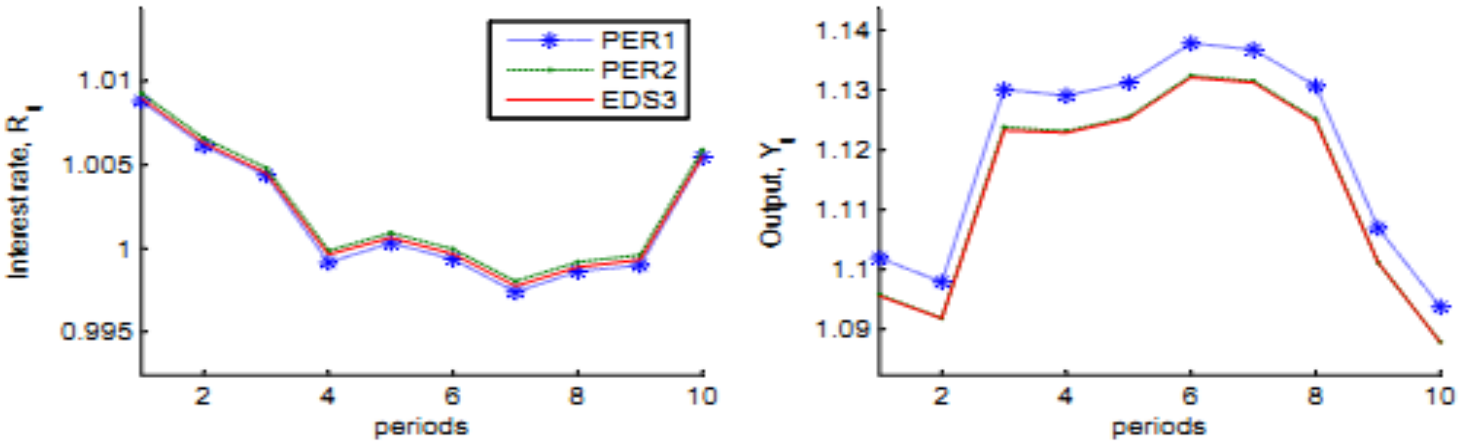
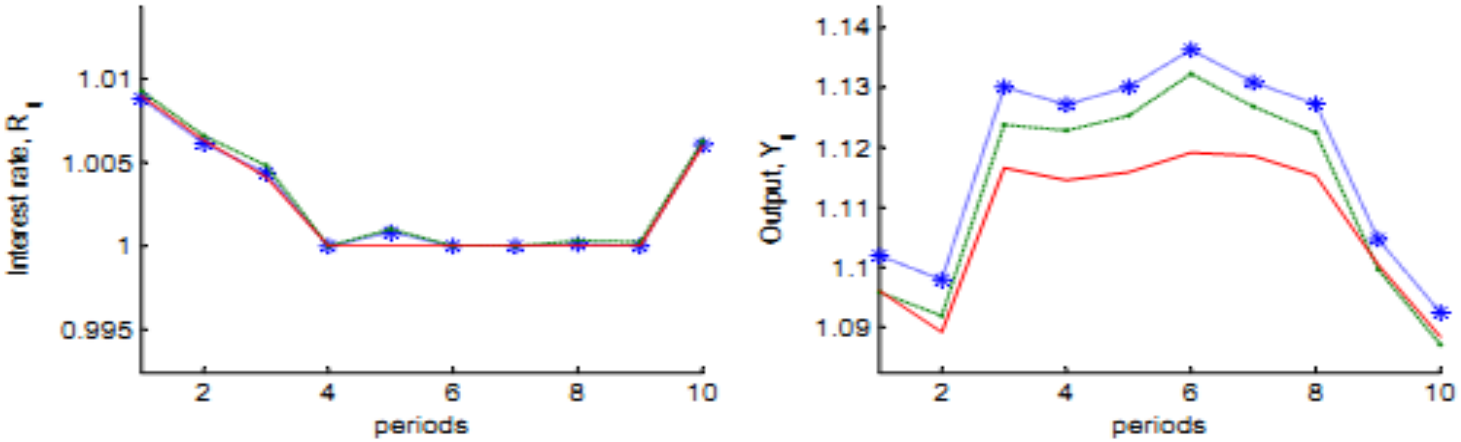


Figure 6b. A time-series solution to a new Keynesian model: ZLB is imposed



Conclusion

- The EDS algorithm accurately solves models that were considered to be unfeasible until now.
- A mix of techniques taken together allows us to address the challenges of high-dimensional problems:
  - EDS grid domain - a tiny fraction of the standard hypercube domain;
  - monomial and one-node integration rules;
  - fixed-point iteration for finding policy functions;
  - iteration-on-allocation and precomputation approaches for solving for intratemporal choice.
- A proper coordination of the above techniques is crucial for accuracy and speed.

