Numerically Stable and Accurate Stochastic Simulation Approaches for Solving Dynamic Economic Models

Kenneth Judd, Lilia Maliar and Serguei Maliar

April 17, 2011

Introduction

- Dynamic models are used nowadays in all fields and areas of economics: macroeconomics, finance, international trade, labor economics, etc.
- A characteristic feature of such models is that they do not admit closed-form (analytical) solutions.
- Many interesting models (e.g., heterogeneous agents, sectors, countries) cannot be solved with the existing methods in spite of modern high-speed computers.

Introduction

- *In this paper:* we describe a class of numerical methods for solving dynamic stochastic models based on stochastic simulation.
- *Stochastic simulation methods* we propose are (i) simpler, (ii) faster and (iii) more accurate than those existing in the previous literature.
- Stochastic simulation methods are especially attractive for high dimensional applications, where other methods are not feasible.

  Example, dynamic stochastic models with 20 heterogeneous countries and 40 state variables.

Background

Three broad classes of numerical methods

1. **Projection methods**; Judd (1992), Christiano and Fisher (2000).
2. **Perturbation methods;** Judd and Guu (1993), Collard and Juillard (2001).
3. **Stochastic-simulation methods;** den Haan and Marcet (1990), Smith (1991).

Illustrative example

A one-sector neoclassical growth model:

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

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

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

where $\epsilon_{t+1} \sim \mathcal{N}\left(0, \sigma^2\right)$; and initial condition $\left(k_0, a_0\right)$ is given;
$u\left(\cdot\right)$ = utility function; $f\left(\cdot\right)$ = production function;
c_t = consumption; k_{t+1} = capital; a_t = productivity;
δ = discount factor; d = depreciation rate of capital;
ρ = autocorrelation coefficient of the productivity level;
σ = standard deviation of the productivity shock.

Projection methods

Characteristic features

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

A projection method for the growth model

- Choose a grid of I points in the state space $\{k_i, a_i\}_{i=1}^{I}$.
- Parameterize the capital policy function by a polynomial

$$k_i' \simeq \Psi(k_i, a_i; \beta) = \beta_0 + \beta_1 k_i + \beta_2 a_i + \beta_3 k_i^2 + \beta_4 a_i^2 + ...$$

and substitute it in the Euler equation to get

$$\min_{\beta} \left\| u_1(c_i(k_i, a_i; \beta)) - E\left\{\delta u_1\left(c_i'(k_i, a_i; \beta)\right)\left[1 - d + a_i' f'\left(\Psi(k_i, a_i; \beta)\right)\right]\right\}\right\|$$

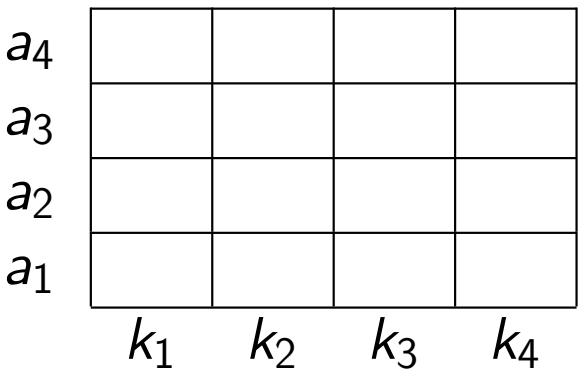
where $\beta \equiv (\beta_0, \beta_1, ...)$ is a vector of coefficients and

$$\begin{aligned} a_i' &= a_i^{\rho} \exp(\epsilon) \\ k_i'' &= \Psi\left(\Psi(k_i, a_i; \beta), a_i'; \beta\right) \\ c_i(k_i, a_i; \beta) &= (1-d) k_i + a_i f(k_i) - k_i' \\ c_i'(k_i, a_i; \beta) &= (1-d) k_i' + a_i' f(k_i') - k_i'' \end{aligned}$$

- Find a vector of coefficients β that minimizes the distance using a Newton's type of solver (this procedure involves evaluating conditional expectations, i.e., numerical integration).

Projection methods: curse of dimensionality

- Very accurate and fast with few state variables but cost grows exponentially with dimensionality!
  (a) Product hypercube domain $\Longrightarrow$ 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 $\Rightarrow 100^{10} = 10^{20}$).

- *Kruger and Kubler (2004):* Smolyak's sparse grid - efficient grid with relatively small number of points in a multidimensional hypercube.

Perturbation methods

Characteristic features

- Compute a solution in just one point (steady state).
- Identify polynomial coefficients of policy functions using k-order Taylor's expansion of the optimality conditions.

Perturbation methods: a log-linearization example

- *Log-linearization* - first-order Taylor's expansion, e.g.,

$$u'(c_t) \simeq u'(c) + u''(c)\, c \frac{(c_t - c)}{c} = u'(c) + u''(c)\, c \widehat{c}_t$$

  where $\widehat{c}_t = \frac{c_t - c}{c}$ = log-deviation of c_t from the steady state c.
- Substitute $\widehat{c}_t$ and $\widehat{k}_t = \frac{k_t - k}{k}$ in the optimality conditions to get a linearized system of equations.
- Postulate specific log-linear form for policy functions $c_t = C(k_t, a_t)$ and $k_t = K(k_t, a_t)$:

$$\widehat{k}_{t+1} = \zeta_{kk} \widehat{k}_t + \zeta_{ka} \widehat{a}_t, \qquad \widehat{c}_t = \zeta_{ck} \widehat{k}_t + \zeta_{ca} \widehat{a}_t$$

  where ζ_{kk}, ζ_{ka}, ζ_{ck} and ζ_{ca} = coefficients to be determined.
- Solve the obtained system of equations $\Longrightarrow$ identify the coefficients ζ_{kk}, ζ_{ka}, ζ_{ck} and ζ_{ca}.

Perturbation methods of higher orders

- **Perturbation is a Taylor's expansion performed numerically.** It is a generalization of the (first-order) log-linearization method to higher orders.
- **Perturbation methods are very fast but the range of their accuracy is uncertain.** This is a local approximation, and the accuracy might deteriorate dramatically away from the steady state.

  JEDC 2011 comparison results: 1st- and 2nd-order perturbation methods, PER1 and PER2, of Kollmann, Kim and Kim (2011) produce errors:

  - *on a stochastic simulation up to 6.3% and 1.4%, respectively;*
  - *on a 30% deviation from steady state up to 65% and 50%, respectively.*

- **These accuracy levels are not acceptable**: a method that produces errors of 6% per quarter in the US GDP is not satisfactory.

