NBER Summer Institute Econometrics Methods Lecture:

GMM and Consumption-Based Asset Pricing

Sydney C. Ludvigson, NYU and NBER

July 14, 2010

GMM and Consumption-Based Models: Themes

- Why care about consumption-based models?

GMM and Consumption-Based Models: Themes

- Why care about consumption-based models?
- True systematic risk factors are macroeconomic in nature–derived from **IMRS over consumption**–asset prices are *derived endogenously* from these.

GMM and Consumption-Based Models: Themes

- Why care about consumption-based models?
- True systematic risk factors are macroeconomic in nature–derived from **IMRS over consumption**–asset prices are *derived endogenously* from these.
- Some cons-based models work better than others, but...

GMM and Consumption-Based Models: Themes

- Why care about consumption-based models?
- True systematic risk factors are macroeconomic in nature–derived from **IMRS over consumption**–asset prices are *derived endogenously* from these.
- Some cons-based models work better than others, but...
- ...emphasize here: all models are misspecified, and macro variables often measured with error. Therefore:

GMM and Consumption-Based Models: Themes

- Why care about consumption-based models?
- True systematic risk factors are macroeconomic in nature–derived from **IMRS over consumption**–asset prices are *derived endogenously* from these.
- Some cons-based models work better than others, but...
- ...emphasize here: all models are misspecified, and macro variables often measured with error. Therefore:
  1. move away from specification tests of perfect fit (given sampling error),

GMM and Consumption-Based Models: Themes

- Why care about consumption-based models?
- True systematic risk factors are macroeconomic in nature–derived from **IMRS over consumption**–asset prices are *derived endogenously* from these.
- Some cons-based models work better than others, but...
- ...emphasize here: all models are misspecified, and macro variables often measured with error. Therefore:
  1. move away from specification tests of perfect fit (given sampling error),
  2. toward estimation and testing that recognize all models are misspecified,

GMM and Consumption-Based Models: Themes

- Why care about consumption-based models?
- True systematic risk factors are macroeconomic in nature–derived from **IMRS over consumption**–asset prices are *derived endogenously* from these.
- Some cons-based models work better than others, but...
- ...emphasize here: all models are misspecified, and macro variables often measured with error. Therefore:
  1. move away from specification tests of perfect fit (given sampling error),
  2. toward estimation and testing that recognize all models are misspecified,
  3. toward methods permit comparison of *magnitude* of misspecification among multiple, competing macro models.

GMM and Consumption-Based Models: Themes

- Why care about consumption-based models?
- True systematic risk factors are macroeconomic in nature–derived from **IMRS over consumption**–asset prices are *derived endogenously* from these.
- Some cons-based models work better than others, but...
- ...emphasize here: all models are misspecified, and macro variables often measured with error. Therefore:
  1. move away from specification tests of perfect fit (given sampling error),
  2. toward estimation and testing that recognize all models are misspecified,
  3. toward methods permit comparison of *magnitude* of misspecification among multiple, competing macro models.
- Themes are important in choosing *which* methods to use.

- GMM estimation of classic representative agent, CRRA utility model

GMM and Consumption-Based Models: Outline

- GMM estimation of classic representative agent, CRRA utility model
- Incorporating conditioning information: scaled consumption-based models

GMM and Consumption-Based Models: Outline

- GMM estimation of classic representative agent, CRRA utility model
- Incorporating conditioning information: scaled consumption-based models
- Generalizations of CRRA utility: recursive utility

GMM and Consumption-Based Models: Outline

- GMM estimation of classic representative agent, CRRA utility model
- Incorporating conditioning information: scaled consumption-based models
- Generalizations of CRRA utility: recursive utility
  - Semi-nonparametric minimum distance estimators: unrestricted LOM for data

GMM and Consumption-Based Models: Outline

- GMM estimation of classic representative agent, CRRA utility model
- Incorporating conditioning information: scaled consumption-based models
- Generalizations of CRRA utility: recursive utility
  - Semi-nonparametric minimum distance estimators: unrestricted LOM for data
  - Simulation methods: restricted LOM

GMM and Consumption-Based Models: Outline

- GMM estimation of classic representative agent, CRRA utility model
- Incorporating conditioning information: scaled consumption-based models
- Generalizations of CRRA utility: recursive utility
  - Semi-nonparametric minimum distance estimators: unrestricted LOM for data
  - Simulation methods: restricted LOM
- Consumption-based asset pricing: concluding thoughts

- Economic model implies set of r population moment restrictions

$$E\{\underbrace{\mathbf{h}\left(\boldsymbol{\theta}, \mathbf{w}_t\right)}_{(r\times 1)}\} = 0$$

GMM Review (Hansen, 1982)

- Economic model implies set of r population moment restrictions

$$E\{\underbrace{\mathbf{h}\left(\boldsymbol{\theta}, \mathbf{w}_t\right)}_{(r\times 1)}\} = 0$$

- $\mathbf{w}_t$ is an $h \times 1$ vector of variables known at t

GMM Review (Hansen, 1982)

- Economic model implies set of r population moment restrictions

$$E\{\underbrace{\mathbf{h}\left(\boldsymbol{\theta}, \mathbf{w}_t\right)}_{(r \times 1)}\} = 0$$

- $\mathbf{w}_t$ is an $h \times 1$ vector of variables known at t
- $\boldsymbol{\theta}$ is an $a \times 1$ vector of coefficients

GMM Review (Hansen, 1982)

- Economic model implies set of r population moment restrictions

$$E\{\underbrace{\mathbf{h}\left(\boldsymbol{\theta}, \mathbf{w}_t\right)}_{(r\times 1)}\} = 0$$

- $\mathbf{w}_t$ is an $h \times 1$ vector of variables known at t
- $\boldsymbol{\theta}$ is an $a \times 1$ vector of coefficients
- Idea: choose $\boldsymbol{\theta}$ to make the sample moment as close as possible to the population moment.

GMM Review (Hansen, 1982)

- Sample moments:

$$\underbrace{\mathbf{g}(\boldsymbol{\theta};\mathbf{y}_T)}_{(r\times 1)} \equiv (1/T)\sum_{t=1}^{T}\mathbf{h}\left(\boldsymbol{\theta},\mathbf{w}_t\right),$$

GMM Review (Hansen, 1982)

- Sample moments:

$$\underbrace{\mathbf{g}(\boldsymbol{\theta};\mathbf{y}_T)}_{(r\times 1)} \equiv (1/T)\sum_{t=1}^{T}\mathbf{h}\left(\boldsymbol{\theta},\mathbf{w}_t\right),$$

- $\mathbf{y}_T \equiv \left(\mathbf{w}'_T, \mathbf{w}'_{T-1}, ...\mathbf{w}'_1\right)'$ is a $T \cdot h \times 1$ vector of observations.

GMM Review (Hansen, 1982)

- Sample moments:

$$\underbrace{\mathbf{g}(\boldsymbol{\theta};\mathbf{y}_T)}_{(r\times 1)} \equiv (1/T)\sum_{t=1}^{T}\mathbf{h}\left(\boldsymbol{\theta},\mathbf{w}_t\right),$$

- $\mathbf{y}_T \equiv \left(\mathbf{w}_T', \mathbf{w}_{T-1}', ...\mathbf{w}_1'\right)'$ is a $T \cdot h \times 1$ vector of observations.

- The GMM estimator $\widehat{\boldsymbol{\theta}}$ minimizes the scalar

$$Q\left(\boldsymbol{\theta};\boldsymbol{y}_{\mathbf{T}}\right) = \underset{(1\times r)}{[\mathbf{g}(\boldsymbol{\theta};\mathbf{y}_T)]'}\underset{(r\times r)}{\mathbf{W}_T}\underset{(r\times 1)}{[\mathbf{g}(\boldsymbol{\theta};\boldsymbol{y}_T)]}, \tag{1}$$

- $\{\mathbf{W}_T\}_{T=1}^{\infty}$ a sequence of $r \times r$ positive definite matrices which may be a function of the data, $\mathbf{y}_T$.

GMM Review (Hansen, 1982)

- Sample moments:

$$\underbrace{\mathbf{g}(\boldsymbol{\theta};\mathbf{y}_T)}_{(r\times 1)} \equiv (1/T)\sum_{t=1}^{T}\mathbf{h}\left(\boldsymbol{\theta},\mathbf{w}_t\right),$$

- $\mathbf{y}_T \equiv \left(\mathbf{w}'_T, \mathbf{w}'_{T-1}, ...\mathbf{w}'_1\right)'$ is a $T \cdot h \times 1$ vector of observations.

- The GMM estimator $\widehat{\boldsymbol{\theta}}$ minimizes the scalar

$$Q\left(\boldsymbol{\theta};\boldsymbol{y}_{\mathbf{T}}\right) = \underbrace{[\mathbf{g}(\boldsymbol{\theta};\mathbf{y}_T)]'}_{(1\times r)}\underbrace{\mathbf{W}_T}_{(r\times r)}\underbrace{[\mathbf{g}(\boldsymbol{\theta};\boldsymbol{y}_T)]}_{(r\times 1)}, \qquad (1)$$

- $\{\mathbf{W}_T\}_{T=1}^{\infty}$ a sequence of $r \times r$ positive definite matrices which may be a function of the data, $\mathbf{y}_T$.
- If $r = a$, $\boldsymbol{\theta}$ estimated by setting each $\mathbf{g}(\boldsymbol{\theta};\mathbf{y}_T)$ to zero.

GMM Review (Hansen, 1982)

- Sample moments:

$$\underbrace{\mathbf{g}(\boldsymbol{\theta};\mathbf{y}_T)}_{(r\times 1)} \equiv (1/T)\sum_{t=1}^{T}\mathbf{h}\left(\boldsymbol{\theta},\mathbf{w}_t\right),$$

- $\mathbf{y}_T \equiv \left(\mathbf{w}'_T, \mathbf{w}'_{T-1}, ...\mathbf{w}'_1\right)'$ is a $T \cdot h \times 1$ vector of observations.

- The GMM estimator $\widehat{\boldsymbol{\theta}}$ minimizes the scalar

$$Q\left(\boldsymbol{\theta};\boldsymbol{y}_{\mathbf{T}}\right) = \underbrace{[\mathbf{g}(\boldsymbol{\theta};\mathbf{y}_T)]'}_{(1\times r)} \underbrace{\mathbf{W}_T}_{(r\times r)} \underbrace{[\mathbf{g}(\boldsymbol{\theta};\boldsymbol{y}_T)]}_{(r\times 1)}, \tag{1}$$

- $\{\mathbf{W}_T\}_{T=1}^{\infty}$ a sequence of $r \times r$ positive definite matrices which may be a function of the data, $\mathbf{y}_T$.
- If $r = a$, $\boldsymbol{\theta}$ estimated by setting each $\mathbf{g}(\boldsymbol{\theta};\mathbf{y}_T)$ to zero.
- GMM refers to use of (1) to estimate $\boldsymbol{\theta}$ when $r > a$.

- Asym. properties (Hansen 1982): $\widehat{\theta}$ consistent, asym. normal.

GMM Review (Hansen, 1982)

- Asym. properties (Hansen 1982): $\widehat{\theta}$ consistent, asym. normal.
- Optimal weighting $\mathbf{W}_T = \mathbf{S}^{-1}$

$$\underset{r\times r}{\mathbf{S}} = \sum_{j=-\infty}^{\infty} E\left\{ \underbrace{\left[\mathbf{h}\left(\boldsymbol{\theta}_o, \mathbf{w}_t\right)\right]}_{r\times 1} \underbrace{\left[\mathbf{h}\left(\boldsymbol{\theta}_o, \mathbf{w}_{t-j}\right)\right]'}_{1\times r} \right\}.$$

GMM Review (Hansen, 1982)

- Asym. properties (Hansen 1982): $\widehat{\theta}$ consistent, asym. normal.
- Optimal weighting $\mathbf{W}_T = \mathbf{S}^{-1}$

$$\underset{r\times r}{\mathbf{S}} = \sum_{j=-\infty}^{\infty} E\left\{ \underbrace{\left[\mathbf{h}\left(\boldsymbol{\theta}_o, \mathbf{w}_t\right)\right]}_{r\times 1} \underbrace{\left[\mathbf{h}\left(\boldsymbol{\theta}_o, \mathbf{w}_{t-j}\right)\right]'}_{1\times r} \right\}.$$

- In many asset pricing applications, it is inappropriate to use $\mathbf{W}_T = \mathbf{S}^{-1}$ (see below).

- $\widehat{\mathbf{S}}_T$ depends on $\widehat{\boldsymbol{\theta}}_T$ which depends on $\widehat{\mathbf{S}}_T$. Employ an iterative procedure:

- $\widehat{\mathbf{S}}_T$ depends on $\widehat{\boldsymbol{\theta}}_T$ which depends on $\widehat{\mathbf{S}}_T$. Employ an iterative procedure:

1. Obtain an initial estimate of $\boldsymbol{\theta} = \widehat{\boldsymbol{\theta}}_T^{(1)}$, by minimizing $Q(\boldsymbol{\theta}; \boldsymbol{y}_{\mathbf{T}})$ subject to arbitrary weighting matrix, e.g., $\mathbf{W} = \mathbf{I}$.

2. Use $\widehat{\boldsymbol{\theta}}_T^{(1)}$ to obtain initial estimate of $\mathbf{S} = \widehat{\mathbf{S}}_T^{(1)}$.

3. Re-minimize $Q(\boldsymbol{\theta}; \boldsymbol{y}_{\mathbf{T}})$ using initial estimate $\widehat{\mathbf{S}}_T^{(1)}$; obtain new estimate $\widehat{\boldsymbol{\theta}}_T^{(2)}$.

4. Continue iterating until convergence, or stop. (Estimators have same asym. dist. but finite sample properties differ.)