Equilibrium conditions

Some derivations:

FOC of a reoptimizing intermediate-good firm with respect to $\widetilde{P}_t$ is

$$E_t \sum_{j=0}^{\infty} (\beta\theta)^j \Lambda_{t+j} Y_{t+j} P_{t+j}^{\varepsilon+1} \left[\frac{\widetilde{P}_t}{P_{t+j}} - \frac{\varepsilon}{\varepsilon - 1} \text{mc}_{t+j} \right] = 0$$

where $\Lambda_{t+j} = \frac{\exp\left(\eta_{u,t+j}\right) C_{t+j}^{-\gamma}}{P_{t+j}}$ follows from the household's FOC.

How should we deal with this infinite sum when solving the model?
Need a recursive representation of this FOC.

Equilibrium conditions

- **Step 1:** Define $\chi_{t,j}$

$$\chi_{t,j} \equiv \begin{cases} 1 \text{ if } j = 0 \\ \frac{1}{\pi_{t+j} \cdot \pi_{t+j-1} \cdots \pi_{t+1}} \text{ if } j \geq 1 \end{cases}.$$

Then $\chi_{t,j} = \chi_{t+1,j-1} \cdot \frac{1}{\pi_{t+1}}$ for $j > 0$. Therefore, FOC with the infinite sum becomes

$$E_t \sum_{j=0}^{\infty} (\beta\theta)^j \exp\left(\eta_{u,t+j}\right) C_{t+j}^{-\gamma} Y_{t+j} \chi_{t,j}^{-\varepsilon} \left[\frac{\widetilde{P}_t}{P_t} \chi_{t,j} - \frac{\varepsilon}{\varepsilon - 1} \text{mc}_{t+j}\right] = 0$$

- **Step 2:** Rewrite the above FOC as

$$\frac{\widetilde{P}_t}{P_t} = \frac{E_t \sum_{j=0}^{\infty} (\beta\theta)^j \exp\left(\eta_{u,t+j}\right) C_{t+j}^{-\gamma} Y_{t+j} \chi_{t,j}^{-\varepsilon} \frac{\varepsilon}{\varepsilon-1} \text{mc}_{t+j}}{E_t \sum_{j=0}^{\infty} (\beta\theta)^j \exp\left(\eta_{u,t+j}\right) C_{t+j}^{-\gamma} Y_{t+j} \chi_{t,j}^{1-\varepsilon}} \equiv \frac{S_t}{F_t}$$

Equilibrium conditions

- **Step 3:** Find recursive formulas for S_t and F_t. For example, for S_t,

$$S_t \equiv E_t \sum_{j=0}^{\infty} (\beta\theta)^j \exp\left(\eta_{u,t+j}\right) C_{t+j}^{-\gamma} Y_{t+j} \chi_{t,j}^{-\varepsilon} \frac{\varepsilon}{\varepsilon - 1} \text{mc}_{t+j}$$

$$= \frac{\varepsilon}{\varepsilon - 1} \exp\left(\eta_{u,t}\right) C_t^{-\gamma} Y_t \text{mc}_t$$

$$+\beta\theta E_t \left\{ \sum_{j=1}^{\infty} (\beta\theta)^{j-1} \exp\left(\eta_{u,t+j}\right) C_{t+j}^{-\gamma} Y_{t+j} \left(\frac{\chi_{t+1,j-1}}{\pi_{t+1}} \right)^{-\varepsilon} \frac{\varepsilon}{\varepsilon - 1} \text{mc}_{t+j} \right\}$$

$$= \frac{\varepsilon}{\varepsilon - 1} \exp\left(\eta_{u,t}\right) C_t^{-\gamma} Y_t \text{mc}_t + \beta\theta E_t \left\{ \frac{1}{\pi_{t+1}^{-\varepsilon}} \times \right.$$

$$\left. \times E_{t+1} \left(\sum_{j=0}^{\infty} (\beta\theta)^j \exp\left(\eta_{u,t+1+j}\right) C_{t+1+j}^{-\gamma} Y_{t+1+j} \chi_{t+1,j}^{-\varepsilon} \frac{\varepsilon}{\varepsilon - 1} \text{mc}_{t+1+j} \right) \right\}$$

$$= \frac{\varepsilon}{\varepsilon - 1} \exp\left(\eta_{u,t}\right) C_t^{-\gamma} Y_t \text{mc}_t + \beta\theta E_t \left\{ \pi_{t+1}^{\varepsilon} S_{t+1} \right\}$$

Aggregate price relationship

- Aggregate price index:

$$P_t \equiv \left(\int_0^1 P_t(i)^{1-\varepsilon} di \right)^{\frac{1}{1-\varepsilon}} =$$

$$\left[\int_{\text{reopt.}} P_t(i)^{1-\varepsilon} di + \int_{\text{non-reopt.}} P_t(i)^{1-\varepsilon} di \right]^{\frac{1}{1-\varepsilon}}$$

  – "reopt." and "non-reopt." denote, resp., the firms that reoptimize and do not reoptimize their prices at t.