Characteristic features

- Compute a solution on simulated series.
- Use Monte Carlo integration for approximating conditional expectations.
- Main steps:
  - *Step* 1. Guess a policy function.
  - *Step* 2. Simulate time series.
  - *Step* 3. Use simulation results to check and to update the guess.
  - Iterate on *Steps* $2-3$ until convergence.

A stochastic simulation method for the growth model

- Parameterize the capital policy function by a polynomial

$$k_{t+1} \simeq \Psi\left(k_t, a_t; \beta\right) = \beta_0 + \beta_1 k_t + \beta_2 a_t + \beta_3 k_t^2 + \beta_4 a_t^2 + ...$$

  and substitute it into the budget constraint to get

$$c_t = \left(1 - d\right) k_t + a_t f\left(k_t\right) - \Psi\left(k_t, a_t; \beta\right).$$

- Fix $\beta \equiv (\beta_0, \beta_1, ...)$. Given shocks $\{a_t\}_{t=0}^T$, simulate $\{c_t, k_{t+1}\}_{t=0}^T$ and construct

$$y_t \equiv \delta \frac{u_1\left(c_{t+1}\right)}{u_1\left(c_t\right)} \left[1 - d + a_{t+1} f'\left(k_{t+1}\right)\right] k_{t+1}.$$

- Regress y_t on $\left(1, k_t, a_t, k_t^2, a_t^2, ...\right) \Longrightarrow$ get $\widehat{\beta}$ (Monte Carlo integration).
- Compute the next-iteration input $\beta^{(j+1)}$ as

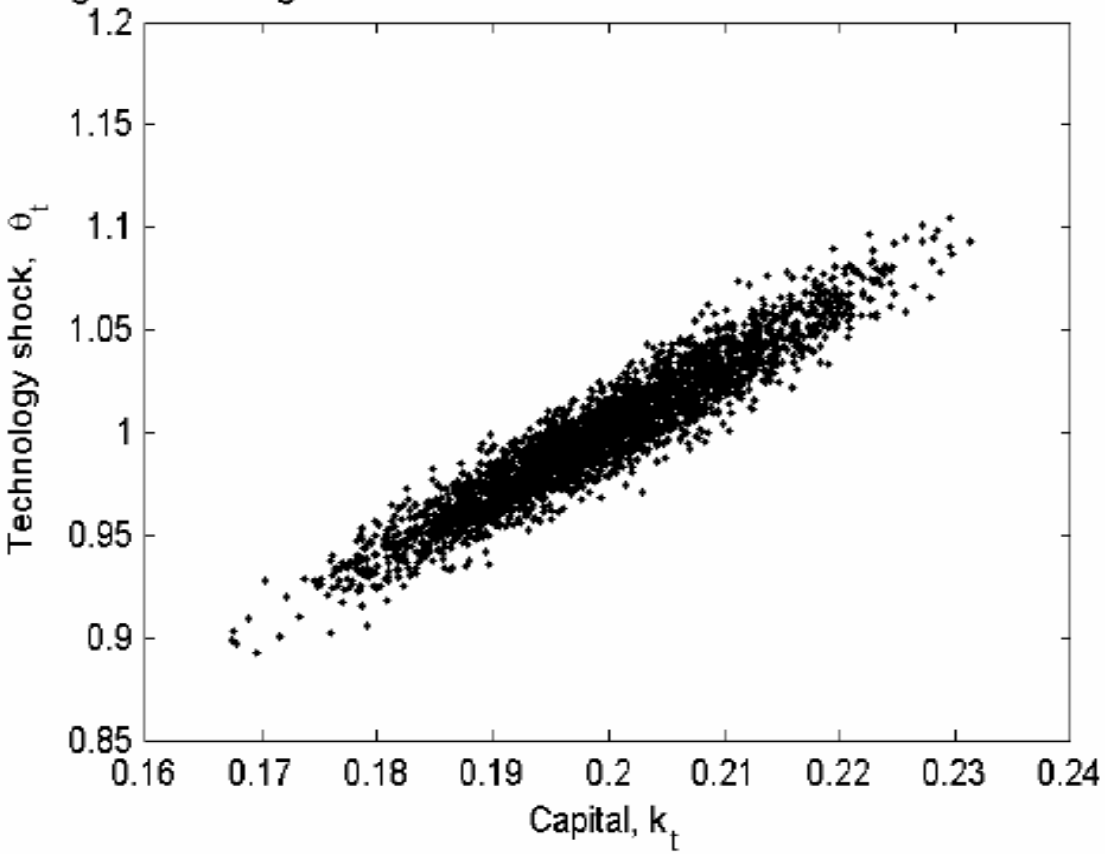
$$\beta^{(j+1)} = \left(1 - \mu\right) \beta^{(j)} + \mu \widehat{\beta},$$

  where $\mu \in (0, 1] =$ damping parameter.

Key advantage of stochastic simulation methods

Stochastic simulation methods have endogenous solution domain: the areas of the state space that are visited in simulation (the ergodic set).
Recall that for projection and perturbation methods: the domain is an exogenous rectangular grid and the steady state point, respectively.

Figure 1. The ergodic distribution in the model with a closed-form solution.



