Review: DP and Econometrics

Leonid Kogan

MIT, Sloan

15.450, Fall 2010
Outline

1. Dynamic Portfolio Choice
2. Financial Econometrics
Dynamic Portfolio Choice

Financial Econometrics
In models in which all options are redundant (e.g., Black-Scholes), dynamic portfolio choice is relatively easy.

Solve a static problem:
- Find the best state contingent payoff (under given utility) which is budget-feasible.
- Replicate the chosen payoff using dynamic trading in available assets.

Merton’s solution: CRRA utility with risk aversion $\gamma$, Black-Scholes model:

$$\phi_t^* = \frac{\mu - r}{\gamma \sigma^2}$$

Myopic portfolio is optimal.
**Problem**

- Consider the Black-Scholes framework with parameters \( r, \mu, \) and \( \sigma \). Your objective is to find an optimal investment strategy maximizing the expected utility of terminal portfolio value

\[
E_0 \left[ \frac{1}{1 - \gamma} (W_T)^{1-\gamma} \right]
\]

subject to a lower bound on terminal wealth:

\[
W_T \geq W
\]

1. Using the static approach, express the optimal terminal wealth as a function of the SPD.
2. Show that one can implement the optimal strategy using European options on the stock.
3. (*) Implement the optimal strategy using dynamic trading.
Dynamic programming principle.
Bellman equation.
Controlled Markov processes. Problem formulation.
Key examples: portfolio choice with time-varying moments of returns; American option pricing.
Outline

1. Dynamic Portfolio Choice
2. Financial Econometrics
Parameter Estimation

GMM

- Estimate parameters using moment restrictions.
- If the true distribution satisfies

\[ E[f(x_t, \theta_0)] = 0, \quad E[f(x_t, \theta)] \neq 0 \quad \text{if} \quad \theta \neq \theta_0 \]

estimate \( \theta_0 \) using a sample analog of the population moments

\[ \hat{E}[f(x_t, \hat{\theta})] \equiv \frac{1}{T} \sum_{t=1}^{T} f(x_t, \hat{\theta}) = 0 \]

- Which moments to choose for estimation?
MLE tells us that a particular choice of moments would work and would produce the most precise estimates.

For IID observations, MLE prescribes estimating parameters as

$$\hat{\theta} = \arg \max_{\theta} \mathbb{E} \left[ \ln p(x_t, \theta) \right]$$

In moment form, this implies

$$\sum_{t=1}^{T} \frac{\partial \ln p(x, \theta)}{\partial \theta} = 0$$

MLE is a special case of GMM with a particular choice of moments, based on the pdf.
MLE for Dependent Observations

- MLE approach works even if observations are dependent.
- Consider a time series \( x_t, x_{t+1}, \ldots \) and assume that the distribution of \( x_{t+1} \) depends only on \( L \) lags: \( x_t, \ldots, x_{t+1-L} \).
- Log likelihood conditional on the first \( L \) observations:
  \[
  \hat{\theta} = \arg \max_{\theta} \mathcal{L}(\theta) = \arg \max_{\theta} \left\{ \sum_{t=L}^{T-1} \ln p(x_{t+1} | x_t, \ldots, x_{t+1-L}; \theta) \right\}
  \]
- **AR(p)** (**AutoRegressive**) time series model with IID Gaussian errors:
  \[
  x_{t+1} = a_0 + a_1 x_t + \ldots a_p x_{t+1-p} + \varepsilon_{t+1}, \quad \varepsilon_{t+1} \sim \mathcal{N}(0, \sigma^2)
  \]
- Construct log likelihood:
  \[
  \mathcal{L}(\theta) = \sum_{t=p}^{T-1} -\ln \sqrt{2\pi\sigma^2} - \frac{(x_{t+1} - a_0 - a_1 x_t - \ldots a_p x_{t+1-p})^2}{2\sigma^2}
  \]
Parameter Estimation
Iterated expectations

- Another approach to forming moment conditions is to use iterated expectations.
- For example, consider a linear model
  \[ y_t = b_0 + b_1 x_t + \varepsilon_t \]
- Assume that
  \[ \mathbb{E}[\varepsilon_t | x_t] = 0 \]
- Using iterated expectations, we can form two moments
  \[ \mathbb{E}[(y_t - b_0 - b_1 x_t) \times 1] = 0 \]
  \[ \mathbb{E}[(y_t - b_0 - b_1 x_t) \times x_t] = 0 \]
- Recover standard OLS formulas.
- \( \varepsilon_t \) could be heteroscedastic, our estimator is still valid since our moment restrictions are valid.
Parameter Estimation

