High-dimensional Problems in Finance and Economics

Thomas M. Mertens

NYU Stern

Risk Economics Lab

April 17, 2012
Motivation

• Many problems in finance and economics are high dimensional.
  • Dynamic Optimization: Multiple kinds of capital stocks
  • DSGE: Multiple consumers/firms/countries
  • Games: Multiple players and states
  • Bayesian analyses compute high-dimensional integrals
  • Bootstrapping: analyze many n-dimensional samples from n data points
  • Simulation of large Markov processes - MCMC, Gibbs sampling, ACE
  • Parameter space searches to find robust conclusions

• Problems for both integration and approximation arise.
Many problems thus display the curse of dimensionality.

The problem arises when you approximate a function on a grid.

Use $n$ points in each dimension.

For a $d$-dimensional problem, you have $n^d$ grid points.

The dimensionality for a good approximation is thus limited.
What is the reaction?

• According to a giant in this field:
  • Response I: Analyze silly models
    • Reduce heterogeneity in tastes, abilities, age, etc.
    • Assume no risk
    • Assume common information, beliefs, and learning rules
  • Response II: Do bad math
  • Response III: Do bad math when analyzing silly models
Bad news

• There is little appreciation for methods.
• A purely technical contribution usually does make its way into top journals.
• However you get a lot of attention if
  • you solve more general models and find new effects
  • if you create new models that couldn’t be solved before.
• There is potentially a big benefit from cross-disciplinary research.
I will talk about three ideas how economists deal with the curse of dimensionality.

Two of the methods are used in my own work.

The first method is grid-based and demonstrates the curse of dimensionality.

Then I will introduce a model with many state variables (and thus dimensions)

and two methods to solve it.
• The first problem deals with estimation of economic models.
• Bayesian estimation involves integration of the posterior.
• We present a numerical quadrature approach for Bayesian estimation.
• We use sparse grids to deal with high dimensionality.
• It is an alternative to the use of simulations.
Bayesian Estimation

- Get data $Y$.
- Obtain likelihood function $L(Y|\theta)$.
- Use prior information $p(\theta)$.
- Posterior: $p(\theta|Y) \sim L(Y|\theta) \cdot p(\theta)$
- We then compute Bayesian estimators:

$$M = \int h(\theta) L(Y|\theta) p(\theta) d\theta$$
Technique: MCMC

- MCMC (Markov-chain Monte Carlo) methods help to compute the integral.
- Create Markov-chain that has true distribution as stationary distribution.
- Pick starting point.
- Pick draws from each marginal distribution to get to the next step.
Technique: MCMC

- Metropolis-Hastings.
- Pick starting point.
- Draw sample $\psi$ from proposal distribution $Q(\theta_t'|\theta_{t-1})$ (i.e. $Q(\theta_t'|\theta_{t-1}) \sim N(\theta_{t-1}, \Sigma)$).
- Accept draw only if $u < \frac{p(\psi) \mathcal{L}(Y|\psi)}{p(\theta_{t-1}) \mathcal{L}(Y|\theta_{t-1})}$ where $u \sim U([0, 1])$ otherwise $\theta_t = \theta_{t-1}$.
- Toss first $N$ draws which is known as the ”burn-in phase”.
• Main difficulty is dimensionality of the problem.
• Monte-Carlo methods converge at a rate independent of dimension — but slowly.
• Slow convergence puts restrictions on the set of models that can be estimated.
• We want a faster method.
• There is hope: the function is very smooth.
• Pseudo-random schemes give you convergence $O(N^{-\frac{1}{2}})$.
• Equidistributional sequences give convergence of order 1 for $C^1$-functions.
• There are non-simulation based methods with convergence of $O(N^{-k})$ for periodic $C^k$-functions.
• For very smooth functions, there is essentially no ”curse of dimensionality”.
Goal: Approximate function $f(x)$ with basis functions.