Classic Example: Hansen and Singleton (1982)

- Investors maximize utility

$$\max_{C_t} \; E_t \left[\sum_{i=0}^{\infty} \beta^i u\left(C_{t+i}\right) \right]$$

Classic Example: Hansen and Singleton (1982)

- Investors maximize utility

$$\max_{C_t} \; E_t \left[\sum_{i=0}^{\infty} \beta^i u\left(C_{t+i}\right) \right]$$

- Power (isoelastic) utility

$$\begin{cases} u\left(C_t\right) = \frac{C_t^{1-\gamma}}{1-\gamma} & \gamma > 0 \\ u\left(C_t\right) = \ln(C_t) & \gamma = 1 \end{cases} \tag{2}$$

Classic Example: Hansen and Singleton (1982)

- Investors maximize utility

$$\max_{C_t} \; E_t \left[\sum_{i=0}^{\infty} \beta^i u\left(C_{t+i}\right) \right]$$

- Power (isoelastic) utility

$$\begin{cases} u\left(C_t\right) = \frac{C_t^{1-\gamma}}{1-\gamma} & \gamma > 0 \\ u\left(C_t\right) = \ln(C_t) & \gamma = 1 \end{cases} \tag{2}$$

- N assets $=>$ N first-order conditions

$$C_t^{-\gamma} = \beta E_t \left\{ \left(1 + \Re_{i,t+1}\right) C_{t+1}^{-\gamma} \right\} \qquad i = 1, ..., N. \tag{3}$$

Asset Pricing Example: Hansen and Singleton (1982)

- Re-write moment conditions

$$0 = E_t \left\{ 1 - \beta \left[\left(1 + \Re_{i,t+1}\right) \frac{C_{t+1}^{-\gamma}}{C_t^{-\gamma}} \right] \right\}. \tag{4}$$

- Re-write moment conditions

$$0 = E_t \left\{ 1 - \beta \left[(1 + \Re_{i,t+1}) \frac{C_{t+1}^{-\gamma}}{C_t^{-\gamma}} \right] \right\}. \tag{4}$$

- 2 params to estimate: β and γ, so $\boldsymbol{\theta} = (\beta, \gamma)'$.

- Re-write moment conditions

$$0 = E_t \left\{ 1 - \beta \left[\left(1 + \Re_{i,t+1}\right) \frac{C_{t+1}^{-\gamma}}{C_t^{-\gamma}} \right] \right\}. \tag{4}$$

- 2 params to estimate: β and γ, so $\boldsymbol{\theta} = (\beta, \gamma)'$.
- $\mathbf{x}_t^*$ denotes info set of investors

$$0 = E \left\{ \left[1 - \left\{ \beta \left(1 + \Re_{i,t+1}\right) C_{t+1}^{-\gamma} / C_t^{-\gamma} \right\} \right] \left| \mathbf{x}_t^* \right. \right\} \qquad i = 1, \ldots N \tag{5}$$

- Re-write moment conditions

$$0 = E_t \left\{ 1 - \beta \left[\left(1 + \Re_{i,t+1}\right) \frac{C_{t+1}^{-\gamma}}{C_t^{-\gamma}} \right] \right\}. \tag{4}$$

- 2 params to estimate: β and γ, so $\boldsymbol{\theta} = (\beta, \gamma)'$.
- $\mathbf{x}_t^*$ denotes info set of investors

$$0 = E \left\{ \left[1 - \left\{ \beta \left(1 + \Re_{i,t+1}\right) C_{t+1}^{-\gamma} / C_t^{-\gamma} \right\} \right] | \mathbf{x}_t^* \right\} \qquad i = 1, \ldots N \tag{5}$$

- $\mathbf{x}_t \subset \mathbf{x}_t^*$. Conditional model (5) $=>$ unconditional model:

$$0 = E \left\{ \left[1 - \left\{ \beta \left(1 + \Re_{i,t+1}\right) \frac{C_{t+1}^{-\gamma}}{C_t^{-\gamma}} \right\} \right] \mathbf{x}_t \right\} \qquad i = 1, \ldots N \tag{6}$$

Asset Pricing Example: Hansen and Singleton (1982)

- Let $\mathbf{x}_t$ be $M \times 1$. Then $r = N \cdot M$ and,

$$\underset{r\times 1}{\mathbf{h}\left(\boldsymbol{\theta}, \mathbf{w}_t\right)} = \begin{bmatrix} \left[1-\beta\left\{\left(1+\Re_{1,t+1}\right)\frac{C_{t+1}^{-\gamma}}{C_t^{-\gamma}}\right\}\right]\mathbf{x}_t \\ \left[1-\beta\left\{\left(1+\Re_{2,t+1}\right)\frac{C_{t+1}^{-\gamma}}{C_t^{-\gamma}}\right\}\right]\mathbf{x}_t \\ \cdot \\ \cdot \\ \cdot \\ \left[1-\beta\left\{\left(1+\Re_{N,t+1}\right)\frac{C_{t+1}^{-\gamma}}{C_t^{-\gamma}}\right\}\right]\mathbf{x}_t \end{bmatrix} \quad (7)$$

Asset Pricing Example: Hansen and Singleton (1982)

- Let $\mathbf{x}_t$ be $M \times 1$. Then $r = N \cdot M$ and,

$$\underset{r\times 1}{\mathbf{h}}\left(\boldsymbol{\theta}, \mathbf{w}_t\right) = \begin{bmatrix} \left[1 - \beta\left\{\left(1 + \Re_{1,t+1}\right)\frac{C_{t+1}^{-\gamma}}{C_t^{-\gamma}}\right\}\right]\mathbf{x}_t \\ \left[1 - \beta\left\{\left(1 + \Re_{2,t+1}\right)\frac{C_{t+1}^{-\gamma}}{C_t^{-\gamma}}\right\}\right]\mathbf{x}_t \\ \cdot \\ \cdot \\ \cdot \\ \left[1 - \beta\left\{\left(1 + \Re_{N,t+1}\right)\frac{C_{t+1}^{-\gamma}}{C_t^{-\gamma}}\right\}\right]\mathbf{x}_t \end{bmatrix} \tag{7}$$

- Model can be estimated, tested as long as $r \geq 2$.

Asset Pricing Example: Hansen and Singleton (1982)

- Let $\mathbf{x}_t$ be $M \times 1$. Then $r = N \cdot M$ and,

$$\underset{r\times 1}{\mathbf{h}}(\boldsymbol{\theta}, \mathbf{w}_t) = \begin{bmatrix} \left[1 - \beta\left\{(1 + \Re_{1,t+1})\frac{C_{t+1}^{-\gamma}}{C_t^{-\gamma}}\right\}\right]\mathbf{x}_t \\ \left[1 - \beta\left\{(1 + \Re_{2,t+1})\frac{C_{t+1}^{-\gamma}}{C_t^{-\gamma}}\right\}\right]\mathbf{x}_t \\ \cdot \\ \cdot \\ \cdot \\ \left[1 - \beta\left\{(1 + \Re_{N,t+1})\frac{C_{t+1}^{-\gamma}}{C_t^{-\gamma}}\right\}\right]\mathbf{x}_t \end{bmatrix} \tag{7}$$

- Model can be estimated, tested as long as $r \geq 2$.
- Take sample mean of (7) to get $\mathbf{g}(\boldsymbol{\theta}; \mathbf{y}_T)$, minimize

$$\min_{\boldsymbol{\theta}} Q(\boldsymbol{\theta}; \mathbf{y}_T) = \underbrace{[\mathbf{g}(\boldsymbol{\theta}; \mathbf{y}_T)]'}_{1\times r} \underset{r\times r}{\mathbf{W}_T} \underbrace{[\mathbf{g}(\boldsymbol{\theta}; \boldsymbol{y}_T)]}_{r\times 1}$$

- Test of Over-identifying (OID) restrictions:

$$TQ\left(\widehat{\boldsymbol{\theta}};\mathbf{y}_T\right)\overset{a}{\sim}\chi^2(r-a)$$

Asset Pricing Example: Hansen and Singleton (1982)

- Test of Over-identifying (OID) restrictions:

$$TQ\left(\widehat{\boldsymbol{\theta}};\mathbf{y}_T\right)\overset{a}{\sim}\chi^2(r-a)$$

- HS use lags of cons growth and returns in $\mathbf{x}_t$; index and industry returns, NDS expenditures.
- Estimates of $\beta \approx .99$, RRA low $=$ .35 to .999. No equity premium puzzle! But....

Asset Pricing Example: Hansen and Singleton (1982)

- Test of Over-identifying (OID) restrictions:

$$TQ\left(\widehat{\boldsymbol{\theta}};\mathbf{y}_T\right) \overset{a}{\sim} \chi^2(r-a)$$

- HS use lags of cons growth and returns in $\mathbf{x}_t$; index and industry returns, NDS expenditures.
- Estimates of $\beta \approx .99$, RRA low $=$ .35 to .999. No equity premium puzzle! But....
- ...model is strongly rejected according to OID test.

Asset Pricing Example: Hansen and Singleton (1982)

- Test of Over-identifying (OID) restrictions:

$$TQ\left(\widehat{\boldsymbol{\theta}};\mathbf{y}_T\right)\overset{a}{\sim}\chi^2(r-a)$$

- HS use lags of cons growth and returns in $\mathbf{x}_t$; index and industry returns, NDS expenditures.
- Estimates of $\beta \approx .99$, RRA low $=$ .35 to .999. No equity premium puzzle! But....
- ...model is strongly rejected according to OID test.
- Campbell, Lo, MacKinlay (1997): OID rejections stronger whenever stock returns and commercial paper are included as test returns. Why?

- Test of Over-identifying (OID) restrictions:

$$TQ\left(\widehat{\boldsymbol{\theta}};\mathbf{y}_T\right)\overset{a}{\sim}\chi^2(r-a)$$

- HS use lags of cons growth and returns in $\mathbf{x}_t$; index and industry returns, NDS expenditures.
- Estimates of $\beta \approx .99$, RRA low $=$ .35 to .999. No equity premium puzzle! But....
- ...model is strongly rejected according to OID test.
- Campbell, Lo, MacKinlay (1997): OID rejections stronger whenever stock returns and commercial paper are included as test returns. Why?
- Model cannot capture predictable variation in *excess returns* over commercial paper $\Rightarrow$

Asset Pricing Example: Hansen and Singleton (1982)

- Test of Over-identifying (OID) restrictions:

$$TQ\left(\widehat{\boldsymbol{\theta}};\mathbf{y}_T\right) \overset{a}{\sim} \chi^2(r-a)$$

- HS use lags of cons growth and returns in $\mathbf{x}_t$; index and industry returns, NDS expenditures.
- Estimates of $\beta \approx .99$, RRA low $=$ .35 to .999. No equity premium puzzle! But....
- ...model is strongly rejected according to OID test.
- Campbell, Lo, MacKinlay (1997): OID rejections stronger whenever stock returns and commercial paper are included as test returns. Why?
- Model cannot capture predictable variation in *excess returns* over commercial paper $\Rightarrow$
- Researchers have turned to other models of preferences.

More GMM results: Euler Equation Errors

- Results in HS use conditioning info $\mathbf{x}_t$–scaled returns.

More GMM results: Euler Equation Errors

- Results in HS use conditioning info $\mathbf{x}_t$–scaled returns.
- Another limitation with classic CCAPM: large *unconditional* Euler equation (pricing) errors *even* when params freely chosen.

More GMM results: Euler Equation Errors

- Results in HS use conditioning info $\mathbf{x}_t$–scaled returns.
- Another limitation with classic CCAPM: large *unconditional* Euler equation (pricing) errors *even* when params freely chosen.
- Let $M_{t+1} = \beta(C_{t+1}/C_t)^{-\gamma}$. Define Euler equation errors:

$$e_R^j \equiv E[M_{t+1}R_{t+1}^j] - 1$$

$$e_X^j \equiv E[M_{t+1}(R_{t+1}^j - R_{t+1}^f)]$$

More GMM results: Euler Equation Errors

- Results in HS use conditioning info $\mathbf{x}_t$–scaled returns.
- Another limitation with classic CCAPM: large *unconditional* Euler equation (pricing) errors *even* when params freely chosen.
- Let $M_{t+1} = \beta(C_{t+1}/C_t)^{-\gamma}$. Define Euler equation errors:

$$e_R^j \equiv E[M_{t+1}R_{t+1}^j] - 1$$

$$e_X^j \equiv E[M_{t+1}(R_{t+1}^j - R_{t+1}^f)]$$

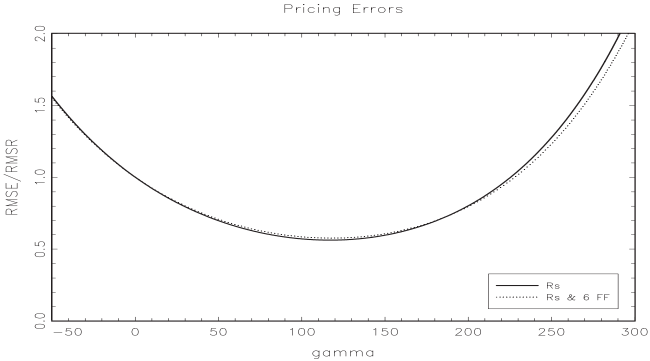
- Choose params: $\min_{\beta,\gamma} \mathbf{g}_T' \mathbf{W}_T \mathbf{g}_T$ where jth element of $\mathbf{g}_T$

$$g_{j,t}(\gamma,\beta) = \frac{1}{T}\sum_{t=1}^{T} e_{R,t}^j$$

$$g_{j,t}(\gamma) = \frac{1}{T}\sum_{t=1}^{T} e_{X,t}^j$$

Unconditional Euler Equation Errors, Excess Returns

- $e_X^j \equiv E[\beta(C_{t+1}/C_t)^{-\gamma}(R_{t+1}^j - R_{t+1}^f)] \qquad j = 1, ..., N$
- $RMSE = \sqrt{\frac{1}{N}\sum_{j=1}^N [e_X^j]^2}, \quad RMSR = \sqrt{\frac{1}{N}\sum_{j=1}^N [E(R_{t+1}^j - R_{t+1}^f)]^2}$



Source: Lettau and Ludvigson (2009). *Rs* is the excess return on CRSP-VW index over 3-Mo T-bill rate. *Rs* & 6 FF refers to this return plus 6 size and book-market sorted portfolios provided by Fama and French. For each value of γ, β is chosen to minimize the Euler equation error for the T-bill rate. U.S. quarterly data, 1954:1-2002:1.