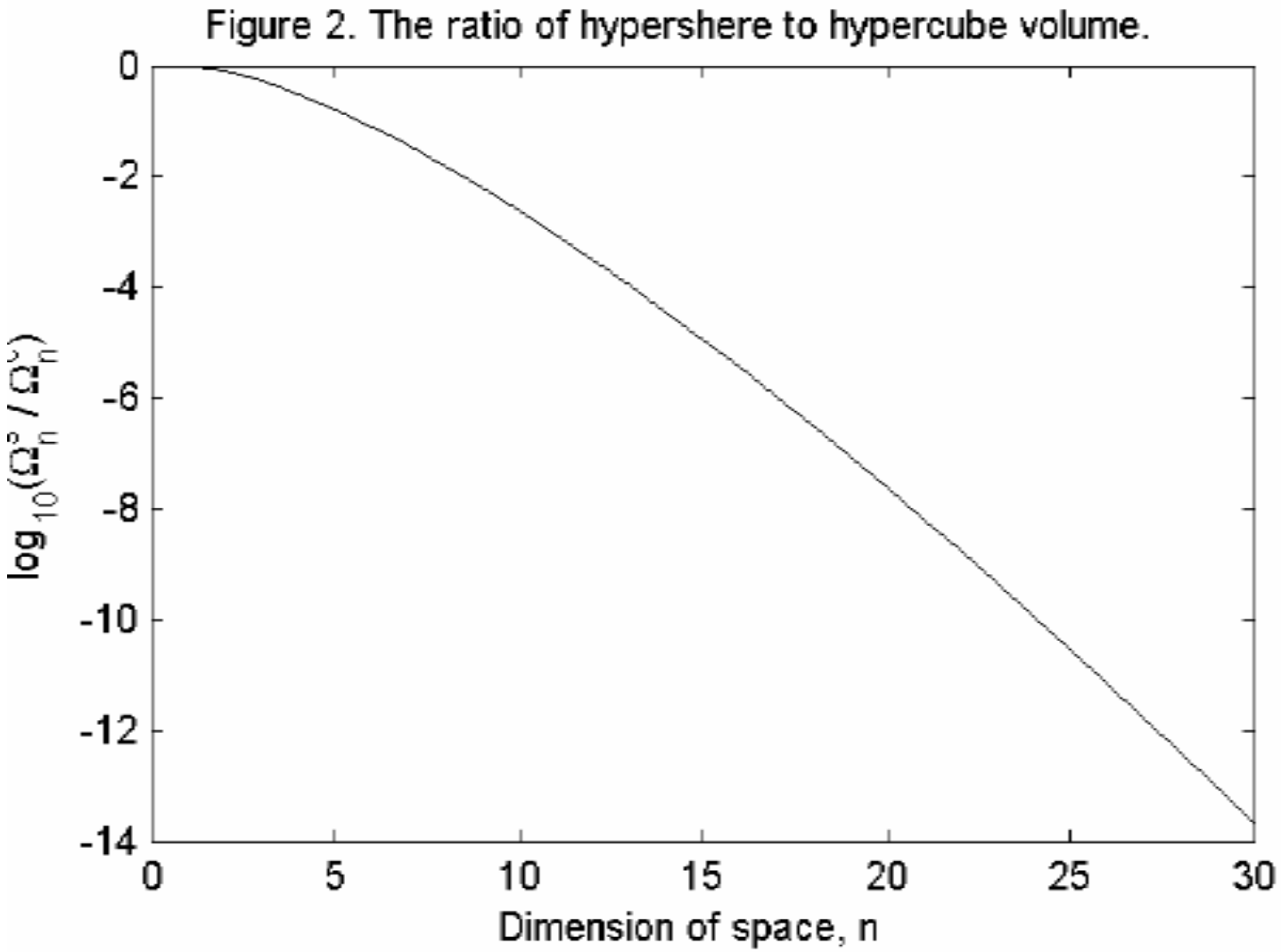
Ratio of hypersphere volume to hypercube volume

- **2-dimensional case:** a circle inscribed within a square occupies about 79% of the area of the square.
- ***p*-dimensional case:** the ratio of a hypersphere's volume Ω_p^s to a hypercube's volume Ω_p^c:

$$\frac{\Omega_p^s}{\Omega_p^c} = \begin{cases} \frac{(\pi/2)^{\frac{p-1}{2}}}{1 \cdot 3 \cdot \ldots \cdot p} \text{ for } p = 1, 3, 5\ldots \\ \\ \frac{(\pi/2)^{\frac{p}{2}}}{2 \cdot 4 \cdot \ldots \cdot p} \text{ for } p = 2, 4, 6\ldots \end{cases}.$$

- **Ratio $\frac{\Omega_p^s}{\Omega_p^c}$ declines rapidly with the dimension of the state space:**
  - when $p = 10$, the ratio $\frac{\Omega_{10}^s}{\Omega_{10}^c} = 3 \cdot 10^{-3}$;
  - when $p = 30$, the ratio $\frac{\Omega_{30}^s}{\Omega_{30}^c} = 2 \cdot 10^{-14}$.

Ergodic set versus tensor-product grid: estimated reduction in cost



Stochastic simulation in problems with high dimensionality

In problems with high dimensionality:

- The hypersphere ergodic set is just a tiny fraction of the hypercube tensor-product grid.
- We avoid computing a solution in an enormously large number of points on the hypercube's edges.
- Stochastic simulation methods are attractive for high-dimensional applications.

But ...

- Previous stochastic simulation methods are numerically unstable. *For example, Marcet's (1988) simulation-based version of parameterized expectation algorithm is unstable even under low (2-nd) degree polynomials; see den Haan and Marcet (1990).*

What causes numerical problems?

1. *Ill-conditioning of the least-squares (LS) problem* solved in the approximation step,

$$\min_{\beta} \left[Y - \Psi\left(k, \theta; \beta\right)\right]' \left[Y - \Psi\left(k, \theta; \beta\right)\right].$$

It arises due to multicollinearity and poor scaling of explanatory variables.

2. In addition, *exponentiated polynomial approximation* $\Psi\left(k, \theta; \beta\right) = \exp\left(\beta_0 + \beta_1 \ln k_t + \beta_2 \ln \theta_t + ...\right)$, used in Marcet (1988), should be estimated with non-linear least-squares (NLLS) methods which

a) require supplying an initial guess;
b) involve computing costly Jacobian and Hessian matrices;
c) can have multiple local minima;
d) often fail to converge.

LS problem

- Consider first a LS problem

$$\min_{\beta} \| Y - \Psi(k, \theta; \beta) \|_2^2 = \min_{\beta} [Y - \Psi(k, \theta; \beta)]' [Y - \Psi(k, \theta; \beta)]$$

  where $\| \cdot \|_2$ denotes L_2 (Euclidean) vector norm.
- Under *the linear regression model*, a solution is

$$\widehat{\beta} = (X'X)^{-1} X'Y$$

- Under *the non-linear regression model*, the LS estimator generally cannot be written explicitly and should be computed with NLLS methods.

Ill-conditioned LS problem in the linear regression model

- The matrix $X'X$ is often *ill-conditioned.* The degree of ill-conditioning of $X'X$ can be measured in terms of a condition number

$$\kappa\left(X'X\right) \equiv \lambda_1 / \lambda_n$$

  $\lambda_1 =$ the largest eigenvalue of $X'X$; $\lambda_n =$ its smallest eigenvalue.
- The eigenvalues of $X'X$ are defined by

$$X'X = V\Lambda V'$$

  $\Lambda =$ an $n \times n$ diagonal matrix with ordered eigenvalues of $X'X$ on its diagonal; $V =$ an $n \times n$ matrix of its eigenvectors.
- $\kappa \uparrow \Longrightarrow \det\left(X'X\right) = \det\left(\Lambda\right) = \lambda_1\lambda_2...\lambda_n \downarrow$ and the closer is $X'X$ to being singular (not invertible).

Multicollinearity problem

Multicollinearity occurs because high-order polynomial terms forming the matrix X are significantly correlated.