We write the approximation $\hat{f}(x)$ as:

$$\hat{f}(x) = \sum_{(l,i)} u_{(l,i)} \phi_{(l,i)}(x) \approx f(x)$$

Basis functions could for instance be:
Approximation
Approximation
Approximation
Exponents

- Why can we truncate the sum at level $l$?
- Look at the approximation again.
- The absolute value of coefficients shrinks to zero.
- More precisely: $u(l,i) \leq c \cdot 2^{-2 \cdot \|l\|_1}$. 
Sparse grids — Construction

- The goal is to generalize the one-dimensional approximation.
- This requires to specify:
  - the basis function
  - the grid
Sparse grids — Construction

- Choice of basis functions and associated grid.

\[ \phi(l,i) = \prod_j \phi(l_j, i_j) \]
Sparse grids — Construction

- "Curse of Dimensionality" for full grids — grid size grows at $n^d$. 
Sparse grids — Construction

Sparse grids:

\[ n=1 \quad n=2 \quad n=3 \]
Approximation

• Note: We can now still write the approximation as

\[ \hat{f}(x) = \sum_{(l,i)} u(l,i) \phi(l,i)(x) \approx f(x) \]

• where \( l = (l_1, \ldots, l_d) \), \( i = (i_1, \ldots, i_d) \)

• \( \phi(l,i) = \prod_j \phi(l_j,i_j) \)

• It is a generalization of the 1-d concept.

• But: sparse grids come at a cost: you have to have bounded second mixed derivatives.
## Features

<table>
<thead>
<tr>
<th></th>
<th>Full grids</th>
<th>Sparse grids</th>
</tr>
</thead>
<tbody>
<tr>
<td># grid points</td>
<td>$n^d$</td>
<td>$O(n \log(n)^{d-1})$</td>
</tr>
<tr>
<td>Accuracy</td>
<td></td>
<td></td>
</tr>
<tr>
<td>$L_2$-error</td>
<td>$O(N^{-2})$</td>
<td>$O(N^{-2} \cdot \log(N)^{d-1})$</td>
</tr>
<tr>
<td>$L_\infty$-error</td>
<td>$O(N^{-2})$</td>
<td>$O(N^{-2} \cdot \log(N)^{d-1})$</td>
</tr>
</tbody>
</table>
Integration and moments

- Having the approximation, we can now integrate
- Just sum up!

\[ \int \sum_{(l,i)} u(l,i) \phi(l,i)(x) = \sum_{(l,i)} u(l,i) \int \phi(l,i)(x) \]

- How about computing moments?
- Use the same grid: since \( \|f - \hat{f}\| < \delta \)

\[ \| \int x^2 (f(x) - \hat{f}(x)) \, dx \| < \delta \int x^2 \, dx \]
Bayesian Estimation

- Let’s look at the problem again.
- Compute expected value of functions of $\theta$

$$M = \int h(\theta) \mathcal{L}(Y|\theta)p(\theta)d\theta$$

- We can integrate directly.
Numerical results

Grid points – log scale

$L^\infty$ error – log scale

Mean
Integral
MCMC

Grid points – log scale

Mean
Integral
MCMC
• This solves three problems:
• 1. Method is fast.
• 2. Error estimation is accurate.
• 3. Get posterior to check for identification.
Motivation

- A large body of literature makes the simplifying assumption of a representative agent.
- In reality, people earn different incomes, have different talents, and hold different expectations.
- For this heterogeneity to impact equilibrium outcomes, asset markets need to be incomplete.
- In reality, we see this type of limited insurance.
- Uninsurable idiosyncratic risk produces heterogeneity across agents.
Difficulties

- With complete markets, agents can insure against idiosyncratic risk.
- As a result, we obtain aggregation.
- With incomplete markets, agents’ choices are affected by their idiosyncratic shocks.
- Hence individual conditions differ across agents.
- The state space consists of one or more distributions.
- We have to deal with the curse of dimensionality.
Standard method

- Dynamic economies are typically solve by approximating the state space.
- Replace the law of motion by a linear function in aggregate state variables.
- You thus impose that prices are only a function of aggregate variables
- and do not depend on the distribution.
- This is an approximation which can produce poor solutions.
Modeling Strategy