- Easy part (all reoptimizers choose the same price):

$$\int_{\text{reopt.}} P_t(i)^{1-\varepsilon} di = (1-\theta) \widetilde{P}_t^{1-\varepsilon}$$

Aggregate price relationship

- Harder part:

$$\int_{\text{non-reopt.}} P_t(i)^{1-\varepsilon} di = \int_0^1 P(j)^{1-\varepsilon} \omega_{t-1,t}(j) dj$$

where $\omega_{t-1,t}(j)$ = measure of non-reoptimizers at t that had $P(j)$ at $t-1$.

Note $\omega_{t-1,t}(j) = \theta \omega_{t-1}(j)$, where $\omega_{t-1}(j)$ = measure of firms with $P(j)$ in $t-1$.

$$\int_{\text{non-reopt.}} P_t(i)^{1-\varepsilon} di = \int_0^1 \theta P(j)^{1-\varepsilon} \omega_{t-1}(j) dj = \theta P_{t-1}^{1-\varepsilon}$$

- Combine the easy and hard parts:

$$P_t = \left[(1-\theta)\widetilde{P}_t^{1-\varepsilon} + \theta P_{t-1}^{1-\varepsilon}\right]^{\frac{1}{1-\varepsilon}}$$

- Rewrite

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

Aggregate output

$$\overline{Y}_t \equiv \int_0^1 Y_t(i)\,di = \int_0^1 \exp\left(\eta_{a,t}\right) L_t(i)\,di = \exp\left(\eta_{a,t}\right) L_t$$

- Use the demand for $Y_t(i)$

$$\overline{Y}_t = \int_0^1 Y_t \left(\frac{P_t(i)}{P_t}\right)^{-\varepsilon} di = Y_t P_t^{\varepsilon} \int_0^1 P_t(i)^{-\varepsilon}\,di = Y_t P_t^{\varepsilon} \left(\overline{P}_t\right)^{-\varepsilon}$$

where

$$\left(\overline{P}_t\right)^{-\varepsilon} \equiv \int_0^1 P_t(i)^{-\varepsilon}\,di$$

- Combine to get

$$Y_t \equiv \overline{Y}_t \left(\frac{\overline{P}_t}{P_t}\right)^{\varepsilon} = \exp\left(\eta_{a,t}\right) L_t \Delta_t \qquad \text{with } \Delta_t \equiv \left(\frac{\overline{P}_t}{P_t}\right)^{\varepsilon}$$

  – Δ_t = measure of price dispersion across firms.

- If $P_t(i) = P_t(i')$ for all i and $i' \in [0,1] \Longrightarrow \Delta_t = 1$.

Law of motion for price distortion

- By analogy with the aggregate price index P_t, define

$$\overline{P}_t \equiv \left[(1-\theta) \widetilde{P}_t^{-\varepsilon} + \theta \left(\overline{P}_{t-1} \right)^{-\varepsilon} \right]^{-\frac{1}{\varepsilon}}$$

- Use $\overline{P}_t$ in the definition of Δ_t

$$\Delta_t = \left(\frac{\left[(1-\theta) \widetilde{P}_t^{-\varepsilon} + \theta \left(\overline{P}_{t-1} \right)^{-\varepsilon} \right]^{-\frac{1}{\varepsilon}}}{P_t} \right)^{\varepsilon}$$

- This implies the law of motion for Δ_t,

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

Aggregate resource constraint

- Combine the household's BC with the government BC

$$P_t C_t + P_t \frac{\overline{G} Y_t}{\exp\left(\eta_{G,t}\right)} = (1 - v) W_t L_t + \Pi_t$$

- Note that the ith intermediate-good firm's profit at t is $\Pi_t(i) \equiv P_t(i) Y_t(i) - (1 - v) W_t L_t(i)$.
- Thus, since $\Pi_t \equiv \int_0^1 \Pi_t(i)\, di$,

$\Pi_t = \int_0^1 P_t(i) Y_t(i)\, di - (1 - v) W_t \int_0^1 L_t(i)\, di = P_t Y_t - (1 - v) W_t L_t$ where $P_t Y_t = \int_0^1 P_t(i) Y_t(i)\, di$ follows by a zero-profit condition of the final-good firms.

- Hence, $P_t C_t + P_t \frac{\overline{G}}{\exp\left(\eta_{G,t}\right)} Y_t = P_t Y_t$.
- In real terms,

$$C_t = \left(1 - \frac{\overline{G}}{\exp\left(\eta_{G,t}\right)}\right) Y_t$$

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**

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