Example

Let $Y = \begin{bmatrix} y_1 \\ y_2 \end{bmatrix}$ and $X = [X_1, X_2] = \begin{bmatrix} x_{11} & x_{11} + \phi \\ x_{12} & x_{12} \end{bmatrix}$, with $x_{12} \neq 0$ and $\phi \neq 0$. In this case, the OLS solution $\widehat{\beta} = (X'X)^{-1} X'Y$ is

$$\widehat{\beta}_1 = \frac{y_2}{x_{12}} - \widehat{\beta}_2 \text{ and } \widehat{\beta}_2 = \frac{y_1}{\phi} - \frac{y_2 x_{11}}{\phi x_{12}}.$$

If $\phi \to 0$, we have $\det(X'X) = x_{12}^2 \phi^2 \to 0$, $\kappa(X'X) = \frac{x_{11}+x_{12}}{\phi} \to \infty$, $\widehat{\beta}_1, \widehat{\beta}_2 \to \pm\infty$, and $\widehat{\beta}_1 \approx -\widehat{\beta}_2$, i.e. a large positive coefficient on one variable is canceled by a similarly large negative coefficient on its correlated counterpart.

Poor scaling problem

- The scaling problem arises when polynomial terms in X have significantly different means and variances.
- *Why?* Due to differential scaling among either the state variables (k_t and θ_t) or polynomial terms of different orders (e.g., k_t and k_t^5).

Example

Let $Y = \begin{bmatrix} y_1 \\ y_2 \end{bmatrix}$ and $X = [X_1, X_2] = \begin{bmatrix} x_{11} & \phi \\ x_{12} & 0 \end{bmatrix}$ with $x_{12} \neq 0$ and $\phi \neq 0$.

In this case, the OLS solution $\widehat{\beta} = (X'X)^{-1} X'Y$ is

$$\widehat{\beta}_1 = \frac{y_2}{x_{12}} \text{ and } \widehat{\beta}_2 = \frac{y_1}{\phi} - \frac{y_2 x_{11}}{\phi x_{12}}.$$

If $\phi \to 0$, we have $\det(X'X) = x_{12}^2 \phi^2 \to 0$, $\kappa(X'X) = \frac{x_{11}}{\phi} \to \infty$, and $\widehat{\beta}_2 \to \pm\infty$, i.e. entries of X_2 that are excessively small in absolute value lead to an excessively large absolute value to coefficient $\widehat{\beta}_2$.

Our strategies for enhancing the numerical stability

In the paper,

1. We introduce LS methods that are more numerically stable than the standard OLS method.
2. We replace the ill-conditioned LS problem with some other less ill-conditioned problem.

LS approaches to the linear regression model

Two LS approaches that are more numerically stable and more suitable for dealing with ill-conditioning than the standard OLS approach.

1. *LS using SVD (LS-SVD)*: infers $(X'X)^{-1}$ matrix included in the OLS formula from a singular value decomposition (SVD) of the matrix X.
2. *Regularized LS using Tikhonov regularization (RLS-Tikhonov)*: relies on a specific (Tikhonov) regularization of the ill-conditioned LS problem that imposes penalties based on the size of the regression coefficients.

The LS-SVD approach finds a solution to the original ill-conditioned LS problem, while the RLS-Tikhonov approach modifies (regularizes) the original ill-conditioned LS problem into a well-conditioned problem.

LS-SVD

- SVD of the matrix $X \in \mathbb{R}^{T \times n}$

$$X = USV'$$

where $U \in \mathbb{R}^{T \times n}$ and $V \in \mathbb{R}^{n \times n}$ = orthogonal matrices; $S \in \mathbb{R}^{n \times n}$ = diagonal matrix with diagonal entries $s_1 \geq s_2 \geq ... \geq s_n \geq 0$, known as *singular values* of X.

- The OLS estimator $\widehat{\beta} = (X'X)^{-1} X'Y$ in terms of the SVD:

$$\widehat{\beta} = (VS'SV')^{-1} VS'U'Y = VS^{-1}U'Y$$

- If $X'X$ is well-conditioned $\Longrightarrow$ the OLS formula and the LS-SVD formula give identical estimates of β.
- However, if $X'X$ is ill-conditioned and the standard OLS estimator cannot be computed $\Longrightarrow$ it is still possible that matrices X and S are sufficiently well-conditioned $\Longrightarrow$ can compute the LS-SVD estimator.

LS-SVD

Example

Let $Y = \begin{bmatrix} y_1 \\ y_2 \end{bmatrix}$ and $X = [X_1, X_2] = \begin{bmatrix} \phi & 0 \\ 0 & 1 \end{bmatrix}$ with $\phi \neq 0$. Then, $S = X$ and $U = V = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}$. The condition numbers of S and $X'X$ are related by $\kappa(S) = \sqrt{\kappa(X'X)} = \phi$. The OLS and the LS-SVD estimators coincide,

$$\widehat{\beta} = (X'X)^{-1} X'Y = \begin{bmatrix} 1/\phi^2 & 0 \\ 0 & 1 \end{bmatrix} \begin{bmatrix} \phi y_1 \\ y_2 \end{bmatrix} =$$

$$= VS^{-1}U'Y = \begin{bmatrix} 1/\phi & 0 \\ 0 & 1 \end{bmatrix} \begin{bmatrix} y_1 \\ y_2 \end{bmatrix} = \begin{bmatrix} y_1/\phi \\ y_2 \end{bmatrix}.$$

If $\phi < 1$, the largest elements of $(X'X)^{-1}$ and S^{-1} are $1/\phi^2$ and $1/\phi$, respectively. When $\phi \approx 0$, the computer has a better chance to compute $1/\phi$ than $1/\phi^2$ since $1/\phi \ll 1/\phi^2$.

RLS-Tikhonov

- Regularization - process of re-formulating an ill-conditioned problem by imposing additional restrictions on the solution.
- Tikhonov regularization - the most commonly used regularization method in approximation theory.
- Impose an L_2 penalty on the size of the regression coefficients:

$$\min_{\beta} \|Y - X\beta\|_2^2 + \eta \|\beta\|_2^2 = \min_{\beta} (Y - X\beta)'(Y - X\beta) + \eta\beta'\beta$$

  where $\eta \geq 0 =$ regularization parameter.
- Find the FOC with respect to β

$$\widehat{\beta}(\eta) = (X'X + \eta I_n)^{-1} X'Y$$

  where $I_n =$ an identity matrix of order n.
- *Note:* add a positive constant to $X'X$ prior to inverting this matrix.
  $\Longrightarrow$ Even if $X'X$ is singular, the matrix $X'X + \eta I_n$ is non-singular.
  $\Longrightarrow$ Can compute its inverse.