More GMM results: Euler Equation Errors

- Magnitude of errors large, even when parameters are freely chosen to minimize errors.
- Unlike the equity premium puzzle of Mehra and Prescott (1985), large Euler eq. errors *cannot be resolved* with high risk aversion.

More GMM results: Euler Equation Errors

- Magnitude of errors large, even when parameters are freely chosen to minimize errors.
- Unlike the equity premium puzzle of Mehra and Prescott (1985), large Euler eq. errors *cannot be resolved* with high risk aversion.
- Lettau and Ludvigson (2009): Leading consumption-based asset pricing theories fail to explain the mispricing of classic CCAPM.

More GMM results: Euler Equation Errors

- Magnitude of errors large, even when parameters are freely chosen to minimize errors.
- Unlike the equity premium puzzle of Mehra and Prescott (1985), large Euler eq. errors *cannot be resolved* with high risk aversion.
- Lettau and Ludvigson (2009): Leading consumption-based asset pricing theories fail to explain the mispricing of classic CCAPM.
- Anomaly is striking b/c early evidence (e.g., Hansen & Singleton) that the classic model's Euler equations were violated provided the impetus for developing these newer models.

More GMM results: Euler Equation Errors

- Magnitude of errors large, even when parameters are freely chosen to minimize errors.
- Unlike the equity premium puzzle of Mehra and Prescott (1985), large Euler eq. errors *cannot be resolved* with high risk aversion.
- Lettau and Ludvigson (2009): Leading consumption-based asset pricing theories fail to explain the mispricing of classic CCAPM.
- Anomaly is striking b/c early evidence (e.g., Hansen & Singleton) that the classic model's Euler equations were violated provided the impetus for developing these newer models.
- Results imply data on consumption and asset returns not jointly lognormal!

- Asset pricing applications often require $\mathbf{W}_T \neq \mathbf{S}^{-1}$. Why?

- Asset pricing applications often require $\mathbf{W}_T \neq \mathbf{S}^{-1}$. Why?
- One reason: assessing specification error, comparing models.

- Asset pricing applications often require $\mathbf{W}_T \neq \mathbf{S}^{-1}$. Why?
- One reason: assessing specification error, comparing models.
- Consider two estimated models of SDF, e.g.,
  1. CCAPM: $M_{t+1}^{(1)} = \beta(C_{t+1}/C_t)^{-\gamma}$, OID restricts not rejected
  2. CAPM: $M_{t+1}^{(2)} = a + bR_{m,t+1}$, OID restricts rejected

GMM Asset Pricing With Non-Optimal Weighting

Comparing specification error: Hansen and Jagannathan, 1997

- Asset pricing applications often require $\mathbf{W}_T \neq \mathbf{S}^{-1}$. Why?
- One reason: assessing specification error, comparing models.
- Consider two estimated models of SDF, e.g.,
  1. CCAPM: $M_{t+1}^{(1)} = \beta(C_{t+1}/C_t)^{-\gamma}$, OID restricts not rejected
  2. CAPM: $M_{t+1}^{(2)} = a + bR_{m,t+1}$, OID restricts rejected
- May we conclude Model 1 is superior?

GMM Asset Pricing With Non-Optimal Weighting

Comparing specification error: Hansen and Jagannathan, 1997

- Asset pricing applications often require $\mathbf{W}_T \neq \mathbf{S}^{-1}$. Why?
- One reason: assessing specification error, comparing models.
- Consider two estimated models of SDF, e.g.,
  1. CCAPM: $M_{t+1}^{(1)} = \beta(C_{t+1}/C_t)^{-\gamma}$, OID restricts not rejected
  2. CAPM: $M_{t+1}^{(2)} = a + bR_{m,t+1}$, OID restricts rejected
- May we conclude Model 1 is superior?
- No. Hansen's J-test of OID restricts depends on model specific $\mathbf{S}$: $J = \mathbf{g}_T' \mathbf{S}^{-1} \mathbf{g}_T$.

- Asset pricing applications often require $\mathbf{W}_T \neq \mathbf{S}^{-1}$. Why?
- One reason: assessing specification error, comparing models.
- Consider two estimated models of SDF, e.g.,
  1. CCAPM: $M_{t+1}^{(1)} = \beta(C_{t+1}/C_t)^{-\gamma}$, OID restricts not rejected
  2. CAPM: $M_{t+1}^{(2)} = a + bR_{m,t+1}$, OID restricts rejected
- May we conclude Model 1 is superior?
- No. Hansen's J-test of OID restricts depends on model specific $\mathbf{S}$: $J = \mathbf{g}_T'\mathbf{S}^{-1}\mathbf{g}_T$.
- Model 1 can look better simply b/c the SDF and pricing errors $\mathbf{g}_T$ are more volatile, not b/c pricing errors are lower.

- HJ: compare models $M_t(\boldsymbol{\theta})$ using distance metric:

$$\text{Dist}_T(\boldsymbol{\theta}) = \sqrt{\min_{\boldsymbol{\theta}} \mathbf{g}_T(\boldsymbol{\theta})' \, \mathbf{G}_T^{-1} \mathbf{g}_T(\boldsymbol{\theta})}, \qquad \mathbf{G}_T \equiv \frac{1}{T}\sum_{t=1}^{T} \underbrace{R_t R_t'}_{N\times N}$$

$$\mathbf{g}_T(\boldsymbol{\theta}) \equiv \frac{1}{T}\sum_{t=1}^{T} \left[M_t(\boldsymbol{\theta}) \mathbf{R}_t - \mathbf{1}_N \right]$$

- HJ: compare models $M_t(\boldsymbol{\theta})$ using distance metric:

$$\mathrm{Dist}_T(\boldsymbol{\theta}) = \sqrt{\min_{\boldsymbol{\theta}} \mathbf{g}_T(\boldsymbol{\theta})' \mathbf{G}_T^{-1} \mathbf{g}_T(\boldsymbol{\theta})}, \qquad \mathbf{G}_T \equiv \frac{1}{T}\sum_{t=1}^{T} \underbrace{R_t R_t'}_{N\times N}$$

$$\mathbf{g}_T(\boldsymbol{\theta}) \equiv \frac{1}{T}\sum_{t=1}^{T} \left[M_t(\boldsymbol{\theta})\mathbf{R}_t - \mathbf{1}_N\right]$$

- Dist_T does not reward SDF volatility => suitable for model comparison.

- HJ: compare models $M_t(\boldsymbol{\theta})$ using distance metric:

$$\text{Dist}_T(\boldsymbol{\theta}) = \sqrt{\min_{\boldsymbol{\theta}} \mathbf{g}_T(\boldsymbol{\theta})' \, \mathbf{G}_T^{-1} \mathbf{g}_T(\boldsymbol{\theta})}, \qquad \mathbf{G}_T \equiv \frac{1}{T} \sum_{t=1}^{T} \underbrace{R_t R_t'}_{N \times N}$$

$$\mathbf{g}_T(\boldsymbol{\theta}) \equiv \frac{1}{T} \sum_{t=1}^{T} \left[M_t(\boldsymbol{\theta}) \mathbf{R}_t - \mathbf{1}_N \right]$$

- Dist_T does not reward SDF volatility => suitable for model comparison.
- Dist_T is a measure of model misspecification:

- HJ: compare models $M_t(\boldsymbol{\theta})$ using distance metric:

$$\text{Dist}_T(\boldsymbol{\theta}) = \sqrt{\min_{\boldsymbol{\theta}} \mathbf{g}_T(\boldsymbol{\theta})' \, \mathbf{G}_T^{-1} \mathbf{g}_T(\boldsymbol{\theta})}, \qquad \mathbf{G}_T \equiv \frac{1}{T} \sum_{t=1}^{T} \underbrace{R_t R_t'}_{N \times N}$$

$$\mathbf{g}_T(\boldsymbol{\theta}) \equiv \frac{1}{T} \sum_{t=1}^{T} \left[M_t(\boldsymbol{\theta}) \mathbf{R}_t - \mathbf{1}_N \right]$$

- Dist_T does not reward SDF volatility $=>$ suitable for model comparison.
- Dist_T is a measure of model misspecification:
  - Gives distance between $M_t(\boldsymbol{\theta})$ and nearest point in space of all SDFs that price assets correctly.

GMM Asset Pricing With Non-Optimal Weighting

Comparing specification error: Hansen and Jagannathan, 1997

- HJ: compare models $M_t(\boldsymbol{\theta})$ using distance metric:

$$\text{Dist}_T(\boldsymbol{\theta}) = \sqrt{\min_{\boldsymbol{\theta}} \mathbf{g}_T(\boldsymbol{\theta})' \mathbf{G}_T^{-1} \mathbf{g}_T(\boldsymbol{\theta})}, \qquad \mathbf{G}_T \equiv \frac{1}{T}\sum_{t=1}^{T} \underbrace{R_t R_t'}_{N\times N}$$

$$\mathbf{g}_T(\boldsymbol{\theta}) \equiv \frac{1}{T}\sum_{t=1}^{T}\left[M_t(\boldsymbol{\theta})\mathbf{R}_t - \mathbf{1}_N\right]$$

- Dist_T does not reward SDF volatility $=>$ suitable for model comparison.
- Dist_T is a measure of model misspecification:
  - Gives distance between $M_t(\boldsymbol{\theta})$ and nearest point in space of all SDFs that price assets correctly.
  - Gives maximum pricing error of any portfolio formed from the N assets.

- Appeal of HJ Distance metric:

- Appeal of HJ Distance metric:
  - Recognizes all models are misspecified.
  - Provides method for comparing models by assessing which is *least* misspecified.

- Appeal of HJ Distance metric:
    - Recognizes all models are misspecified.
    - Provides method for comparing models by assessing which is *least* misspecified.
- Important problem: how to compare HJ distances statistically?

- Appeal of HJ Distance metric:
  - Recognizes all models are misspecified.
  - Provides method for comparing models by assessing which is *least* misspecified.
- Important problem: how to compare HJ distances statistically?
- One possibility developed in Chen and Ludvigson (2009): White's reality check method.

- Chen and Ludvigson (2009) compare HJ distances among K competing models using White's reality check method.

- Chen and Ludvigson (2009) compare HJ distances among K competing models using White's reality check method.

1. Take benchmark model, e.g., model with smallest squared distance $d_{1,T}^2 \equiv \min\{d_{j,T}^2\}_{j=1}^K$.
2. Null: $d_{1,T}^2 - d_{2,T}^2 \leq 0$, where $d_{2,T}^2$ is competing model with the next smallest squared distance.
3. Test statistic $\mathcal{T}^W = \sqrt{T}(d_{1,T}^2 - d_{2,T}^2)$.
4. If null is true, test statistic should not be unusually large, given sampling error.
5. Given distribution for $\mathcal{T}^W$, reject null if historical value $\widehat{\mathcal{T}}^W$ is $>$ 95th percentile.

GMM Asset Pricing With Non-Optimal Weighting

Statistical comparison of HJ distance: Chen and Ludvigson, 2009

- Chen and Ludvigson (2009) compare HJ distances among K competing models using White's reality check method.
  1. Take benchmark model, e.g., model with smallest squared distance $d_{1,T}^2 \equiv \min\{d_{j,T}^2\}_{j=1}^K$.
  2. Null: $d_{1,T}^2 - d_{2,T}^2 \leq 0$, where $d_{2,T}^2$ is competing model with the next smallest squared distance.
  3. Test statistic $\mathcal{T}^W = \sqrt{T}(d_{1,T}^2 - d_{2,T}^2)$.
  4. If null is true, test statistic should not be unusually large, given sampling error.
  5. Given distribution for $\mathcal{T}^W$, reject null if historical value $\widehat{\mathcal{T}}^W$ is $>$ 95th percentile.
- Method applies generally to **any stationary** law of motion for data, **multiple** competing possibly **nonlinear**, SDF models.

- Distribution of $\mathcal{T}^W$ is computed via block bootstrap.
  - $\mathcal{T}^W$ has complicated limiting distribution.

- Distribution of $\mathcal{T}^W$ is computed via block bootstrap.
  - $\mathcal{T}^W$ has complicated limiting distribution.
- Bootstrap works only if have a multivariate, joint, continuous, limiting distribution under null.

- Distribution of $\mathcal{T}^W$ is computed via block bootstrap.
  - $\mathcal{T}^W$ has complicated limiting distribution.
- Bootstrap works only if have a multivariate, joint, continuous, limiting distribution under null.
- Proof of limiting distributions exists for applications to most asset pricing models:

- Distribution of $\mathcal{T}^W$ is computed via block bootstrap.
  - $\mathcal{T}^W$ has complicated limiting distribution.
- Bootstrap works only if have a multivariate, joint, continuous, limiting distribution under null.
- Proof of limiting distributions exists for applications to most asset pricing models:
  - For parametric models (Hansen, Heaton, Luttmer '95)
  - For semiparametric models (Ai and Chen '07).

GMM Asset Pricing With Non-Optimal Weighting

Reasons to use identity matrix: econometric problems

- Econometric problems: near singular $\mathbf{S}_T^{-1}$ or $\mathbf{G}_T^{-1}$.
    - Asset returns are highly correlated.
    - We have large N and modest T.
    - If $T < N$ covariance matrix for N asset returns is singular. Unless $T >> N$, matrix can be near-singular.

- Original test assets may have economically meaningful characteristics (e.g., size, value).

- Original test assets may have economically meaningful characteristics (e.g., size, value).
- Using $\mathbf{W}_T = \mathbf{S}_T^{-1}$ or $\mathbf{G}_T^{-1}$ same as using $\mathbf{W}_T = \mathbf{I}$ and doing GMM on *re-weighted portfolios* of original test assets.
  - Triangular factorization $\mathbf{S}^{-1} = (\mathbf{P}'\mathbf{P})$, $\mathbf{P}$ lower triangular

$$\min \mathbf{g}_T' \mathbf{S}^{-1} \mathbf{g}_T \Leftrightarrow (\mathbf{g}_T' \mathbf{P}') \mathbf{I} (\mathbf{P} \mathbf{g}_T)$$

GMM Asset Pricing With Non-Optimal Weighting

Reasons to use identity matrix: economically interesting portfolios

- Original test assets may have economically meaningful characteristics (e.g., size, value).
- Using $\mathbf{W}_T = \mathbf{S}_T^{-1}$ or $\mathbf{G}_T^{-1}$ same as using $\mathbf{W}_T = \mathbf{I}$ and doing GMM on *re-weighted portfolios* of original test assets.
    - Triangular factorization $\mathbf{S}^{-1} = (\mathbf{P}'\mathbf{P})$, $\mathbf{P}$ lower triangular

$$\min \mathbf{g}_T' \mathbf{S}^{-1} \mathbf{g}_T \Leftrightarrow (\mathbf{g}_T' \mathbf{P}') \mathbf{I} (\mathbf{P} \mathbf{g}_T)$$

- Re-weighted portfolios may not provide large spread in average returns.