Example: Standard Dynamic Stochastic General Equilibrium model

- **Households** Consumption-saving, capital-bond, subject to 
  *uninsurable idiosyncratic labor income shocks*.
- **Representative Firm** Standard production.
Households (1/2)

- $I$ households live in time periods $t = 0, \ldots, \infty$.
- Households choose consumption path and capital holdings to maximize

$$
\max_{c^i_t, k^i_{t+1}, b^i_{t+1}} \sum_{t=0}^{\infty} \beta^t E_0 \left[ u_c(c^i_t) - u_k(k^i_{t+1}, b^i_{t+1}) \right] \quad i = 1, \ldots, I
$$

subject to

$$
c^i_t + k^i_{t+1} + b^i_{t+1} = (1 + r^k_t)k^i_t + (1 + r^b_t)b^i_t + w_t e^{\psi^i_t}.
$$

- Labor income is subject to the individual shock $\psi^i_t$.

$$
\psi^i_{t+1} = \rho^i_\psi \psi_t + \theta^i_{t+1}
$$
Households (2/2)

\[ u_k(k_{t+1}^i, b_{t+1}^i) = \nu_1 \frac{1}{(k_{t+1}^i + b_{t+1}^i - k)^2} \]

\[ u_k(k_{t+1}^i, b_{t+1}^i) = \nu_1 \frac{1}{(k_{t+1}^i + b_{t+1}^i - k)^2} + \nu_2 (k_{t+1}^i - \bar{k})^2 \]

\[ u_k(k_{t+1}^i, b_{t+1}^i) = \nu_1 \frac{1}{(k_{t+1}^i + b_{t+1}^i - k)^2} + \nu_2 (k_{t+1}^i - \bar{k})^2 + \nu_3 (b_{t+1}^i)^2 \]

\[ u_k(k_{t+1}^i, b_{t+1}^i) = \nu_1 \frac{1}{(k_{t+1}^i + b_{t+1}^i - k)^2} + \nu_2 (k_{t+1}^i - \bar{k})^2 + \nu_3 (b_{t+1}^i)^2 + \nu_4 (k_{t+1}^i + b_{t+1}^i) \]

- The utility function \( u_k(k_{t+1}, b_{t+1}) \) has several components:
  - Endogenous borrowing constraint.
  - Defined steady-state distribution of capital holdings.
  - Defined steady-state portfolio weights.
  - Last term ensures that penalty function has global minimum at \( \bar{k} \).
Firm

- Standard production sector.
- Perfect Competition, zero profits.
- Produce the final good according to
  \( Y = f(K, L, z) = e^z K^\alpha L^{1-\alpha} \) where
  \[
  z_{t+1} = \rho z_t + \eta_{t+1}
  \]
- Maximizing profits
  \[
  \max_{K_t, L_t} Y_t - r_t^K K_t - w_t L_t
  \]
- returns first-order conditions:
  \[
  r_t^K = \alpha e^{zt} K_t^{\alpha-1} L_t^{1-\alpha}
  \]
  \[
  w_t = (1 - \alpha) e^{zt} K_t^\alpha L_t^{-\alpha}
  \]
State space

- The state space of this economy is extremely large. For each agent, we need to keep track of:
  - capital holdings
  - bond holdings
  - individual labor income shock
- Furthermore, we need to keep track of aggregate productivity.
- A more complicated model might have many more state variables for each agent.
- The challenge is to find the solution for economic behavior as a function of all state variables.
- In turn, next period’s state variables are determined by the behavior.
- We need to solve a fixed-point problem on the state space.
Projection Methods

- One could build an approximation if the number of agents is relatively small.
- Therefore, one could use sparse grid methods.
- Still, there is a lot of space wasted.
- To see this, note that shocks are typically normally distributed in economics.
- A good area for approximation is not a cube but a sphere!
• The ergodic distribution can be at a different location in the state space than the deterministic version.
• To get a good approximation, one would like to approximate well around the ergodic mean.
• Stochastic simulation methods provide a way to solve the economy where it matters the most.
• We are looking for a consumption function for every single agent in the economy.
• To compute the equilibrium, we build an approximation to the choice functions.
• Therefore, pick a class of approximating functions (e.g. polynomials of degree $n$).
• We seek to find the best approximation in this class of functions.
• Once we have a guess, we can simulate the economy and check for accuracy.
Method