QMLE

- QMLE helps formulate moment conditions when the exact form of the pdf is not known.
- Pretend that errors are Gaussian and use MLE to form moment restrictions.
- Make sure that the moment restrictions we have derived are valid, based on what we know about the model.
- Intuition: we may only need limited information, e.g., a couple of moments, to estimate the parameters. No need to know the entire distribution.
- QMLE is a valid (consistent) approach, less precise than MLE but more robust.
Example: Interest Rate Model

Iterated expectations

- Interest rate model:

\[ r_{t+1} = a_0 + a_1 r_t + \varepsilon_{t+1}, \quad E(\varepsilon_{t+1}|r_t) = 0, \quad E(\varepsilon_{t+1}^2|r_t) = b_0 + b_1 r_t \]

- GMM with moment conditions derived using iterated expectations

\[
\begin{align*}
\mathbb{E} \left[ (r_{t+1} - a_0 - a_1 r_t) \times 1 \right] &= 0 \\
\mathbb{E} \left[ (r_{t+1} - a_0 - a_1 r_t) \times r_t \right] &= 0 \\
\mathbb{E} \left\{ \left[ (r_{t+1} - a_0 - a_1 r_t)^2 - b_0 - b_1 r_t \right] \times 1 \right\} &= 0 \\
\mathbb{E} \left\{ \left[ (r_{t+1} - a_0 - a_1 r_t)^2 - b_0 - b_1 r_t \right] \times r_t \right\} &= 0
\end{align*}
\]

- \((a_0, a_1)\) can be estimated from the first pair of moment conditions. Equivalent to OLS, ignore information about second moment.
Example: Interest Rate Model

QMLE

- Treat $\varepsilon_t$ as Gaussian $\mathcal{N}(0, b_0 + b_1 r_{t-1})$.
- Construct log likelihood:

$$
\mathcal{L}(\theta) = \sum_{t=1}^{T-1} -\ln \sqrt{2\pi(b_0 + b_1 r_t)} - \frac{(r_{t+1} - a_0 - a_1 r_t)^2}{2(b_0 + b_1 r_t)}
$$

- $(a_0, a_1)$ can no longer be estimated separately from $(b_0, b_1)$.
- Optimality conditions for $(a_0, a_1)$:

$$
\sum_{t=1}^{T-1} (1, r_t)^t \frac{(r_{t+1} - a_0 - a_1 r_t)}{b_0 + b_1 r_t} = 0
$$

- This is no longer OLS, but GLS. More precise estimates of $(a_0, a_1)$.
- Down-weight residuals with high variance.
Under mild regularity conditions, GMM estimates are consistent: asymptotically, as the sample size $T$ approaches infinity, $\hat{\theta} \to \theta_0$ (in probability).

Define

$$\hat{d} = \frac{\partial \hat{E}(f(x_t, \theta))}{\partial \theta'} \bigg|_{\hat{\theta}}$$
$$\hat{S} = \hat{E}[f(x_t, \hat{\theta})f(x_t, \hat{\theta})']$$

GMM estimates are asymptotically normal:

$$\sqrt{T}(\hat{\theta} - \theta_0) \Rightarrow \mathcal{N} \left[ 0, \left( \hat{d}' \hat{S}^{-1} \hat{d} \right)^{-1} \right]$$

Standard errors are based on the asymptotic var-cov matrix of the estimates,

$$T\text{Var}[^{\theta}] = \left( \hat{d}' \hat{S}^{-1} \hat{d} \right)^{-1}$$
Problem

Suppose we observe a sequence of IID random variables $X_t \geq 0$, $t = 1, \ldots, T$, with probability density

$$pdf(X) = \lambda e^{-\lambda X}, \quad X \geq 0$$

1. Write down the log-likelihood function $\mathcal{L}(\lambda)$.
2. Compute the maximum likelihood estimate $\hat{\lambda}$.
3. Derive the standard error for $\hat{\lambda}$. 
Problem

Suppose you observe a series of observations \( X_t, t = 1, \ldots, T \). You need to fit a model

\[
X_{t+1} = f(X_t, X_{t-1}; \theta) + \varepsilon_{t+1}
\]

where \( E[\varepsilon_{t+1}|X_t, X_{t-1}, \ldots, X_1] = 0 \). Innovations \( \varepsilon_{t+1} \) have zero mean conditionally on \( X_t, X_{t-1}, \ldots, X_1 \). You also know that innovations \( \varepsilon_{t+1} \) have constant conditional variance:

\[
E[\varepsilon_{t+1}^2|X_t, X_{t-1}, \ldots, X_1] = \sigma^2
\]

The parameter \( \sigma \) is not known. \( \theta \) is the scalar parameter affecting the shape of the function \( f(X_t, X_{t-1}; \theta) \).