- Original test assets may have economically meaningful characteristics (e.g., size, value).
- Using $\mathbf{W}_T = \mathbf{S}_T^{-1}$ or $\mathbf{G}_T^{-1}$ same as using $\mathbf{W}_T = \mathbf{I}$ and doing GMM on *re-weighted portfolios* of original test assets.
  - Triangular factorization $\mathbf{S}^{-1} = (\mathbf{P}'\mathbf{P})$, $\mathbf{P}$ lower triangular

$$\min \mathbf{g}_T'\mathbf{S}^{-1}\mathbf{g}_T \Leftrightarrow (\mathbf{g}_T'\mathbf{P}')\mathbf{I}(\mathbf{P}\mathbf{g}_T)$$

- Re-weighted portfolios may not provide large spread in average returns.
- May imply implausible long and short positions in test assets.

GMM Asset Pricing With Non-Optimal Weighting

Reasons not to use $\mathbf{W}_T = \mathbf{I}$: objective function dependence on test asset choice

- Using $\mathbf{W}_T = [E_T(\mathbf{R}'\mathbf{R})]^{-1}$, GMM objective function is invariant to initial choice of test assets.

- Using $\mathbf{W}_T = [E_T(\mathbf{R}'\mathbf{R})]^{-1}$, GMM objective function is invariant to initial choice of test assets.

- Form a portfolio, $\mathbf{AR}$ from initial returns $\mathbf{R}$. (Note, portfolio weights sum to 1 so $\mathbf{A1}_N = \mathbf{1}_N$).

$$
\begin{aligned}
& \left[E\left(M\mathbf{R}\right) - \mathbf{1}_N\right]' E\left(\mathbf{R}\mathbf{R}'\right)^{-1}\left[E\left(M\mathbf{R} - \mathbf{1}_N\right)\right] \\
= \quad & \left[E\left(M\mathbf{A}\mathbf{R}\right) - \mathbf{A1}_N\right]' E\left(\mathbf{A}\mathbf{R}\mathbf{R}'\mathbf{A}\right)^{-1}\left[E\left(M\mathbf{A}\mathbf{R} - \mathbf{A1}_N\right)\right].
\end{aligned}
$$

- Using $\mathbf{W}_T = [E_T(\mathbf{R}'\mathbf{R})]^{-1}$, GMM objective function is invariant to initial choice of test assets.

- Form a portfolio, $\mathbf{AR}$ from initial returns $\mathbf{R}$. (Note, portfolio weights sum to 1 so $\mathbf{A1}_N = \mathbf{1}_N$).

$$
\begin{aligned}
& [E(M\mathbf{R}) - \mathbf{1}_N]' E\left(\mathbf{R}\mathbf{R}'\right)^{-1} [E(M\mathbf{R} - \mathbf{1}_N)] \\
= \quad & [E(M\mathbf{AR}) - \mathbf{A1}_N]' E\left(\mathbf{A}\mathbf{R}\mathbf{R}'\mathbf{A}\right)^{-1} [E(M\mathbf{AR} - \mathbf{A1}_N)] .
\end{aligned}
$$

- With $\mathbf{W}_T = \mathbf{I}$ or other fixed weighting, GMM objective depends on choice of test assets.

More Complex Preferences: Scaled Consumption-Based Models

- Consumption-based models may be approximated:

$$M_{t+1} \approx a_t + b_t \Delta c_{t+1}, \qquad c_{t+1} \equiv \ln(C_{t+1})$$

More Complex Preferences: Scaled Consumption-Based Models

- Consumption-based models may be approximated:

$$M_{t+1} \approx a_t + b_t \Delta c_{t+1}, \qquad c_{t+1} \equiv \ln(C_{t+1})$$

- Example: Classic CCAPM with CRRA utility

$$u(C_t) = \frac{C_t^{1-\gamma}}{1-\gamma} \quad \Rightarrow \quad M_{t+1} \approx \underbrace{\beta}_{a_t = a_0} - \underbrace{\beta\gamma}_{b_t = b_0} \Delta c_{t+1}$$

More Complex Preferences: Scaled Consumption-Based Models

- Consumption-based models may be approximated:

$$M_{t+1} \approx a_t + b_t \Delta c_{t+1}, \qquad c_{t+1} \equiv \ln(C_{t+1})$$

- Example: Classic CCAPM with CRRA utility

$$u(C_t) = \frac{C_t^{1-\gamma}}{1-\gamma} \quad \Rightarrow \quad M_{t+1} \approx \underbrace{\beta}_{a_t = a_0} - \underbrace{\beta\gamma}_{b_t = b_0} \Delta c_{t+1}$$

- Model with habit and time-varying risk aversion: Campbell and Cochrane '99, Menzly et. al '04

$$u(C_t, S_t) = \frac{(C_t S_t)^{1-\gamma}}{1-\gamma}, \qquad S_{t+1} \equiv \frac{C_t - X_t}{C_t}$$

$$\Rightarrow M_{t+1} \approx \beta \Big(\underbrace{1 - \gamma(\phi - 1)(s_t - \bar{s})}_{a_t} - \underbrace{\gamma(1 + \psi(s_t))}_{b_t} \Delta c_{t+1} \Big)$$

More Complex Preferences: Scaled Consumption-Based Models

- Consumption-based models may be approximated:

$$M_{t+1} \approx a_t + b_t \Delta c_{t+1}, \qquad c_{t+1} \equiv \ln(C_{t+1})$$

- Example: Classic CCAPM with CRRA utility

$$u(C_t) = \frac{C_t^{1-\gamma}}{1-\gamma} \quad \Rightarrow \quad M_{t+1} \approx \underbrace{\beta}_{a_t = a_0} - \underbrace{\beta\gamma}_{b_t = b_0} \Delta c_{t+1}$$

- Model with habit and time-varying risk aversion: Campbell and Cochrane '99, Menzly et. al '04

$$u(C_t, S_t) = \frac{(C_t S_t)^{1-\gamma}}{1-\gamma}, \qquad S_{t+1} \equiv \frac{C_t - X_t}{C_t}$$

$$\Rightarrow M_{t+1} \approx \beta\Big(\underbrace{1 - \gamma(\phi - 1)(s_t - \bar{s})}_{a_t} - \underbrace{\gamma(1 + \psi(s_t))}_{b_t} \Delta c_{t+1}\Big)$$

- Proxies for time-varying risk-premia should be good proxies for time-variation in a_t and b_t.

Scaled Consumption-Based Models

- $M_{t+1} \approx a_t + b_t \Delta c_{t+1}$
- Empirical specification: Lettau and Ludvigson (2001a, 2001b):
  - $a_t = a_0 + a_1 z_t, \quad b_t = b_0 + b_1 z_t$
  - $z_t = cay_t \equiv c_t - \alpha_a a_t - \alpha_y y_t$, (cointegrating residual)
  - cay_t related to log consumption-(aggregate) wealth ratio.
  - cay_t strong predictor of excess stock market returns

Scaled Consumption-Based Models

- $M_{t+1} \approx a_t + b_t \Delta c_{t+1}$
- Empirical specification: Lettau and Ludvigson (2001a, 2001b):
  - $a_t = a_0 + a_1 z_t, \quad b_t = b_0 + b_1 z_t$
  - $z_t = cay_t \equiv c_t - \alpha_a a_t - \alpha_y y_t$, (cointegrating residual)
  - cay_t related to log consumption-(aggregate) wealth ratio.
  - cay_t strong predictor of excess stock market returns
- Other examples: including housing consumption

$$U(C_t, H_t) = \frac{\widetilde{C}_t^{1-\frac{1}{\sigma}}}{1-\frac{1}{\sigma}} \qquad \widetilde{C}_t = \left[\chi C_t^{\frac{\varepsilon-1}{\varepsilon}} + (1-\chi) H_t^{\frac{\varepsilon-1}{\varepsilon}}\right]^{\frac{\varepsilon}{\varepsilon-1}},$$

$$\Rightarrow \ln M_{t+1} \approx a_t + b_t \Delta \ln C_{t+1} + d_t \Delta \ln S_{t+1}, \qquad S_{t+1} \equiv \frac{p_t^C C_t}{p_t^C C_t + p_t^H H_t}$$

- Lustig and Van Nieuwerburgh '05 (incomplete markets):
  - $a_t = a_0 + a_1 z_t, \quad b_t = b_0 + b_1 z_t, \quad d_t = d_0 + d_1 z_t$
  - $z_t =$ housing collateral ratio (measures quantity of risk sharing)

- Two kinds of conditioning are often confused.

- Two kinds of conditioning are often confused.
- Euler equation: $E\left[M_{t+1}R_{t+1}|\mathbf{z}_t\right] = 1$
- Unconditional version: $E\left[M_{t+1}R_{t+1}\right] = 1$

- Two kinds of conditioning are often confused.
- Euler equation: $E\left[M_{t+1}R_{t+1}|\mathbf{z}_t\right] = 1$
- Unconditional version: $E\left[M_{t+1}R_{t+1}\right] = 1$
- Two forms of conditionality:
  1. scaling *returns*: $E\left[M_{t+1}\left(R_{i,t+1} \otimes (1\ \mathbf{z}_t)'\right)\right] = 1$
  2. scaling *factors* $\mathbf{f}_{t+1}$, e.g., $\mathbf{f}_{t+1} = \Delta \ln C_{t+1}$:

$$\begin{aligned} M_{t+1} &= \mathbf{b}_t' \mathbf{f}_{t+1} \text{ with } \mathbf{b}_t = \mathbf{b}_0 + \mathbf{b}_1 \mathbf{z}_t \\ &= \mathbf{b}'\left(\mathbf{f}_{t+1} \otimes (1\ \mathbf{z}_t)'\right) \end{aligned}$$

Scaled Consumption-Based Models

Distinguishing two types of conditioning, or state dependence

- Two kinds of conditioning are often confused.
- Euler equation: $E\left[M_{t+1}R_{t+1}|\mathbf{z}_t\right] = 1$
- Unconditional version: $E\left[M_{t+1}R_{t+1}\right] = 1$
- Two forms of conditionality:
  1. scaling *returns*: $E\left[M_{t+1}\left(R_{i,t+1} \otimes (1\ \mathbf{z}_t)'\right)\right] = 1$
  2. scaling *factors* $\mathbf{f}_{t+1}$, e.g., $\mathbf{f}_{t+1} = \Delta \ln C_{t+1}$:

$$\begin{aligned} M_{t+1} &= \mathbf{b}_t' \mathbf{f}_{t+1} \text{ with } \mathbf{b}_t = \mathbf{b}_0 + \mathbf{b}_1 \mathbf{z}_t \\ &= \mathbf{b}' \left(\mathbf{f}_{t+1} \otimes (1\ \mathbf{z}_t)'\right) \end{aligned}$$

- Scaled consumption-based models are *conditional* in sense that M_{t+1} is a *state-dependent* function of $\Delta \ln C_{t+1}$
  - $\Rightarrow$ scaled *factors*

- Two kinds of conditioning are often confused.
- Euler equation: $E\left[M_{t+1}R_{t+1}|\mathbf{z}_t\right] = 1$
- Unconditional version: $E\left[M_{t+1}R_{t+1}\right] = 1$
- Two forms of conditionality:
  1. scaling *returns*: $E\left[M_{t+1}\left(R_{i,t+1} \otimes (1\ \mathbf{z}_t)'\right)\right] = 1$
  2. scaling *factors* $\mathbf{f}_{t+1}$, e.g., $\mathbf{f}_{t+1} = \Delta \ln C_{t+1}$:

$$\begin{aligned} M_{t+1} &= \mathbf{b}_t' \mathbf{f}_{t+1} \text{ with } \mathbf{b}_t = \mathbf{b}_0 + \mathbf{b}_1 \mathbf{z}_t \\ &= \mathbf{b}' \left(\mathbf{f}_{t+1} \otimes (1\ \mathbf{z}_t)'\right) \end{aligned}$$

- Scaled consumption-based models are *conditional* in sense that M_{t+1} is a *state-dependent* function of $\Delta \ln C_{t+1}$
  - $\Rightarrow$ scaled *factors*
- Scaled consumption-based models have been *tested* on unconditional moments, $E\left[M_{t+1}R_{t+1}\right] = 1$
  - $\Rightarrow$ NO scaled *returns*.

- Scaled CCAPM turns a *single* factor model with *state-dependent* weights into *multi-factor* model $\mathbf{f_t}$ with *constant* weights:

$$\begin{aligned} M_{t+1} &= (a_0 + a_1 z_t) + (b_0 + b_1 z_t)\,\Delta \ln C_{t+1} \\ &= a_0 + a_1 \underbrace{z_t}_{f_{1,t+1}} + b_0 \underbrace{\Delta \ln C_{t+1}}_{f_{2,t+1}} + b_1 (\underbrace{z_t \Delta \ln C_{t+1}}_{f_{3,t+1}}) \end{aligned}$$

- Scaled CCAPM turns a *single* factor model with *state-dependent* weights into *multi-factor* model $\mathbf{f_t}$ with *constant* weights:

$$
\begin{aligned}
M_{t+1} &= (a_0 + a_1 z_t) + (b_0 + b_1 z_t)\,\Delta \ln C_{t+1} \\
&= a_0 + a_1 \underbrace{z_t}_{f_{1,t+1}} + b_0 \underbrace{\Delta \ln C_{t+1}}_{f_{2,t+1}} + b_1 (\underbrace{z_t \Delta \ln C_{t+1}}_{f_{3,t+1}})
\end{aligned}
$$

- Multiple risk factors $\mathbf{f_t'} \equiv (z_t, \Delta \ln C_{t+1}, z_t \Delta \ln C_{t+1})$.