LAD approaches to the linear regression model

- Replace the ill-conditioned LS problem with a least absolute deviation (LAD) problem

$$\min_{\beta} \|Y - X\beta\|_1 = \min_{\beta} 1_T' |Y - X\beta|$$

  where $\|\cdot\|_1$ denotes L_1 vector norm.
- The LAD problem does not require computing $(X'X)^{-1}$.
- No explicit solution. However, we can re-formulate the LAD problem to consist of a linear objective function and linear constraints $\Longrightarrow$ Solve with standard linear programming techniques.
- Substitute $|Y - X\beta|$ with a vector $w \in \mathbb{R}^T$ to obtain

$$\min_{\beta,\ w} 1_T' w$$

$$\text{s.t.} \quad -w \leq Y - X\beta \leq w$$

- This problem has $n + T$ unknowns. We argue that it is not the most suitable for a numerical analysis.

LAD: primal problem (LAD-PP)

- Charnes et al. (1955): express the deviation for each observation as a difference between two non-negative variables u_t and v_t,

$$y_t - \sum_{i=0}^{n} \beta_i x_{it} = u_t - v_t \tag{1}$$

- u_t and v_t can be interpreted as non-negative vertical deviations above and below the fitted line, $\widehat{y}_t = X_t \widehat{\beta}$, respectively; $u_t + v_t =$ absolute deviation between the fit $\widehat{y}_t$ and the observation y_t.
- *Primal problem:* minimize the total sum of absolute deviations subject to (1),

$$\min_{\beta,u,v} \; 1_T' u + 1_T' v$$

$$\text{s.t. } u - v + X\beta = Y$$

$$u \geq 0, \quad v \geq 0$$

where $u,\ v \in \mathbb{R}^T$.

- This formulation is more simple to solve than the direct formulation.

LAD: dual problem (LAD-DP)

- Every primal problem can be converted into a dual problem.
- *Dual problem* corresponding to the primal problem:

$$\max_{q} \; Y'q$$
$$\text{s.t. } X'q = 0$$
$$-1_T \leq q \leq 1_T$$

  where $q \in \mathbb{R}^T$ is a vector of unknowns.
- If the number of observations, T, is sizable (i.e. $T \gg n$), the dual problem is less computationally cumbersome than the primal problem.

Regularized LAD (RLAD)

- Modify the original LAD problem to incorporate an L_1 penalty on β.
- *The RLAD problem:*

$$\min_{\beta} \|Y - X\beta\|_1 + \eta \|\beta\|_1 = \min_{\beta} 1_T' |Y - X\beta| + \eta 1_n' |\beta|$$

  where $\eta \geq 0 =$ regularization parameter.
- We develop a linear programming formulation of the RLAD problem parallel to the LAD-PP: replace $|\beta_i|$ with two variables.
- Wang, Gordon and Zhu (2006): represent $|\beta_i|$ as $sign(\beta_i)\beta_i$.

RLAD: primal problem (RLAD-PP)

- To cast the RLAD problem into a linear programming form, we represent β as $\beta_i = a_i - b_i$, with $a_i \geq 0$, $b_i \geq 0$ for $i = 1, ..., n$.
- We then impose a linear penalty on each a_i and b_i.
- The resulting regularized version of the primal problem:

$$\begin{aligned} \min_{a,b,u,v} \quad & 1_T' u + 1_T' v + \eta 1_n' a + \eta 1_n' b \\ \text{s.t.} \quad & u - v + Xa - Xb = Y \\ & u \geq 0, \quad v \geq 0 \\ & a \geq 0, \quad b \geq 0 \end{aligned}$$

  where a, $b \in \mathbb{R}^n$ = vectors that define β.
- This problem has $2T + 2n$ unknowns, as well as T equality restrictions and $2T + 2n$ lower bounds.

RLAD: dual problem (RLAD-DP)

- The dual problem corresponding to the RLAD-PP:

$$\begin{aligned} &\max_{q} \; Y'q \\ &\text{s.t. } X'q \leqslant \eta \cdot 1_n \\ &-X'q \leqslant \eta \cdot 1_n \\ &-1_T \leq q \leq 1_T \end{aligned}$$

  where $q \in \mathbb{R}^T$ = vector of unknowns.

- Here, $2n$ linear inequality restrictions and $2T$ lower and upper bounds on T unknown components of q.

LS and LAD approaches to the non-linear regression model

- Extensions to the case of the non-linear regression model,

$$Y = \Psi\left(k, \theta; \beta\right) + \varepsilon$$

- NLLS computes a Taylor's expansion of $\Psi\left(k, \theta; \beta\right)$ around a initial guess, β and makes a step $\Delta\beta$ toward a solution, $\widehat{\beta}$,

$$\widehat{\beta} \simeq \beta + \Delta\beta$$

- The step $\Delta\beta$ is a solution to the system of normal equations,

$$J'J\Delta\beta = J'\Delta Y$$

where $J \equiv \begin{pmatrix} \frac{\partial\Psi(k_1,\theta_1;\beta)}{\partial\beta_0} & \cdots & \frac{\partial\Psi(k_1,\theta_1;\beta)}{\partial\beta_n} \\ \cdots & \cdots & \cdots \\ \frac{\partial\Psi(k_T,\theta_T;\beta)}{\partial\beta_0} & \cdots & \frac{\partial\Psi(k_T,\theta_T;\beta)}{\partial\beta_n} \end{pmatrix}$ is Jacobian and

$$\Delta Y \equiv \begin{pmatrix} y_1 - \Psi\left(k_1, \theta_1; \beta\right) \\ \cdots \\ y_T - \Psi\left(k_T, \theta_T; \beta\right) \end{pmatrix}$$

LS and LAD approaches to the non-linear regression model

Gauss-Newton method,

$$\Delta\beta = \left(J'J\right)^{-1} J'\Delta Y \text{ looks like OLS } \beta = \left(X'X\right)^{-1} X'Y$$

$J'J$ is ill-conditioned $\Longrightarrow$ Employ the described approaches developed for the linear regression model.

1. Compute an inverse of the ill-conditioned matrix $J'J$ by using LS methods based on *SVD or QR factorization of* J.
2. Tikhonov type of regularization leading to the *Levenberg-Marquart* method,
$$\Delta\beta = \left(J'J + \eta I_{n+1}\right)^{-1} J'\Delta Y$$
3. Replace the ill-conditioned NLLS problem with *a non-linear LAD* (NLLAD) problem,
$$\min_{\beta} 1'_T \left|Y - \Psi\left(k, \theta; \beta\right)\right| \simeq \min_{\Delta\beta} 1'_T \left|\Delta Y - J\Delta\beta\right|$$
Formulate NLLAD problem as a linear programming problem: non-regularized & regularized primal and dual problems.