- Choose a flexible class of functional forms $\psi(s; b)$.
- $b$ parameterizes the approximating function.
- To set up, re-write the optimality conditions for all agents in the form

$$k_{t+1}(s_t) = E[\beta \frac{u'(c_{t+1})}{u'(c_t)} (1 + e^{z_{t+1} f'(k_{t+1}, l_{t+1}'))}]$$
Algorithm

- Pick an initial set of parameters $b_0$.
- Choose an initial point in the state space $s_0$.
- In iteration $i$:
- Simulate the economy for $T$ periods using the policy rule $\psi(s, b_i)$
Algorithm

- For each $t$, define the optimal choice under the rule $k_{t+1} = \psi(s; b)$:

$$y_t = E[\beta \frac{u'(c_{t+1})}{u'(c_t)}(1 + e^{z_{t+1}} f'(k_{t+1}, l_{t+1}))]$$

- Find the best approximation $b^*$ by solving

$$\min_b \|\varepsilon_{T-1}^{T-1}\|$$ (1)

- for some norm $\| \cdot \|$.

- Updating:

$$b^{i+1} = (1 - \xi)b^i + \xi b^*$$

- Iterate until convergence.
Discussion

- Good method for medium-scale economies.
- Easy to program
- Typically each iteration can be solved quickly
- but it can potentially take many iterations.
Back to first principles

- Usually little is known about the solution (or even existence) for incomplete market economies.
- We focus on a broad class of models with aggregate and uninsurable idiosyncratic risk.

**What do we know?**

1. Solution at the deterministic steady-state.
2. Solution to the deterministic case exists.
3. Equilibrium conditions are typically "well-behaved".

**How can we use this information to construct a solution to the stochastic problem?**
Graphical illustration (1/3)

- Desired solution.
Graphical illustration (2/3)

- We embed this solution in a larger space.
• We start the solution at the no-risk case.
Using perturbation methods turns out to be particularly convenient.

The key observations are:

1. At the deterministic steady-state, all agents are identical.
2. Expansions around this steady-state are symmetrical:
3. Only need to expand optimality conditions for one agent.
4. Only need to expand in state variables for two agents (first-order).
• Equilibrium conditions in the form

\[ E_t \left[ g^1(X_t, z_t, C_t, P_t, X_{t+1}, z_{t+1}, C_{t+1}, P_{t+1}) \right] = 0 \]

\[ X_{t+1} = g^2(X_t, z_t, C_t, P_t) \]

• \( F \) contains:
  • first-order conditions of all agents (Euler equations)
  • Law of motions
  • Market clearing conditions

• First-order conditions are identical for all agents.
• Thus: expand only one optimality condition.
In our example:

\[
\begin{pmatrix}
  x_1^1 & x_1^2 & \cdots & x_1^N \\
  \vdots & \vdots & \ddots & \vdots \\
  x_i^1 & x_i^2 & \cdots & x_i^N
\end{pmatrix}
\times
\begin{pmatrix}
  z_1 \\
  \vdots \\
  z_Z
\end{pmatrix}
= \left( c_1^i \right)
\begin{pmatrix}
  x_1^1 & x_1^2 & \cdots & x_1^N \\
  \vdots & \vdots & \ddots & \vdots \\
  x_i^1 & x_i^2 & \cdots & x_i^N
\end{pmatrix}
\times
\begin{pmatrix}
  z_1 \\
  \vdots \\
  z_Z
\end{pmatrix},
\sigma
\]

\[
\begin{pmatrix}
  c_1^i \\
  \vdots \\
  c_C^i
\end{pmatrix}
\begin{pmatrix}
  x_1^1 & x_1^2 & \cdots & x_1^N \\
  \vdots & \vdots & \ddots & \vdots \\
  x_i^1 & x_i^2 & \cdots & x_i^N
\end{pmatrix}
\times
\begin{pmatrix}
  z_1 \\
  \vdots \\
  z_Z
\end{pmatrix},
\sigma
\]

\[
\begin{pmatrix}
  k_1^1 & b_1^2 & \psi_1 \\
  \vdots & \vdots & \vdots \\
  k_i^1 & b_i^2 & \psi_i
\end{pmatrix}
\]
Deterministic steady-state