- Scaled CCAPM turns a *single* factor model with *state-dependent* weights into *multi-factor* model $\mathbf{f_t}$ with *constant* weights:

$$
\begin{aligned}
M_{t+1} &= (a_0 + a_1 z_t) + (b_0 + b_1 z_t)\,\Delta \ln C_{t+1} \\
&= a_0 + a_1 \underbrace{z_t}_{f_{1,t+1}} + b_0 \underbrace{\Delta \ln C_{t+1}}_{f_{2,t+1}} + b_1 (\underbrace{z_t \Delta \ln C_{t+1}}_{f_{3,t+1}})
\end{aligned}
$$

- Multiple risk factors $\mathbf{f_t'} \equiv (z_t, \Delta \ln C_{t+1}, z_t \Delta \ln C_{t+1})$.

- Scaled consumption models have *multiple, constant* betas for each factor, rather than a single time-varying beta for $\Delta \ln C_{t+1}$.

Deriving the "beta"-representation

Let $\mathbf{F} = (1 \;\; \mathbf{f}')'$, $M = \mathbf{b}'\mathbf{F}$, ignore time indices

$$\begin{aligned} 1 &= E[MR^i] \\ &= E[R^i\mathbf{F}']\mathbf{b} \quad \Rightarrow \quad \text{unconditional moments} \\ &= E[R^i]E[\mathbf{F}']\mathbf{b} + Cov(R^i, \mathbf{F}')\mathbf{b} \Rightarrow \end{aligned}$$

$$\begin{aligned} E[R^i] &= \frac{1 - Cov(R^i, \mathbf{F}')\mathbf{b}}{E[\mathbf{F}']\mathbf{b}} \\ &= \frac{1 - Cov(R^i, \mathbf{f}')\overline{\mathbf{b}}}{E[\mathbf{F}']\mathbf{b}} \\ &= \frac{1 - Cov(R^i, \mathbf{f}')Cov(\mathbf{f}, \mathbf{f}')^{-1}Cov(\mathbf{f}, \mathbf{f}')\overline{\mathbf{b}}}{E[\mathbf{F}']\mathbf{b}} \\ &= R^0 - R^0\boldsymbol{\beta}'Cov(\mathbf{f}, \mathbf{f}')\overline{\mathbf{b}} \\ &= R^0 - \boldsymbol{\beta}'\boldsymbol{\lambda} \quad \Rightarrow \quad \text{multiple, constant betas} \end{aligned}$$

- Estimate cross-sectional model using Fama-MacBeth (see Brandt lecture).

Fama-MacBeth Methodology: Preview–See Brandt

Step 1: Estimate β's in time-series regression for each portfolio i:

$$\boldsymbol{\beta}_i \equiv Cov(\mathbf{f}_{t+1}, \mathbf{f}'_{t+1})^{-1} Cov(\mathbf{f}_{t+1}, R_{i,t+1})$$

Step 2: Cross-sectional regressions (T of them):

$$R_{i,t+1} - R_{0,t} = \alpha_{i,t} + \boldsymbol{\beta}'_i \boldsymbol{\lambda}_t$$

$$\boldsymbol{\lambda} = 1/T \sum_{t=1}^{T} \boldsymbol{\lambda}_t; \qquad \sigma^2(\boldsymbol{\lambda}) = 1/T \sum_{t=1}^{T} (\boldsymbol{\lambda}_t - \boldsymbol{\lambda})^2$$

Note: report Shanken t-statistics (corrected for estimation error of betas in first stage)

Scaled Consumption-Based Models

Distinguishing two types of conditioning, or state dependence

- Scaled models: conditioning done in SDF:

$$M_{t+1} = a_t + b_t \Delta \ln C_{t+1},$$

not in Euler equation: $E(M\mathbf{R}) = \mathbf{1}_N$.

Scaled Consumption-Based Models

Distinguishing two types of conditioning, or state dependence

- Scaled models: conditioning done in SDF:

$$M_{t+1} = a_t + b_t \Delta \ln C_{t+1},$$

not in Euler equation: $E(M\mathbf{R}) = \mathbf{1}_N$.

- Gives rise to a *restricted* conditional consumption beta model:

$$R_t^i = a + \beta_{\Delta c} \Delta c_t + \beta_{\Delta c,z} \Delta c_t z_{t-1} + \beta_z z_{t-1}$$

- Scaled models: conditioning done in SDF:

$$M_{t+1} = a_t + b_t \Delta \ln C_{t+1},$$

not in Euler equation: $E(M\mathbf{R}) = \mathbf{1}_N$.

- Gives rise to a *restricted* conditional consumption beta model:

$$R_t^i = a + \beta_{\Delta c} \Delta c_t + \beta_{\Delta c,z} \Delta c_t z_{t-1} + \beta_z z_{t-1}$$

- Rewrite as

$$R_t^i = a + \underbrace{\left(\beta_c + \beta_{c,z} z_{t-1}\right)}_{\text{time-varying beta}} \Delta c_t + \beta_z z_{t-1}$$

- Scaled models: conditioning done in SDF:

$$M_{t+1} = a_t + b_t \Delta \ln C_{t+1},$$

not in Euler equation: $E(M\mathbf{R}) = \mathbf{1}_N$.

- Gives rise to a *restricted* conditional consumption beta model:

$$R_t^i = a + \beta_{\Delta c} \Delta c_t + \beta_{\Delta c,z} \Delta c_t z_{t-1} + \beta_z z_{t-1}$$

- Rewrite as

$$R_t^i = a + \underbrace{(\beta_c + \beta_{c,z} z_{t-1})}_{\text{time-varying beta}} \Delta c_t + \beta_z z_{t-1}$$

- Unlikely the same time-varying beta as obtained from modeling conditional mean $E_t(M_{t+1}\mathbf{R}_{t+1}) = 1$.

- Distinction is important.

- Distinction is important.
- Conditioning in SDF: theory provides guidance:
  - typically a few variables that capture risk-premia.

- Distinction is important.
- Conditioning in SDF: theory provides guidance:
  - typically a few variables that capture risk-premia.
- Conditioning in Euler eqn: model joint dist. $(M_{t+1}\mathbf{R_{t+1}})$.

- Distinction is important.
- Conditioning in SDF: theory provides guidance:
    - typically a few variables that capture risk-premia.
- Conditioning in Euler eqn: model joint dist. $(M_{t+1}\mathbf{R_{t+1}})$.
- Latter may require variables beyond a few that capture risk-premia.

- Distinction is important.
- Conditioning in SDF: theory provides guidance:
  - typically a few variables that capture risk-premia.
- Conditioning in Euler eqn: model joint dist. $(M_{t+1}\mathbf{R_{t+1}})$.
- Latter may require variables beyond a few that capture risk-premia.
- Approximating condition mean well requires large number of instruments (misspecified information sets)
  - Results sensitive to chosen conditioning variables, may fail to span information sets of market participants.

- Distinction is important.
- Conditioning in SDF: theory provides guidance:
  - typically a few variables that capture risk-premia.
- Conditioning in Euler eqn: model joint dist. $(M_{t+1}\mathbf{R_{t+1}})$.
- Latter may require variables beyond a few that capture risk-premia.
- Approximating condition mean well requires large number of instruments (misspecified information sets)
  - Results sensitive to chosen conditioning variables, may fail to span information sets of market participants.
- Partial solution: summarize information in large number of time-series with few estimated dynamic factors (e.g., Ludvigson and Ng '07, '09).

- Bottom lines:

- Bottom lines:

  1. Conditional moments of $M_{t+1}\mathbf{R}_{t+1}$ difficult to model $\Rightarrow$ reason to focus on unconditional moments $E[M_{t+1}\mathbf{R}_{t+1}] = 1$.

- Bottom lines:

  1. Conditional moments of $M_{t+1}\mathbf{R}_{t+1}$ difficult to model $\Rightarrow$ reason to focus on unconditional moments $E[M_{t+1}\mathbf{R}_{t+1}] = 1$.

  2. Models are misspecified: interesting question is whether state-dependence of M_{t+1} on consumption growth $\Rightarrow$ less misspecification than standard, fixed-weight CCAPM.

- Bottom lines:

  1. Conditional moments of $M_{t+1}\mathbf{R}_{t+1}$ difficult to model $\Rightarrow$ reason to focus on unconditional moments $E[M_{t+1}\mathbf{R}_{t+1}] = 1$.
  2. Models are misspecified: interesting question is whether state-dependence of M_{t+1} on consumption growth $\Rightarrow$ less misspecification than standard, fixed-weight CCAPM.
  3. As before, can compare models on basis of HJ distances, using White "reality check" method to compare statistically.

Asset Pricing Models With Recursive Preferences

- Growing interest in asset pricing models with recursive preferences, e.g., Epstein and Zin '89, '91, Weil '89, (EZW).

Asset Pricing Models With Recursive Preferences

- Growing interest in asset pricing models with recursive preferences, e.g., Epstein and Zin '89, '91, Weil '89, (EZW).

- Two reasons recursive utility is of interest:

Asset Pricing Models With Recursive Preferences

- Growing interest in asset pricing models with recursive preferences, e.g., Epstein and Zin ’89, ’91, Weil ’89, (EZW).
- Two reasons recursive utility is of interest:
  - More flexibility as regards attitudes toward risk and intertemporal substitution.

Asset Pricing Models With Recursive Preferences

- Growing interest in asset pricing models with recursive preferences, e.g., Epstein and Zin '89, '91, Weil '89, (EZW).
- Two reasons recursive utility is of interest:
  - More flexibility as regards attitudes toward risk and intertemporal substitution.
  - Preferences deliver an added risk factor for explaining asset returns.

- Growing interest in asset pricing models with recursive preferences, e.g., Epstein and Zin '89, '91, Weil '89, (EZW).
- Two reasons recursive utility is of interest:
  - More flexibility as regards attitudes toward risk and intertemporal substitution.
  - Preferences deliver an added risk factor for explaining asset returns.
- But, only a small amount of econometric work on recursive preferences $\Rightarrow$ gap in the literature.

- Here discuss two examples of estimating EZW models:

- Here discuss two examples of estimating EZW models:

  1. For general stationary, consumption growth and cash flow dynamics: Chen, Favilukis, Ludvigson '07.

- Here discuss two examples of estimating EZW models:

  1. For general stationary, consumption growth and cash flow dynamics: Chen, Favilukis, Ludvigson '07.
  2. When *restricting* cash flow dynamics (e.g., "long-run risk"): Bansal, Gallant, Tauchen '07.

- Here discuss two examples of estimating EZW models:
    1. For general stationary, consumption growth and cash flow dynamics: Chen, Favilukis, Ludvigson '07.
    2. When *restricting* cash flow dynamics (e.g., "long-run risk"): Bansal, Gallant, Tauchen '07.
- In (1) DGP is left *unrestricted*, as is joint distribution of consumption and returns (distribution-free estimation procedure).

- Here discuss two examples of estimating EZW models:

  1. For general stationary, consumption growth and cash flow dynamics: Chen, Favilukis, Ludvigson '07.

  2. When *restricting* cash flow dynamics (e.g., "long-run risk"): Bansal, Gallant, Tauchen '07.

- In (1) DGP is left *unrestricted*, as is joint distribution of consumption and returns (distribution-free estimation procedure).

- In (2) DGP *and* distribution of shocks explicitly modeled.

- Recursive utility (Epstein, Zin ('89, '91) & Weil ('89)):

$$V_t = \left[(1-\beta) C_t^{1-\rho} + \beta \mathcal{R}_t \left(V_{t+1} \right)^{1-\rho} \right]^{\frac{1}{1-\rho}}$$

$$\mathcal{R}_t \left(V_{t+1} \right) = \left(E \left[V_{t+1}^{1-\theta} | \mathcal{F}_t \right] \right)^{\frac{1}{1-\theta}}$$

- Recursive utility (Epstein, Zin ('89, '91) & Weil ('89)):

$$V_t = \left[(1-\beta) C_t^{1-\rho} + \beta \mathcal{R}_t \left(V_{t+1} \right)^{1-\rho} \right]^{\frac{1}{1-\rho}}$$

$$\mathcal{R}_t \left(V_{t+1} \right) = \left(E \left[V_{t+1}^{1-\theta} | \mathcal{F}_t \right] \right)^{\frac{1}{1-\theta}}$$

- V_{t+1} is continuation value, θ is RRA, $1/\rho$ is EIS.

- Recursive utility (Epstein, Zin ('89, '91) & Weil ('89)):

$$V_t = \left[(1-\beta) C_t^{1-\rho} + \beta \mathcal{R}_t \left(V_{t+1} \right)^{1-\rho} \right]^{\frac{1}{1-\rho}}$$

$$\mathcal{R}_t \left(V_{t+1} \right) = \left(E \left[V_{t+1}^{1-\theta} | \mathcal{F}_t \right] \right)^{\frac{1}{1-\theta}}$$

- V_{t+1} is continuation value, θ is RRA, $1/\rho$ is EIS.
- Rescale utility function (Hansen, Heaton, Li '05):

$$\frac{V_t}{C_t} = \left[(1-\beta) + \beta \mathcal{R}_t \left(\frac{V_{t+1}}{C_{t+1}} \frac{C_{t+1}}{C_t} \right)^{1-\rho} \right]^{\frac{1}{1-\rho}}$$

- Recursive utility (Epstein, Zin ('89, '91) & Weil ('89)):

$$V_t = \left[(1-\beta) C_t^{1-\rho} + \beta \mathcal{R}_t \left(V_{t+1} \right)^{1-\rho} \right]^{\frac{1}{1-\rho}}$$

$$\mathcal{R}_t \left(V_{t+1} \right) = \left(E \left[V_{t+1}^{1-\theta} | \mathcal{F}_t \right] \right)^{\frac{1}{1-\theta}}$$

- V_{t+1} is continuation value, θ is RRA, $1/\rho$ is EIS.
- Rescale utility function (Hansen, Heaton, Li '05):

$$\frac{V_t}{C_t} = \left[(1-\beta) + \beta \mathcal{R}_t \left(\frac{V_{t+1}}{C_{t+1}} \frac{C_{t+1}}{C_t} \right)^{1-\rho} \right]^{\frac{1}{1-\rho}}$$

- Special case: $\rho = \theta \Rightarrow$ CRRA separable utility $V_t = \beta \frac{C_t^{1-\theta}}{1-\theta}$.