Unified principal component method

- We merge our approximation methods & the principal component analysis into a unified approach.
  $\Longrightarrow$ Can handle any degree of ill-conditioning.
- $Z \equiv XV$, where $X \in \mathbb{R}^{T \times n}$, $Z \in \mathbb{R}^{T \times n}$ and $V \in \mathbb{R}^{n \times n}$.
- $Z_1, ..., Z_n$ are called *principal components* of X and are orthogonal, $Z_i' Z_i = s_i^2$ and $Z_j' Z_i = 0$ for any $j \neq i$, where $s_i = i$-th singular value of X.
- *Idea:* reduce ill-conditioning of X to a "desired" level by excluding low variance principle components corresponding to small singular values.
- Let $\overline{\kappa}$ = largest condition number of X that we are willing to accept.
- Compute $\frac{s_1}{s_2}, ..., \frac{s_1}{s_n}$, where s_1 = largest singular value.
- $\kappa(X) = \kappa(S) = \frac{s_1}{s_n}$ = actual condition number of the matrix X.

Unified principal component method

- Let $Z^r \equiv (Z_1, ..., Z_r) \in \mathbb{R}^{T \times r}$ be the first r principal components for which $\frac{s_1}{s_i} \leq \overline{\kappa}$.
- Remove the last $n - r$ principal components for which $\frac{s_1}{s_i} > \overline{\kappa}$.
- By construction, $\kappa(Z^r) \leq \overline{\kappa}$.
- Re-write the linear regression model in terms of Z^r,

$$Y = Z^r \vartheta + \varepsilon$$

  where $\vartheta \in \mathbb{R}^r$ = vector of coefficients.
- Estimate ϑ using any of the LS and LAD methods described.
- Find $\widehat{\beta} = V^r \widehat{\vartheta} \in \mathbb{R}^n$, where $V^r = (V_1, ..., V_r) \in \mathbb{R}^{n \times r}$ contains the first r right singular vectors of X.

Other factors affecting the numerical stability

1. The choice of policy functions to parameterize (e.g., capital versus marginal-utility policy functions).
2. The choice of a polynomial family (e.g., ordinary versus Hermite polynomials).
3. Normalization of variables in the regression.

Choosing a policy function to parameterize

- *Our benchmark case:* parameterize capital decision function $k_{t+1} = K(k_t, \theta_t)$,

$$k_{t+1} = E_t \left\{ \delta \frac{u'(c_{t+1})}{u'(c_t)} \left[1 - d + \theta_{t+1} f'(k_{t+1}) \right] k_{t+1} \right\} \simeq \Psi(k_t, \theta_t; \beta)$$

  where $\Psi(k_t, \theta_t; \beta)$ = flexible functional form that depends on a vector of coefficients β.

- *Many other parameterizations exist: e.g.,* parameterize marginal utility as in den Haan and Marcet (1990)

$$u'(c_t) = \delta E_t \left\{ u'(c_{t+1}) \left[1 - d + \theta_{t+1} f'(k_{t+1}) \right] \right\} \simeq \delta \Psi(k_t, \theta_t; \beta)$$

Choosing a polynomial representation

- Polynomial space of functions.
- Ordinary polynomial representation - standard.
- We consider Hermite polynomial representation.
- Ordinary polynomials $P_m(x)$ versus Hermite polynomials $H_m(x)$ up to degree five:

$$
\begin{aligned}
P_0(x) &= 1 & H_0(x) &= 1 \\
P_1(x) &= x & H_1(x) &= x \\
P_2(x) &= x^2 & H_2(x) &= x^2 - 1 \\
P_3(x) &= x^3 & H_3(x) &= x^3 - 3x \\
P_4(x) &= x^4 & H_4(x) &= x^4 - 6x^2 + 3 \\
P_5(x) &= x^5 & H_5(x) &= x^5 - 10x^3 + 15x.
\end{aligned}
$$

- $P_m(x)$, $m = 1, ..., 5$ appear very similar $\Longrightarrow$ the explanatory variables for the regression are likely to be correlated.
- $H_m(x)$, $m = 1, ..., 5$ are different in the shapes $\Longrightarrow$ the multicollinearity problem manifests to a much lesser degree, if at all.

Choosing a polynomial representation

Figure 3. Ordinary polynomials.