- Deterministic steady-state is defined.
- It features a degenerate distribution of capital.
- Fixing this distribution is important to avoid unit roots.
- Once we move to a stochastic economy, we no longer need to fix the portfolio exogenously.
Deterministic steady-state

- We start the solution at the no-risk case.
Deterministic steady-state

- Deterministic steady-state.
We want to solve the equation

\[ F(y(x), x) = 0 \]

\[ \frac{1}{y(x)} - x + \frac{x^2 + x - 1}{y(x)} = 0 \]

to which the solution is \( y(x) = x + 1 \).

Expand this equation around \( x_0 = 1 \):

\[ \frac{1}{y_0} - x_0 + \frac{x_0^2 + x_0 - 1}{y_0} = 0 \]

Expand this equation around \( x_0 = 1 \):

\[ \frac{1}{y_0} - 1 + \frac{1}{y_0} = 0 \]

Expand this equation around \( x_0 = 1 \):

\[ y_0 = 2 \]

First-order term at \( x_0 = 1 \):

\[ \frac{dF(y(x), x)}{dx} \bigg|_{x=x_0} = 0 \]
Perturbation methods

- Using perturbation methods, we build a *higher-order approximation* around the deterministic steady-state.
- Perturbation methods build a Taylor series approximation of the policy functions and prices around the point of expansion.

\[
\begin{align*}
C^{i} = & \sum_{o=1}^{\infty} \sum_{\|l\|+|j|+|k|=o} \frac{1}{l! \cdot j! \cdot k!} \frac{\partial^{o} C^{i}}{\partial X^{l} \partial z^{j} \partial \sigma^{k}} \bigg|_{(X^{0}, z^{0}, 0)} \langle \langle (X - X^{0})^{l} \rangle \rangle \langle \langle z - z^{0} \rangle \rangle \cdot \sigma^{k} \\
C^{i} = & \sum_{o=1}^{\infty} \sum_{\|l\|+|j|+|k|=o} \frac{1}{l! \cdot j! \cdot k!} \frac{\partial^{o} C^{i}}{\partial X^{l} \partial z^{j} \partial \sigma^{k}} \bigg|_{(X^{0}, z^{0}, 0)} \langle \langle (X - X^{0})^{l} \rangle \rangle \langle \langle z - z^{0} \rangle \rangle \cdot \sigma^{k}
\end{align*}
\]
Higher-order approximation (1/4)

- Obtain derivatives as a solution to
  \[ g^1_i(X_t, z_t, C_t, P_t, X_{t+1}, z_{t+1}, C_{t+1}, P_{t+1}) = 0 \]

  \[ \frac{dg^1_i(X_t, z_t, C_t, P_t, X_{t+1}, z_{t+1}, C_{t+1}, P_{t+1})}{dx^1_1} \bigg|_{(X^0, z^0, 0)} = 0 \]

  \[
  \frac{\partial g^1_i}{\partial x^1_1} + \frac{\partial g^1_i}{\partial C_t} \frac{\partial C_t}{\partial x^1_1} + \frac{\partial g^1_i}{\partial P_t} \frac{\partial P_t}{\partial x^1_1} + \frac{\partial g^1_i}{\partial X_{t+1}} \frac{\partial X_{t+1}}{\partial x^1_1} + \frac{\partial g^1_i}{\partial z_{t+1}} \frac{\partial z_{t+1}}{\partial x^1_1} \\
  + \frac{\partial g^1_i}{\partial C_{t+1}} \frac{\partial C_{t+1}}{\partial x^1_1} + \frac{\partial g^1_i}{\partial P_{t+1}} \frac{\partial P_{t+1}}{\partial x^1_1} = 0
  \]