- The MRS is pricing kernel (SDF) with added risk factor:

$$M_{t+1} = \beta \left(\frac{C_{t+1}}{C_t} \right)^{-\rho} \left(\frac{\frac{V_{t+1}}{C_{t+1}} \frac{C_{t+1}}{C_t}}{\mathcal{R}_t \left(\frac{V_{t+1}}{C_{t+1}} \frac{C_{t+1}}{C_t} \right)} \right)^{\rho - \theta}$$

- The MRS is pricing kernel (SDF) with added risk factor:

$$M_{t+1} = \beta \left(\frac{C_{t+1}}{C_t} \right)^{-\rho} \left(\frac{\frac{V_{t+1}}{C_{t+1}} \frac{C_{t+1}}{C_t}}{\mathcal{R}_t \left(\frac{V_{t+1}}{C_{t+1}} \frac{C_{t+1}}{C_t} \right)} \right)^{\rho - \theta}$$

- **Difficulty**: MRS a function of V/C, unobservable, embeds $\mathcal{R}_t(\cdot)$.

- The MRS is pricing kernel (SDF) with added risk factor:

$$M_{t+1} = \beta \left(\frac{C_{t+1}}{C_t} \right)^{-\rho} \left(\frac{\frac{V_{t+1}}{C_{t+1}} \frac{C_{t+1}}{C_t}}{\mathcal{R}_t \left(\frac{V_{t+1}}{C_{t+1}} \frac{C_{t+1}}{C_t} \right)} \right)^{\rho - \theta}$$

- **Difficulty**: MRS a function of V/C, unobservable, embeds $\mathcal{R}_t(\cdot)$.
- Epstein-Zin '91 use alt. rep. of SDF, uses agg. wealth return $R_{w,t}$:

$$M_{t+1} = \left\{ \beta \left(\frac{C_{t+1}}{C_t} \right)^{-\rho} \right\}^{\frac{1-\theta}{1-\rho}} \left\{ \frac{1}{R_{w,t+1}} \right\}^{\frac{\theta-\rho}{1-\rho}}$$

where $R_{w,t}$ *proxied* by stock market return.

- The MRS is pricing kernel (SDF) with added risk factor:

$$M_{t+1} = \beta \left(\frac{C_{t+1}}{C_t} \right)^{-\rho} \left(\frac{\frac{V_{t+1}}{C_{t+1}} \frac{C_{t+1}}{C_t}}{\mathcal{R}_t \left(\frac{V_{t+1}}{C_{t+1}} \frac{C_{t+1}}{C_t} \right)} \right)^{\rho - \theta}$$

- **Difficulty**: MRS a function of V/C, unobservable, embeds $\mathcal{R}_t(\cdot)$.
- Epstein-Zin '91 use alt. rep. of SDF, uses agg. wealth return $R_{w,t}$:

$$M_{t+1} = \left\{ \beta \left(\frac{C_{t+1}}{C_t} \right)^{-\rho} \right\}^{\frac{1-\theta}{1-\rho}} \left\{ \frac{1}{R_{w,t+1}} \right\}^{\frac{\theta-\rho}{1-\rho}}$$

  where $R_{w,t}$ *proxied* by stock market return.
- Problem: $R_{w,t+1}$ represents a claim to future C_t, itself unobservable.

1. If **EIS=1**, and $\Delta \log C_{t+1}$ follows a loglinear time-series process, $\log(V/C)$ has an analytical solution.

1. If **EIS=1**, and $\Delta \log C_{t+1}$ follows a loglinear time-series process, $\log(V/C)$ has an analytical solution.
2. If returns, C_t are jointly lognormal and homoscedastic, risk premia are **approx. log-linear** functions of COV between returns, and news about current and future C_t growth.

1. If **EIS=1**, and $\Delta \log C_{t+1}$ follows a loglinear time-series process, $\log(V/C)$ has an analytical solution.
2. If returns, C_t are jointly lognormal and homoscedastic, risk premia are **approx. log-linear** functions of COV between returns, and news about current and future C_t growth.

- But....

1. If **EIS=1**, and $\Delta \log C_{t+1}$ follows a loglinear time-series process, $\log(V/C)$ has an analytical solution.
2. If returns, C_t are jointly lognormal and homoscedastic, risk premia are **approx. log-linear** functions of COV between returns, and news about current and future C_t growth.

- But....
  - EIS=1 $\Rightarrow$ consumption-wealth ratio is constant, contradicting statistical evidence.

1. If **EIS=1**, and $\Delta \log C_{t+1}$ follows a loglinear time-series process, $\log(V/C)$ has an analytical solution.
2. If returns, C_t are jointly lognormal and homoscedastic, risk premia are **approx. log-linear** functions of COV between returns, and news about current and future C_t growth.

- But....
  - EIS=1 $\Rightarrow$ consumption-wealth ratio is constant, contradicting statistical evidence.
  - Joint lognormality strongly rejected in quarterly data.

1. If **EIS=1**, and $\Delta \log C_{t+1}$ follows a loglinear time-series process, $\log(V/C)$ has an analytical solution.
2. If returns, C_t are jointly lognormal and homoscedastic, risk premia are **approx. log-linear** functions of COV between returns, and news about current and future C_t growth.

- But....
  - EIS=1 $\Rightarrow$ consumption-wealth ratio is constant, contradicting statistical evidence.
  - Joint lognormality strongly rejected in quarterly data.
- Points to need for estimation method feasible under less restrictive assumptions.

- CFL: semiparametric approach to estimate EZW model *without*:

- CFL: semiparametric approach to estimate EZW model *without*:
  - Need to proxy $R_{w,t+1}$ with observable returns.

- CFL: semiparametric approach to estimate EZW model *without*:
  - Need to proxy $R_{w,t+1}$ with observable returns.
  - Loglinearizing the model.

- CFL: semiparametric approach to estimate EZW model *without*:
  - Need to proxy $R_{w,t+1}$ with observable returns.
  - Loglinearizing the model.
  - Parametric restrictions on law of motion *or* joint dist. of C_t and $R_{i,t}$, *or* on value of key preference parameters.
- Obtain estimates of β, RRA θ, EIS ρ^{-1}

- CFL: semiparametric approach to estimate EZW model *without*:
  - Need to proxy $R_{w,t+1}$ with observable returns.
  - Loglinearizing the model.
  - Parametric restrictions on law of motion *or* joint dist. of C_t and $R_{i,t}$, *or* on value of key preference parameters.
- Obtain estimates of β, RRA θ, EIS ρ^{-1}
- Evaluate EZW model's ability to fit asset return data relative to competing model specifications.

- CFL: semiparametric approach to estimate EZW model *without*:
  - Need to proxy $R_{w,t+1}$ with observable returns.
  - Loglinearizing the model.
  - Parametric restrictions on law of motion *or* joint dist. of C_t and $R_{i,t}$, *or* on value of key preference parameters.
- Obtain estimates of β, RRA θ, EIS ρ^{-1}
- Evaluate EZW model's ability to fit asset return data relative to competing model specifications.
- Investigate implications for $R_{w,t+1}$ and return to human wealth.

- CFL: semiparametric approach to estimate EZW model *without*:
  - Need to proxy $R_{w,t+1}$ with observable returns.
  - Loglinearizing the model.
  - Parametric restrictions on law of motion *or* joint dist. of C_t and $R_{i,t}$, *or* on value of key preference parameters.
- Obtain estimates of β, RRA θ, EIS ρ^{-1}
- Evaluate EZW model's ability to fit asset return data relative to competing model specifications.
- Investigate implications for $R_{w,t+1}$ and return to human wealth.
- Semiparametric approach is **sieve minimum distance** (SMD) procedure.

- First order conditions for optimal consumption choice:

$$E_t\left[\beta\left(\frac{C_{t+1}}{C_t}\right)^{-\rho}\left(\frac{\frac{V_{t+1}}{C_{t+1}}\frac{C_{t+1}}{C_t}}{\mathcal{R}_t\left(\frac{V_{t+1}}{C_{t+1}}\frac{C_{t+1}}{C_t}\right)}\right)^{\rho-\theta}R_{i,t+1}-1\right]=0 \tag{8}$$

- First order conditions for optimal consumption choice:

$$E_t\left[\beta\left(\frac{C_{t+1}}{C_t}\right)^{-\rho}\left(\frac{\frac{V_{t+1}}{C_{t+1}}\frac{C_{t+1}}{C_t}}{\mathcal{R}_t\left(\frac{V_{t+1}}{C_{t+1}}\frac{C_{t+1}}{C_t}\right)}\right)^{\rho-\theta}R_{i,t+1}-1\right]=0 \qquad (8)$$

- **CFL**: plug $\frac{V_t}{C_t}=\left[(1-\beta)+\beta\mathcal{R}_t\left(\frac{V_{t+1}}{C_{t+1}}\frac{C_{t+1}}{C_t}\right)^{1-\rho}\right]^{\frac{1}{1-\rho}}$ into (8):

$$E_t\left[\beta\left(\frac{C_{t+1}}{C_t}\right)^{-\rho}\left(\frac{\frac{V_{t+1}}{C_{t+1}}\frac{C_{t+1}}{C_t}}{\left\{\frac{1}{\beta}\left[\frac{V_t}{C_t}^{1-\rho}-(1-\beta)\right]\right\}^{\frac{1}{1-\rho}}}\right)^{\rho-\theta}R_{i,t+1}-1\right]=0 \qquad i=1,\ldots,N. \qquad (9)$$

- First order conditions for optimal consumption choice:

$$E_t\left[\beta\left(\frac{C_{t+1}}{C_t}\right)^{-\rho}\left(\frac{\frac{V_{t+1}}{C_{t+1}}\frac{C_{t+1}}{C_t}}{\mathcal{R}_t\left(\frac{V_{t+1}}{C_{t+1}}\frac{C_{t+1}}{C_t}\right)}\right)^{\rho-\theta}R_{i,t+1}-1\right]=0 \qquad (8)$$

- **CFL**: plug $\frac{V_t}{C_t}=\left[(1-\beta)+\beta\mathcal{R}_t\left(\frac{V_{t+1}}{C_{t+1}}\frac{C_{t+1}}{C_t}\right)^{1-\rho}\right]^{\frac{1}{1-\rho}}$ into (8):

$$E_t\left[\beta\left(\frac{C_{t+1}}{C_t}\right)^{-\rho}\left(\frac{\frac{V_{t+1}}{C_{t+1}}\frac{C_{t+1}}{C_t}}{\left\{\frac{1}{\beta}\left[\frac{V_t}{C_t}^{1-\rho}-(1-\beta)\right]\right\}^{\frac{1}{1-\rho}}}\right)^{\rho-\theta}R_{i,t+1}-1\right]=0 \qquad i=1,...,N. \qquad (9)$$

- N test asset returns, $\{R_{i,t+1}\}_{i=1}^{N}$. (9) is a x-sect asset pricing model.

EZW Recursive Preferences

Unrestricted Dynamics, Distribution-Free Estimation: Chen, Favilukis, Ludvigson '07

- First order conditions for optimal consumption choice:

$$E_t\left[\beta\left(\frac{C_{t+1}}{C_t}\right)^{-\rho}\left(\frac{\frac{V_{t+1}}{C_{t+1}}\frac{C_{t+1}}{C_t}}{\mathcal{R}_t\left(\frac{V_{t+1}}{C_{t+1}}\frac{C_{t+1}}{C_t}\right)}\right)^{\rho-\theta}R_{i,t+1}-1\right]=0 \tag{8}$$

- **CFL**: plug $\frac{V_t}{C_t}=\left[(1-\beta)+\beta\mathcal{R}_t\left(\frac{V_{t+1}}{C_{t+1}}\frac{C_{t+1}}{C_t}\right)^{1-\rho}\right]^{\frac{1}{1-\rho}}$ into (8):

$$E_t\left[\beta\left(\frac{C_{t+1}}{C_t}\right)^{-\rho}\left(\frac{\frac{V_{t+1}}{C_{t+1}}\frac{C_{t+1}}{C_t}}{\left\{\frac{1}{\beta}\left[\frac{V_t}{C_t}^{1-\rho}-(1-\beta)\right]\right\}^{\frac{1}{1-\rho}}}\right)^{\rho-\theta}R_{i,t+1}-1\right]=0 \qquad i=1,...,N. \tag{9}$$

- N test asset returns, $\{R_{i,t+1}\}_{i=1}^{N}$. (9) is a x-sect asset pricing model.
- Moment restrictions (9) form the basis of empirical investigation.

EZW Recursive Preferences

Unrestricted Dynamics, Distribution-Free Estimation: Chen, Favilukis, Ludvigson '07

- First order conditions for optimal consumption choice:

$$E_t \left[\beta \left(\frac{C_{t+1}}{C_t} \right)^{-\rho} \left(\frac{\frac{V_{t+1}}{C_{t+1}} \frac{C_{t+1}}{C_t}}{\mathcal{R}_t \left(\frac{V_{t+1}}{C_{t+1}} \frac{C_{t+1}}{C_t} \right)} \right)^{\rho-\theta} R_{i,t+1} - 1 \right] = 0 \qquad (8)$$

- **CFL**: plug $\frac{V_t}{C_t} = \left[(1-\beta) + \beta \mathcal{R}_t \left(\frac{V_{t+1}}{C_{t+1}} \frac{C_{t+1}}{C_t} \right)^{1-\rho} \right]^{\frac{1}{1-\rho}}$ into (8):

$$E_t \left[\beta \left(\frac{C_{t+1}}{C_t} \right)^{-\rho} \left(\frac{\frac{V_{t+1}}{C_{t+1}} \frac{C_{t+1}}{C_t}}{\left\{ \frac{1}{\beta} \left[\frac{V_t}{C_t}^{1-\rho} - (1-\beta) \right] \right\}^{\frac{1}{1-\rho}}} \right)^{\rho-\theta} R_{i,t+1} - 1 \right] = 0 \qquad i = 1, ..., N. \qquad (9)$$

- N test asset returns, $\{R_{i,t+1}\}_{i=1}^{N}$. (9) is a x-sect asset pricing model.
- Moment restrictions (9) form the basis of empirical investigation.
- Empirical model is *semiparametric*: $\delta \equiv (\beta, \theta, \rho)'$ denote finite dimensional parameter vector; V_t/C_t unknown function.