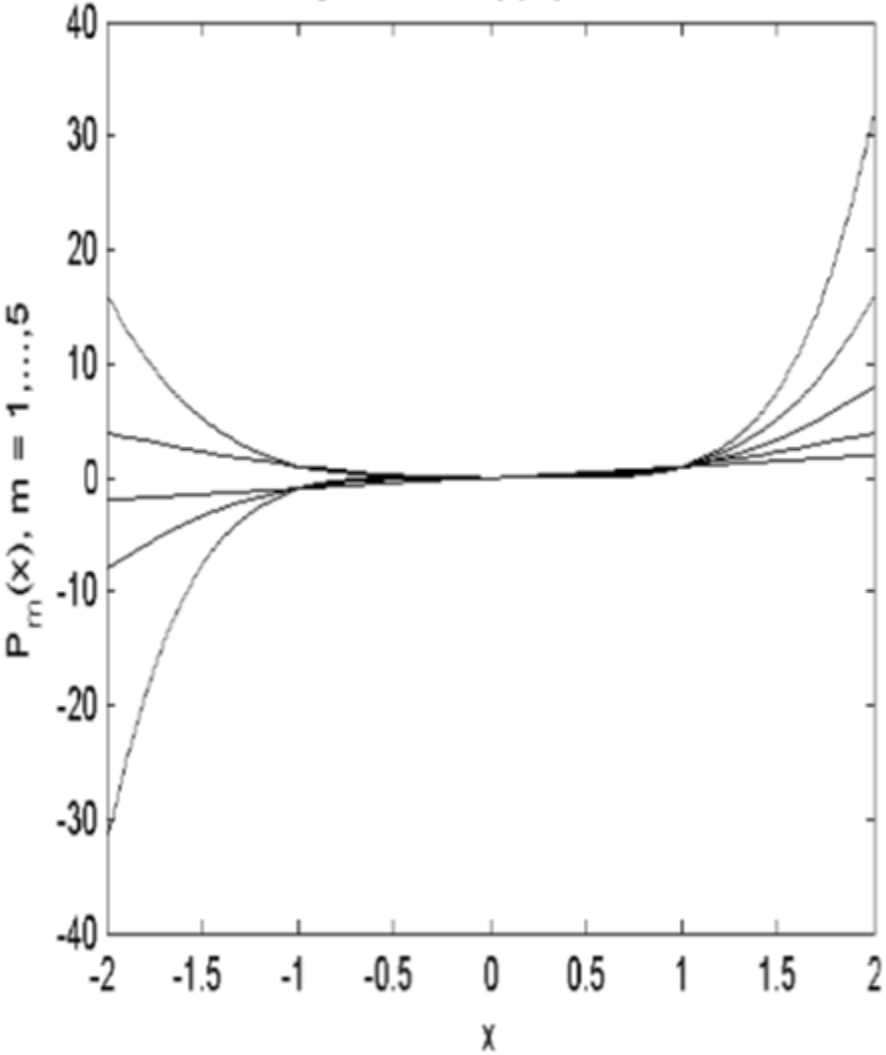
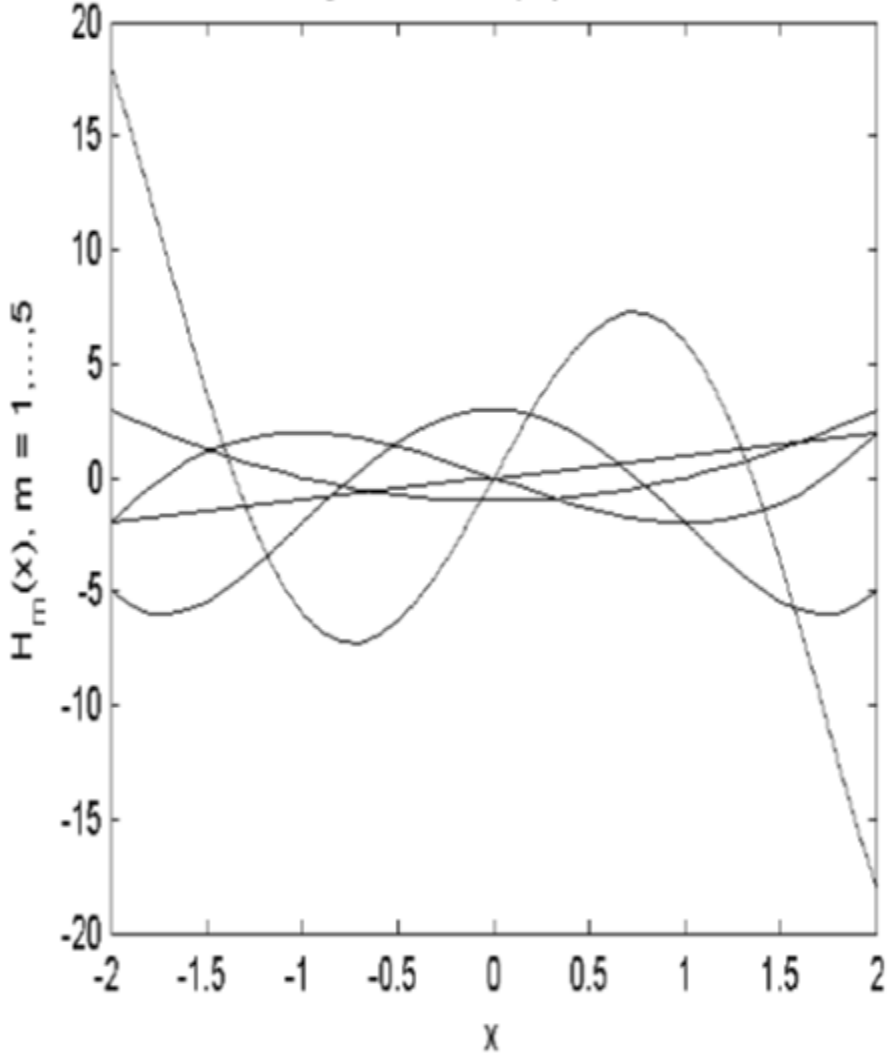


Figure 4. Hermite polynomials.



Choosing a polynomial representation

- We approximate the function Ψ by *a complete set of polynomials* in k_t and θ_t.
- Example, the complete set of Hermite polynomials of degree three in levels, k_t and θ_t:

$$\Psi(k_t, \theta_t; \beta) = \beta_0 + \beta_1 k_t + \beta_2 \theta_t + \beta_3 \left(k_t^2 - 1\right) + \beta_4 k_t \theta_t + \beta_5 \left(\theta_t^2 - 1\right)$$
$$+ \beta_6 \left(3k_t^3 - 3k_t\right) + \beta_7 \left(k_t^2 - 1\right) \theta_t + \beta_8 k_t \left(\theta_t^2 - 1\right) + \beta_9 \left(3\theta_t^3 - 3\theta_t\right)$$

- The polynomial approximations of orders one, two, three, four and five have 3, 6, 10, 15 and 21 coefficients, respectively.

Normalizing the variables

- *Center* - subtract the sample mean from each observation.
- *Scale* - divide each observation by the sample standard deviation.
- By construction, a centered variable has a zero mean, and a scaled variable has a unit standard deviation.
- After a regression model is estimated, the coefficients in the original (unnormalized) regression model are restored.

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: $\delta = 0.99$.
- Depreciation rate: $d = 1$ and $d = 0.02$.
- Under $\gamma = 1$ and $d = 1 \Longrightarrow$ *closed-form solution*.
- Sequence of shocks of length $T = 10,000$, and $T = 3,000$.
- Unit-free Euler equation errors

$$e(k_t, \theta_t) \equiv E_t \left[\frac{c_{t+1}^{-\gamma}}{c_t^{-\gamma}} \left(1 - d + \alpha \theta_{t+1} k_{t+1}^{\alpha-1}\right) \right] - 1$$