- **Key steps**
  - Only expand one optimality condition.
  - Expand only in few state variables.
Higher-order approximation (2/4)

- We exploit the symmetry of the expansion.
- In a first-order, we only need to expand in two directions for a given state variable:
  - with respect to the person’s own state and
  - with respect to any other person’s state.
- Higher-order expansions become slightly more complicated.
Higher-order approximation (3/4)

- First-order approximation corresponds to linearization.
- Solving the system for linear coefficients can be challenging.
- There are potentially multiple solutions (Riccati-type equations).
- But the system is typically polynomial and we resort to standard methods.
• Build linear approximation.
Higher-order approximation (4/4)

- The number of coefficients grows but it remains manageable.
- Expansions of order higher than one lead to linear systems.
- Higher orders are important because
  - Effects of heterogeneity
  - Effects of stochasticity
Higher-order approximation

- Build higher-order derivatives in state variables
Uncertainty

- So far: approximate solution to the deterministic case.
- Last step: move to the stochastic economy.
- Therefore, build expansion with respect to standard deviation of shocks.
- If there are multiple shocks, we scale them proportionately.
- With higher derivatives, we get a mean-variance-skewness theory.
Uncertainty

- Build derivatives with respect to standard deviation of shocks.
- Thereby, we move from the deterministic to the stochastic economy.
• Build derivatives with respect to standard deviation of shocks.
Range of applications (2/2)

- Dynamic Programming problems
- We can solve a planner’s problem
- General form of the Dynamic Program

\[ V(X_t, z_t, \sigma) = u(C_t) + \beta E_t[V(X_{t+1}, z_{t+1}, \sigma)] \]
We see convergence.
• We analyze the "aggregate" stochastic discount factor

$$\sum_{i=1}^{l} \beta \frac{u'(c_{t+1}^i)}{u'(c_t^i)} = c(X_t, z_t, k_{t+1}, b_{t+1})$$

$$\quad + c_z^{(1)} e^{z_{t+1}} + c_z^{(2)} \text{var}(e^{z_{t+1}}) + \ldots$$

$$\quad + c_{\psi}^{(2)} \text{var}(\psi^i) + \ldots$$

• The expansion can be used to price assets.
Impact of heterogeneity

- Compared with the representative agent counterpart, the steady-state level of capital in the stochastic economy is higher.
- There is more heterogeneity and thus more reason to build up precautionary savings.
- The introduction of a bond market mitigates the effects on the steady-state.
- The second-order term consists of two parts:
  - The variance of the distribution of capital
  - The comovement of individual with aggregate capital.
- This tells us how to design solution methods. Simply adding more moments to the state space is not sufficient.
Example: Lucas tree

- Endowment economy with preference shocks.
- Preferences are given by $U = E_0 \left[ \sum \beta^j A_j B_j \frac{C_j^{1-\gamma}}{1-\gamma} \right]$.
- The stochastic processes are

  $\log(C_{t+1}) = \log(C_t) + \varepsilon_{t+1}, \; \varepsilon_{t+1} \sim N(\mu, \sigma_\varepsilon^2)$

  $\log(A_{t+1}) = \rho_A \log(A_t) + \sigma_A \eta_{t+1}, \; \eta_{t+1} \sim (0, 1)$

  $\log(B_{t+1}) = \rho_B \log(B_t) + \sigma_B \eta_{t+1}, \; \eta_{t+1}$

- Claim to tree trades at $P_t$ solvable in closed form.
- Linear law: $\frac{P_{t+1}}{C_{t+1}} = \alpha_0 + \alpha_1 \frac{P_t}{C_t} + \alpha_2 \eta_{t+1}$
- We show example for a particular parameterization.
Results

- Here are the PD-ratios for this example.
- The $R^2$ diagnostic leads to more than 98% for the linear law.
Results

- Here are the PD-ratios for this example.
- The $R^2$ diagnostic leads to more than 98% for the linear law.
Further questions?

- Feel free to contact me.
- mertens@ster.nyu.edu