- Assume $\frac{V_t}{C_t}$ an unknown function F: $\mathbb{R}^2 \to \mathbb{R}$ of form

$$\frac{V_t}{C_t} = F\left(\frac{V_{t-1}}{C_{t-1}}, \frac{C_t}{C_{t-1}}\right),$$

- Assume $\frac{V_t}{C_t}$ an unknown function F: $\mathbb{R}^2 \to \mathbb{R}$ of form

$$\frac{V_t}{C_t} = F\left(\frac{V_{t-1}}{C_{t-1}}, \frac{C_t}{C_{t-1}}\right),$$

- Assume $\{C_t/C_{t-1} : t = 1, ...\}$ is strictly stationary ergodic; and $F(\cdot)$ is such that the process$\{V_t/C_t : t = 1, ...\}$ is asymptotically stationary ergodic.

- Assume $\frac{V_t}{C_t}$ an unknown function F: $\mathbb{R}^2 \rightarrow \mathbb{R}$ of form

$$\frac{V_t}{C_t} = F\left(\frac{V_{t-1}}{C_{t-1}}, \frac{C_t}{C_{t-1}}\right),$$

- Assume $\{C_t/C_{t-1} : t = 1, ...\}$ is strictly stationary ergodic; and $F(\cdot)$ is such that the process$\{V_t/C_t : t = 1, ...\}$ is asymptotically stationary ergodic.
- Justified if, for example,

- Assume $\frac{V_t}{C_t}$ an unknown function F: $\mathbb{R}^2 \to \mathbb{R}$ of form

$$\frac{V_t}{C_t} = F\left(\frac{V_{t-1}}{C_{t-1}}, \frac{C_t}{C_{t-1}}\right),$$

- Assume $\{C_t/C_{t-1} : t = 1, ...\}$ is strictly stationary ergodic; and $F(\cdot)$ is such that the process$\{V_t/C_t : t = 1, ...\}$ is asymptotically stationary ergodic.
- Justified if, for example,
  - $\Delta \log(C_{t+1})$ is (possibly nonlinear) function of a hidden first-order Markov process x_t.
  - Under general assumptions, information in x_t is summarized by V_{t-1}/C_{t-1} and C_t/C_{t-1}.

- Assume $\frac{V_t}{C_t}$ an unknown function F: $\mathbb{R}^2 \rightarrow \mathbb{R}$ of form

$$\frac{V_t}{C_t} = F\left(\frac{V_{t-1}}{C_{t-1}}, \frac{C_t}{C_{t-1}}\right),$$

- Assume $\{C_t/C_{t-1} : t = 1, ...\}$ is strictly stationary ergodic; and $F(\cdot)$ is such that the process$\{V_t/C_t : t = 1, ...\}$ is asymptotically stationary ergodic.
- Justified if, for example,
  - $\Delta \log(C_{t+1})$ is (possibly nonlinear) function of a hidden first-order Markov process x_t.
  - Under general assumptions, information in x_t is summarized by V_{t-1}/C_{t-1} and C_t/C_{t-1}.
- With a nonlinear Markov process for x_t, $F(\cdot)$ can display *nonmonotonicities* in both arguments.

- Assume $\frac{V_t}{C_t}$ an unknown function F: $\mathbb{R}^2 \to \mathbb{R}$ of form

$$\frac{V_t}{C_t} = F\left(\frac{V_{t-1}}{C_{t-1}}, \frac{C_t}{C_{t-1}}\right),$$

- Assume $\{C_t/C_{t-1} : t = 1, ...\}$ is strictly stationary ergodic; and $F(\cdot)$ is such that the process$\{V_t/C_t : t = 1, ...\}$ is asymptotically stationary ergodic.
- Justified if, for example,
    - $\Delta \log(C_{t+1})$ is (possibly nonlinear) function of a hidden first-order Markov process x_t.
    - Under general assumptions, information in x_t is summarized by V_{t-1}/C_{t-1} and C_t/C_{t-1}.
- **Note**: Markov assumption only a *motivation* for arguments of $F(\cdot)$. Econometric methodology itself leaves LOM for $\Delta \ln C_t$ unspecified.

- $\mathcal{F}_t$ denotes agents information set at time t.

- $\mathcal{F}_t$ denotes agents information set at time t.
- $\mathbf{z}_{t+1}$ contains all observations at $t+1$ and

$$\gamma_i(\mathbf{z}_{t+1}, \delta, F) \equiv \beta \left(\frac{C_{t+1}}{C_t} \right)^{-\rho} \left(\frac{F\left(\frac{V_t}{C_t}, \frac{C_{t+1}}{C_{t+1}} \right) \frac{C_{t+1}}{C_t}}{\left\{ \frac{1}{\beta} \left[\left\{ F\left(\frac{V_{t-1}}{C_{t-1}}, \frac{C_t}{C_{t-1}} \right) \right\}^{1-\rho} - (1-\beta) \right] \right\}^{\frac{1}{1-\rho}}} \right)^{\rho-\theta} R_{i,t+1} - 1$$

- $\mathcal{F}_t$ denotes agents information set at time t.
- $\mathbf{z}_{t+1}$ contains all observations at $t+1$ and

$$\gamma_i(\mathbf{z}_{t+1}, \delta, F) \equiv \beta \left(\frac{C_{t+1}}{C_t} \right)^{-\rho} \left(\frac{F\left(\frac{V_t}{C_t}, \frac{C_{t+1}}{C_{t+1}}\right) \frac{C_{t+1}}{C_t}}{\left\{ \frac{1}{\beta} \left[\left\{ F\left(\frac{V_{t-1}}{C_{t-1}}, \frac{C_t}{C_{t-1}} \right) \right\}^{1-\rho} - (1-\beta) \right] \right\}^{\frac{1}{1-\rho}}} \right)^{\rho-\theta} R_{i,t+1} - 1$$

- $\delta_o \equiv (\beta_o, \theta_o, \rho_o)'$, $F_o \equiv F_o(\mathbf{z}_t, \delta_o)$ denote true parameters that uniquely solve the conditional moment restrictions (Euler equations):

$$E\left\{ \gamma_i(\mathbf{z}_{t+1}, \delta_o, F_o(\cdot, \delta_o)) | \mathcal{F}_t \right\} = 0 \qquad i = 1, ..., N, \quad (10)$$

- $\mathcal{F}_t$ denotes agents information set at time t.
- $\mathbf{z}_{t+1}$ contains all observations at $t+1$ and

$$\gamma_i(\mathbf{z}_{t+1}, \delta, F) \equiv \beta \left(\frac{C_{t+1}}{C_t}\right)^{-\rho} \left(\frac{F\left(\frac{V_t}{C_t}, \frac{C_{t+1}}{C_{t+1}}\right) \frac{C_{t+1}}{C_t}}{\left\{\frac{1}{\beta}\left[\left\{F\left(\frac{V_{t-1}}{C_{t-1}}, \frac{C_t}{C_{t-1}}\right)\right\}^{1-\rho} - (1-\beta)\right]\right\}^{\frac{1}{1-\rho}}} \right)^{\rho-\theta} R_{i,t+1} - 1$$

- $\delta_o \equiv (\beta_o, \theta_o, \rho_o)'$, $F_o \equiv F_o(\mathbf{z}_t, \delta_o)$ denote true parameters that uniquely solve the conditional moment restrictions (Euler equations):

$$E\left\{\gamma_i(\mathbf{z}_{t+1}, \delta_o, F_o(\cdot, \delta_o)) | \mathcal{F}_t\right\} = 0 \qquad i = 1, ..., N, \quad (10)$$

- Let $\mathbf{w}_t \subseteq \mathcal{F}_t$. Equation (10) $\Rightarrow$

$$E\left\{\gamma_i(\mathbf{z}_{t+1}, \delta_o, F_o(\cdot, \delta_o)) | \mathbf{w}_t\right\} = 0. \qquad i = 1, ..., N.$$

- Intuition behind minimum distance procedure:

- Intuition behind minimum distance procedure:
- Theory $\Rightarrow$

$$m_t \equiv E\left\{\gamma_i(\mathbf{z}_{t+1}, \delta_o, F_o(\cdot, \delta_o)) | \mathbf{w}_t\right\} = 0. \qquad i = 1, ..., N. \tag{11}$$

- Intuition behind minimum distance procedure:
- Theory $\Rightarrow$

$$m_t \equiv E\left\{\gamma_i(\mathbf{z}_{t+1}, \delta_o, F_o\left(\cdot, \delta_o\right))|\mathbf{w}_t\right\} = 0. \qquad i = 1, ..., N. \tag{11}$$

- Since $m_t = 0$, m_t must have zero variance, mean.

- Intuition behind minimum distance procedure:
- Theory $\Rightarrow$

$$m_t \equiv E\left\{\gamma_i(\mathbf{z}_{t+1}, \delta_o, F_o(\cdot, \delta_o)) | \mathbf{w}_t\right\} = 0. \qquad i = 1, ..., N. \tag{11}$$

- Since $m_t = 0$, m_t must have zero variance, mean.
- Thus can find params by minimizing variance or quadratic norm: $\min E[(m_t)^2]$. Don't observe $m_t \Rightarrow$ need estimate $\widehat{m}_t$.

- Intuition behind minimum distance procedure:
- Theory $\Rightarrow$

$$m_t \equiv E\left\{\gamma_i(\mathbf{z}_{t+1}, \delta_o, F_o(\cdot, \delta_o)) | \mathbf{w}_t\right\} = 0. \qquad i = 1, ..., N. \tag{11}$$

- Since $m_t = 0$, m_t must have zero variance, mean.
- Thus can find params by minimizing variance or quadratic norm: $\min E[(m_t)^2]$. Don't observe $m_t \Rightarrow$ need estimate $\widehat{m}_t$.
- Since (11) is cond. mean, must hold for each observation, t.

- Intuition behind minimum distance procedure:
- Theory $\Rightarrow$

$$m_t \equiv E\left\{\gamma_i(\mathbf{z}_{t+1}, \delta_o, F_o(\cdot, \delta_o)) | \mathbf{w}_t\right\} = 0. \qquad i = 1, ..., N. \tag{11}$$

- Since $m_t = 0$, m_t must have zero variance, mean.
- Thus can find params by minimizing variance or quadratic norm: $\min E[(m_t)^2]$. Don't observe $m_t \Rightarrow$ need estimate $\widehat{m}_t$.
- Since (11) is cond. mean, must hold for each observation, t.

- Obs $>$ params, need way to weight each obs; using sample mean is one way: $\min \mathbf{E_T}[(m_t)^2]$.

- Intuition behind minimum distance procedure:
- Theory $\Rightarrow$

$$m_t \equiv E\left\{\gamma_i(\mathbf{z}_{t+1}, \delta_o, F_o(\cdot, \delta_o)) | \mathbf{w}_t\right\} = 0. \qquad i = 1, ..., N. \tag{11}$$

- Since $m_t = 0$, m_t must have zero variance, mean.
- Thus can find params by minimizing variance or quadratic norm: $\min E[(m_t)^2]$. Don't observe $m_t \Rightarrow$ need estimate $\widehat{m}_t$.
- Since (11) is cond. mean, must hold for each observation, t.

- Obs $>$ params, need way to weight each obs; using sample mean is one way: $\min \mathbf{E_T}[(m_t)^2]$.
- Minimum distance procedure useful for distribution-free estimation involving *conditional* moments.

- Minimum distance procedure useful for distribution-free estimation involving *conditional* moments: min $\mathbf{E_T}[(m_t)^2]$.

- Minimum distance procedure useful for distribution-free estimation involving *conditional* moments: min $\mathbf{E_T}[(m_t)^2]$.
- Contrast with GMM, used for *unconditional* moments: $E[f(x_t, \alpha)] = 0$.

- Minimum distance procedure useful for distribution-free estimation involving *conditional* moments: min $\mathbf{E_T}[(m_t)^2]$.
- Contrast with GMM, used for *unconditional* moments: $E[f(x_t, \alpha)] = 0$.
- With GMM take sample counterpart to population mean: $g_T = \sum_{t=1}^{T} f(x_t, \alpha) = 0$.

- Minimum distance procedure useful for distribution-free estimation involving *conditional* moments: min $\mathbf{E_T}[(m_t)^2]$.
- Contrast with GMM, used for *unconditional* moments: $E[f(x_t, \alpha)] = 0$.
- With GMM take sample counterpart to population mean: $g_T = \sum_{t=1}^{T} f(x_t, \alpha) = 0$.
- Then choose parameters α to min $g_T' W g_T$.

- Minimum distance procedure useful for distribution-free estimation involving *conditional* moments: min $\mathbf{E_T}[(m_t)^2]$.
- Contrast with GMM, used for *unconditional* moments: $E[f(x_t, \alpha)] = 0$.
- With GMM take sample counterpart to population mean: $g_T = \sum_{t=1}^{T} f(x_t, \alpha) = 0$.
- Then choose parameters α to $\min g_T' W g_T$.
- With GMM we average and then square.
- With SMD, we square and then average.