One-country model with closed-form solution

Polynomial	Non-regularization methods									Regularization methods		
	Unnormalized data			Normalized data			Hermite polynomials					
	e_{mean}	e_{max}	*CPU*	e_{mean}	e_{max}	*CPU*	e_{mean}	e_{max}	*CPU*	e_{mean}	e_{max}	*CPU*
	OLS									RLS-Tikhonov, $\overline{\eta}$ = 1(-7)		
1st degree	3.29(-4)	3.35(-3)	2(-1)	3.29(-4)	3.35(-3)	1(-1)	3.29(-4)	3.35(-3)	2(-1)	3.29(-4)	3.35(-3)	2(-1)
2nd degree	3.92(-6)	8.38(-5)	4	3.92(-6)	8.38(-5)	5	3.92(-6)	8.38(-5)	6	3.92(-6)	8.37(-5)	5
3rd degree	-	-	-	1.76(-7)	5.71(-6)	4	1.71(-7)	5.94(-6)	3	1.69(-6)	3.94(-5)	2
4th degree	-	-	-	-	-	-	1.22(-8)	7.83(-7)	1	9.23(-7)	2.72(-5)	2
5th degree	-	-	-	-	-	-	1.04(-9)	8.34(-8)	2	7.83(-7)	2.84(-5)	4
	LS-SVD									RLS-TSVD, $\overline{\kappa}$ = 1(6)		
1st degree	3.29(-4)	3.35(-3)	2(-1)	3.29(-4)	3.35(-3)	2(-1)	3.29(-4)	3.35(-3)	2(-1)	3.29(-4)	3.35(-3)	2(-1)
2nd degree	3.92(-6)	8.38(-5)	5	3.92(-6)	8.38(-5)	5	3.92(-6)	8.38(-5)	6	3.92(-6)	8.38(-5)	8
3rd degree	1.71(-7)	5.94(-6)	3	1.71(-7)	5.94(-6)	3	1.71(-7)	5.94(-6)	3	1.71(-7)	5.94(-6)	3
4th degree	1.23(-8)	8.14(-7)	2	1.22(-8)	7.82(-7)	1	1.22(-8)	7.82(-7)	2	7.69(-8)	2.89(-6)	2
5th degree	-	-	-	1.59(-9)	1.07(-7)	2(-1)	1.06(-9)	9.57(-8)	2	3.01(-8)	2.01(-6)	4
	LAD-DP									RLAD-DP, $\overline{\eta}$ = 1(-4)		
1st degree	2.95(-4)	3.58(-3)	1(1)	3.29(-4)	3.36(-3)	1(1)	2.95(-4)	3.58(-3)	1(1)	3.29(-4)	3.36(-3)	1(1)
2nd degree	3.40(-6)	9.26(-5)	1(1)	3.39(-6)	9.31(-5)	1(1)	3.40(-6)	9.26(-5)	1(1)	3.39(-6)	9.24(-5)	2(1)
3rd degree	1.40(-7)	7.73(-6)	8	1.42(-7)	7.67(-6)	7	1.41(-7)	7.77(-6)	8	1.65(-7)	7.23(-6)	4(1)
4th degree	-	-	-	1.23(-8)	1.20(-6)	4	9.92(-9)	9.55(-7)	4	2.37(-7)	1.07(-5)	6
5th degree	-	-	-	-	-	-	7.45(-10)	1.13(-7)	3	3.12(-7)	1.52(-5)	3(1)

The model with a closed-form solution

- *OLS & unnormalized ordinary polynomials*: cannot go beyond a 2-degree polynomial approximation, and the Euler equation errors are of order 10^{-6}.
- OLS & *normalized variables:* the 3-degree approximation and accuracy of order 10^{-7}.
- *Dual formulation* of the LAD problem & *normalized variables*: the 4-degree approximation and accuracy of order 10^{-8}.
- *Primal formulation*: costly in terms of memory and time.
- *Hermite polynomials:* approximation of all five orders and accuracy of order 10^{-9}.
- The same results for *LS using SVD,* Tikhonov regularization, principal component method, and primal and dual LAD regularization methods.

The general one-country model

- Do not achieve such a remarkable accuracy: sample size restricts accuracy.
- However, our stochastic simulation algorithm is still stable and can compute polynomial approximations up to the 5-degree.
- The model with a highly risk averse agent: fixed-point iteration is fragile under parameterization of capital decision function, however, stability is restored if the marginal utility function instead of capital function is parameterized.

Multi-country model

J countries ($2J$ state variables). Planner solves the following maximization problem:

$$\max_{\left\{\left\{c_t^j, k_{t+1}^j\right\}_{j=1}^J\right\}_{t=0}^{\infty}} E_0 \sum_{j=1}^{J} v^j \left[\sum_{t=0}^{\infty} \delta^t \ln\left(c_t^j\right)\right]$$

s.t. aggregate budget constraint,

$$\sum_{j=1}^{J} c_t^j + \sum_{j=1}^{J} k_{t+1}^j = \sum_{j=1}^{J} k_t^j (1-d) + \sum_{j=1}^{J} \theta_t^j A f^j\left(k_t^j\right),$$

and the process for technology level,

$$\ln \theta_t^j = \rho \ln \theta_{t-1}^j + \left(\epsilon_t + \xi_t^j\right), \quad \epsilon_t \sim \mathcal{N}\left(0, \sigma^2\right), \text{ and } \xi_t^j \sim \mathcal{N}\left(0, \sigma^2\right),$$

where initial condition $\left\{k_0^j, \theta_0^j\right\}_{j=1}^J$ is given.

Model with up to 200 countries (400 state variables)

- Running time ranges from 3 min (2 countries) to 3 hours (200 countries). Linearly additive polynomial is crucial for speed!
- The largest error of our least accurate solution is less than 0.2% (for 200 countries).

# of countries	# of coefficients	T = 10,000		
		e_{mean}	e_{max}	CPU
J=2	5x2	7.9(-5)	6.3(-4)	2(2)
J=4	9x4	8.4(-5)	6.4(-4)	2(2)
J=6	13x6	8.3(-5)	5.7(-4)	3(2)
J=8	17x8	9.1(-5)	6.4(-4)	3(2)
J=10	21x10	1.1(-4)	7.1(-4)	5(2)
J=12	25x12	1.1(-4)	6.8(-4)	4(2)
J=16	33x16	1.3(-4)	8.3(-4)	5(2)
J=20	41x20	1.5(-4)	7.9(-4)	6(2)
J=30	61x30	1.8(-4)	8.7(-4)	9(2)
J=40	81x40	2.2(-4)	1.0(-3)	2(3)
J=60	121x60	2.5(-4)	1.2(-3)	3(3)
J=100	201x100	2.9(-4)	1.3(-3)	6(3)
J=150	301x150	3.4(-4)	1.6(-3)	1(4)
J=200	401x200	3.9(-4)	1.9(-3)	1(4)

Conclusion

- Standard LS methods (OLS and Gauss-Newton methods) fail when problems are ill-conditioned.
- We use more powerful and stable approximation methods.
- *Lessons:*
  1. Normalize the variables, as it never hurts.
  2. Look for basis polynomial functions that do not automatically give multicollinearity.
  3. Use approximation methods that can handle ill-conditioned problems.
  4. Apply the unified principal component method if degrees of ill-conditioning are very high.
  5. Explore alternative decision functions to parameterize.
- We solve problems of much higher dimensionality than the literature and achieve high accuracy.