1. Describe how to estimate the parameter \( \theta \) using the quasi maximum likelihood approach. Derive the relevant equations.
2. Derive the standard error for \( \hat{\theta} \) using GMM standard error formulas.
GMM Standard Errors

Dependent observations

- The relation

\[
\text{Var}[\hat{\theta}] = \frac{1}{T} \left( \hat{d}^{-1} \hat{S} \left( \hat{d}' \right)^{-1} \right) = \frac{1}{T} \left( \hat{d} \hat{S}^{-1} \hat{d}' \right)^{-1}
\]

is still valid. But need to modify the estimate \( \hat{S} \).

- In an infinite sample,

\[
S = \sum_{j=-\infty}^{\infty} E \left[ f(x_t, \theta_0) f(x_{t-j}, \theta_0)' \right]
\]

- Newey-West procedure for computing standard errors prescribes

\[
\hat{S} = \sum_{j=-k}^{k} \frac{k - |j|}{k} \frac{1}{T} \sum_{t=1}^{T} f(x_t, \hat{\theta}) f(x_{t-j}, \hat{\theta})' \quad \text{(Drop out-of-range terms)}
\]

- Of special importance: OLS with Newey-West errors.
Additional Results

- Delta method: distribution of $h(\hat{\theta})$ is approximately

$$\mathcal{N}(h(\theta), V(h)), \quad V(h) = \left( \frac{\partial h(\hat{\theta})}{\partial \hat{\theta}} \right)' V(\hat{\theta}) \left( \frac{\partial h(\hat{\theta})}{\partial \hat{\theta}} \right)$$

- Hypothesis testing: construct a $\chi^2$ test of the hypothesis $h(\theta) = 0$
  - Derive the var-cov of $h(\theta)$, $V(h)$.
  - Construct the test statistic

$$\xi = h(\hat{\theta})' V(h)^{-1} h(\hat{\theta}) \sim \chi^2(\text{dim } h(\hat{\theta}))$$

- Model selection: pick an order of the AR(p) model using an AIC or BIC criterion.
Bootstrap: General Principle

- Bootstrap is a re-sampling method which can be used to evaluate properties of statistical estimators.
- Bootstrap is effectively a Monte Carlo study which uses the empirical distribution as if it were the true distribution.
- Key applications of bootstrap methodology:
  - Evaluate distributional properties of complicated estimators, perform bias adjustment;
  - Improve the precision of asymptotic approximations in small samples (confidence intervals, test rejection regions, etc.)
- Bootstrap bias correction (e.g., predictive regressions):

\[ E \left[ \hat{\theta} - \theta_0 \right] \approx E_R \left[ \hat{\theta}^* - \hat{\theta} \right] \]
Basic bootstrap confidence interval. Nonparametric approach in IID samples.
For non-IID samples, use parametric bootstrap.
Problem

Suppose you observe a series of observations $X_t$, $t = 1, \ldots, T$. You need to fit a model

$$X_{t+1} = f(X_t, X_{t-1}; \theta) + \varepsilon_{t+1}$$

where $\mathbb{E}[\varepsilon_{t+1}|X_t, X_{t-1}, \ldots, X_1] = 0$. Innovations $\varepsilon_{t+1}$ have zero mean conditionally on $X_t, X_{t-1}, \ldots, X_1$. You also know that innovations $\varepsilon_{t+1}$ have constant conditional variance:

$$\mathbb{E}[\varepsilon_{t+1}^2|X_t, X_{t-1}, \ldots, X_1] = \sigma^2$$

The parameter $\sigma$ is not known. $\theta$ is the scalar parameter affecting the shape of the function $f(X_t, X_{t-1}; \theta)$.

1. Describe in detail how to use parametric bootstrap to estimate a 95% confidence interval for $\theta$.
2. Describe how to estimate the bias in your estimate of $\theta$ using parametric bootstrap.