- True parameters δ_o and $F_o(\cdot, \delta_o)$ solve:

$$\min_{\delta \in \mathcal{D}} \inf_{F \in \mathcal{V}} E\left[m(\mathbf{w}_t, \delta, F)' m(\mathbf{w}_t, \delta, F)\right],$$

- where $m(\mathbf{w}_t, \delta, F) = E\{\gamma(\mathbf{z}_{t+1}, \delta, F) | \mathbf{w}_t\}$
- $\gamma(\mathbf{z}_{t+1}, \delta, F) = (\gamma_1(\mathbf{z}_{t+1}, \delta, F), ..., \gamma_N(\mathbf{z}_{t+1}, \delta, F))'$

- True parameters δ_o and $F_o(\cdot, \delta_o)$ solve:

$$\min_{\delta \in \mathcal{D}} \inf_{F \in \mathcal{V}} E\left[m(\mathbf{w}_t, \delta, F)' m(\mathbf{w}_t, \delta, F)\right],$$

- where $m(\mathbf{w}_t, \delta, F) = E\{\gamma(\mathbf{z}_{t+1}, \delta, F) | \mathbf{w}_t\}$
- $\gamma(\mathbf{z}_{t+1}, \delta, F) = (\gamma_1(\mathbf{z}_{t+1}, \delta, F), ..., \gamma_N(\mathbf{z}_{t+1}, \delta, F))'$
- For any candidate $\delta \equiv (\beta, \theta, \rho)' \in \mathcal{D}$, define $V^* \equiv F^*(\mathbf{z}_t, \delta) \equiv F^*(\cdot, \delta)$ as:

$$F^*(\cdot, \delta) = \arg \inf_{F \in \mathcal{V}} E\left[m(\mathbf{w}_t, \delta, F)' m(\mathbf{w}_t, \delta, F)\right]$$

- True parameters δ_o and $F_o(\cdot, \delta_o)$ solve:

$$\min_{\delta \in \mathcal{D}} \inf_{F \in \mathcal{V}} E\left[m(\mathbf{w}_t, \delta, F)' m(\mathbf{w}_t, \delta, F)\right],$$

- where $m(\mathbf{w}_t, \delta, F) = E\{\gamma(\mathbf{z}_{t+1}, \delta, F) | \mathbf{w}_t\}$
- $\gamma(\mathbf{z}_{t+1}, \delta, F) = (\gamma_1(\mathbf{z}_{t+1}, \delta, F), ..., \gamma_N(\mathbf{z}_{t+1}, \delta, F))'$
- For any candidate $\delta \equiv (\beta, \theta, \rho)' \in \mathcal{D}$, define $V^* \equiv F^*(\mathbf{z}_t, \delta) \equiv F^*(\cdot, \delta)$ as:

$$F^*(\cdot, \delta) = \arg\inf_{F \in \mathcal{V}} E\left[m(\mathbf{w}_t, \delta, F)' m(\mathbf{w}_t, \delta, F)\right]$$

- It is clear that $F_o(\mathbf{z}_t, \delta_o) = F^*(\mathbf{z}_t, \delta_o)$

- First step: For any candidate $\delta \in \mathcal{D}$, an initial estimate of $F^*(\cdot, \delta)$ obtained using SMD that consists of two parts: (Newey-Powell '03, Ai-Chen '03, Ai-Chen '07).

- First step: For any candidate $\delta \in \mathcal{D}$, an initial estimate of $F^*(\cdot, \delta)$ obtained using SMD that consists of two parts: (Newey-Powell '03, Ai-Chen '03, Ai-Chen '07).

  1. Replace the conditional expectation with a consistent, nonparametric estimator (specified later).

- First step: For any candidate $\delta \in \mathcal{D}$, an initial estimate of $F^*(\cdot, \delta)$ obtained using SMD that consists of two parts: (Newey-Powell '03, Ai-Chen '03, Ai-Chen '07).

  1. Replace the conditional expectation with a consistent, nonparametric estimator (specified later).
  2. Approximate the unknown function F by a sequence of finite dimensional unknown parameters (sieves) F_{K_T}.
     - Approximation error decreases as K_T increases with T.

EZW Recursive Preferences: Two-Step Procedure

Unrestricted Dynamics, Distribution-Free Estimation: Chen, Favilukis, Ludvigson '07

- First step: For any candidate $\delta \in \mathcal{D}$, an initial estimate of $F^*(\cdot, \delta)$ obtained using SMD that consists of two parts: (Newey-Powell '03, Ai-Chen '03, Ai-Chen '07).
  1. Replace the conditional expectation with a consistent, nonparametric estimator (specified later).
  2. Approximate the unknown function F by a sequence of finite dimensional unknown parameters (sieves) F_{K_T}.
     - Approximation error decreases as K_T increases with T.
- Second step: estimates of δ_o is obtained by solving a sample minimum distance problem such as GMM.

- Approximate $\frac{V_t}{C_t} = F\left(\frac{V_{t-1}}{C_{t-1}}, \frac{C_t}{C_{t-1}}; \delta\right)$ with a bivariate sieve:

$$F\left(\frac{V_{t-1}}{C_{t-1}}, \frac{C_t}{C_{t-1}}; \delta\right) \approx F_{K_T}(\cdot, \delta) = a_0(\delta) + \sum_{j=1}^{K_T} a_j(\delta) B_j\left(\frac{V_{t-1}}{C_{t-1}}, \frac{C_t}{C_{t-1}}\right)$$

- Sieve coefficients $\{a_0, a_1, ..., a_{K_T}\}$ depend on δ
- Basis functions $\{B_j(\cdot, \cdot) : j = 1, ..., K_T\}$ have known functional forms independent of δ
- Initial value for $\frac{V_t}{C_t}$ at time $t = 0$, denoted $\frac{V_0}{C_0}$, taken as a unknown scalar parameter to be estimated.
- Given $\frac{V_0}{C_0}$, $\{a_j\}_{j=1}^{K_T}$, $\{B_j\}_{j=1}^{K_T}$ and data on consumption $\left\{\frac{C_t}{C_{t-1}}\right\}_{t=1}^{T}$, use F_{K_T} to generate a sequence $\left\{\frac{V_i}{C_i}\right\}_{i=1}^{T}$.

- Recall $m(\mathbf{w}_t, \delta_o, F^*(\cdot, \delta_o)) \equiv E\left\{\gamma(\mathbf{z}_{t+1}, \delta_o, F^*\left(\cdot, \delta_o\right)) | \mathbf{w}_t\right\} = 0.$

- First-step SMD estimate $\widehat{F}\left(\cdot\right)$ for $F^*\left(\cdot\right)$ based on

$$\widehat{F}\left(\cdot, \delta\right) = \arg\min_{F_{K_T}} \frac{1}{T}\sum_{t=1}^{T} \widehat{m}(\mathbf{w}_t, \delta, F_{K_T}(\cdot, \delta))'\widehat{m}(\mathbf{w}_t, \delta, F_{K_T}(\cdot, \delta)),$$

- $\widehat{m}(\mathbf{w}_t, \delta, F_{K_T}(\cdot, \delta))$ any nonpara. estimator of m.

- Do this for a three dimensional grid of values of $\delta = (\beta, \theta, \rho)'$.

- Example of nonparametric estimator of m:
- Let $\{p_{0j}(\mathbf{w}_t),\ j = 1, 2, ..., J_T\}$, $\mathbb{R}^{d_w} \to \mathbb{R}$ be *instruments*. $p^{J_T}(\cdot) \equiv (p_{01}(\cdot), ..., p_{0J_T}(\cdot))'$
- Define $T \times J_T$ matrix $\mathbf{P} \equiv \left(p^{J_T}(w_1), ..., p^{J_T}(w_T)\right)'$. Then:

$$\widehat{m}(\mathbf{w}, \delta, F) = \left(\sum_{t=1}^{T} \gamma(\mathbf{z}_{t+1}, \delta, F) p^{J_T}(\mathbf{w}_t)' (\mathbf{P}'\mathbf{P})^{-1}\right) p^{J_T}(\mathbf{w})$$

- $\widehat{m}(\cdot)$ a sieve LS estimator of $m(\mathbf{w}, \delta, F)$.
- Procedure equivalent to regressing each γ_i on instruments and taking fitted values as estimate of conditional mean.

- $\widehat{m}(\cdot)$ a sieve LS estimator of $m(\mathbf{w}, \delta, F)$.

- $\widehat{m}(\cdot)$ a sieve LS estimator of $m(\mathbf{w}, \delta, F)$.
- Attractive feature of this estimator of F^*: implemented as GMM

$$\widehat{F}_T(\cdot, \delta) = \underset{F_T \in \mathcal{V}_T}{\arg\min} \left[\mathbf{g}_T(\delta, F_T; \mathbf{y}^T)\right]' \underbrace{\left\{\mathbf{I}_N \otimes \left(\mathbf{P}'\mathbf{P}\right)^{-1}\right\}}_{\mathbf{W}} \left[\mathbf{g}_T(\delta, F_T; \mathbf{y}^T)\right], \tag{12}$$

where $\mathbf{y}^T = \left(\mathbf{z}'_{T+1}, ...\mathbf{z}'_2, \mathbf{w}'_T, ...\mathbf{w}'_1\right)'$ denotes vector of all obs and

$$\mathbf{g}_T(\delta, F_T; \mathbf{y}^T) = \frac{1}{T} \sum_{t=1}^{T} \gamma(\mathbf{z}_{t+1}, \delta, F_T) \otimes p^{J_T}(\mathbf{w}_t) \tag{13}$$

- $\widehat{m}(\cdot)$ a sieve LS estimator of $m(\mathbf{w}, \delta, F)$.
- Attractive feature of this estimator of F^*: implemented as GMM

$$\widehat{F}_T(\cdot, \delta) = \underset{F_T \in \mathcal{V}_T}{\arg\min} \left[\mathbf{g}_T(\delta, F_T; \mathbf{y}^T)\right]' \underbrace{\left\{\mathbf{I}_N \otimes \left(\mathbf{P}'\mathbf{P}\right)^{-1}\right\}}_{\mathbf{W}} \left[\mathbf{g}_T(\delta, F_T; \mathbf{y}^T)\right], \tag{12}$$

where $\mathbf{y}^T = \left(\mathbf{z}'_{T+1}, ...\mathbf{z}'_2, \mathbf{w}'_T, ...\mathbf{w}'_1\right)'$ denotes vector of all obs and

$$\mathbf{g}_T(\delta, F_T; \mathbf{y}^T) = \frac{1}{T} \sum_{t=1}^{T} \gamma(\mathbf{z}_{t+1}, \delta, F_T) \otimes p^{J_T}(\mathbf{w}_t) \tag{13}$$

- Weighting gives greater weight to moments more highly correlated with instruments $p^{J_T}(\cdot)$.

- $\widehat{m}(\cdot)$ a sieve LS estimator of $m(\mathbf{w}, \delta, F)$.
- Attractive feature of this estimator of F^*: implemented as GMM

$$\widehat{F}_T(\cdot,\delta) = \arg\min_{F_T \in \mathcal{V}_T} \left[\mathbf{g}_T(\delta, F_T; \mathbf{y}^T)\right]' \underbrace{\left\{\mathbf{I}_N \otimes \left(\mathbf{P}'\mathbf{P}\right)^{-1}\right\}}_{\mathbf{W}} \left[\mathbf{g}_T(\delta, F_T; \mathbf{y}^T)\right], \tag{12}$$

where $\mathbf{y}^T = \left(\mathbf{z}'_{T+1}, ...\mathbf{z}'_2, \mathbf{w}'_T, ...\mathbf{w}'_1\right)'$ denotes vector of all obs and

$$\mathbf{g}_T(\delta, F_T; \mathbf{y}^T) = \frac{1}{T}\sum_{t=1}^{T} \gamma(\mathbf{z}_{t+1}, \delta, F_T) \otimes p^{J_T}(\mathbf{w}_t) \tag{13}$$

- Weighting gives greater weight to moments more highly correlated with instruments $p^{J_T}(\cdot)$.
- Weighting can be understood intuitively by noting that variation in conditional mean $m(\mathbf{w}_t, \delta, F)$ is what identifies $F^*(\cdot, \delta)$.

EZW Preferences: Second Step GMM Est of δ_o

Unrestricted Dynamics, Distribution-Free Estimation: Chen, Favilukis, Ludvigson '07

- Under correct specification, δ_o satisfies :

$$E\left\{\gamma_i(\mathbf{z}_{t+1}, \delta_o, F^*(\cdot, \delta_o)) \otimes \mathbf{x}_t\right\} = 0, \qquad i = 1, ..., N.$$

- Under correct specification, δ_o satisfies :

$$E\left\{\gamma_i(\mathbf{z}_{t+1},\delta_o,F^*\left(\cdot,\delta_o\right))\otimes\mathbf{x}_t\right\}=0,\qquad\qquad i=1,...,N.$$

- Sample moments:
$\mathbf{g}_T(\delta,\widehat{F}\left(\cdot,\delta\right);\mathbf{y}^T)\equiv\frac{1}{T}\sum_{t=1}^T\gamma(\mathbf{z}_{t+1},\delta,\widehat{F}\left(\cdot,\delta\right))\otimes\mathbf{x}_t.$

EZW Preferences: Second Step GMM Est of δ_o

Unrestricted Dynamics, Distribution-Free Estimation: Chen, Favilukis, Ludvigson '07

- Under correct specification, δ_o satisfies :

$$E\left\{\gamma_i(\mathbf{z}_{t+1}, \delta_o, F^*\left(\cdot, \delta_o\right)) \otimes \mathbf{x}_t\right\} = 0, \qquad i = 1, ..., N.$$

- Sample moments:
$\mathbf{g}_T(\delta, \widehat{F}\left(\cdot, \delta\right); \mathbf{y}^T) \equiv \frac{1}{T}\sum_{t=1}^T \gamma(\mathbf{z}_{t+1}, \delta, \widehat{F}\left(\cdot, \delta\right)) \otimes \mathbf{x}_t.$
- Regardless the model is correctly or incorrectly specified, estimate δ by minimizing GMM objective:

$$\widehat{\delta} = \arg\min_{\delta \in \mathcal{D}} \left[\mathbf{g}_T(\delta, \widehat{F}\left(\cdot, \delta\right); \mathbf{y}^T)\right]' \mathbf{W} \left[\mathbf{g}_T(\delta, \widehat{F}\left(\cdot, \delta\right); \mathbf{y}^T)\right]$$

EZW Preferences: Second Step GMM Est of δ_o

Unrestricted Dynamics, Distribution-Free Estimation: Chen, Favilukis, Ludvigson '07

- Under correct specification, δ_o satisfies :

$$E\left\{\gamma_i(\mathbf{z}_{t+1}, \delta_o, F^*\left(\cdot, \delta_o\right)) \otimes \mathbf{x}_t\right\} = 0, \qquad i = 1, ..., N.$$

- Sample moments:
$\mathbf{g}_T(\delta, \widehat{F}\left(\cdot, \delta\right); \mathbf{y}^T) \equiv \frac{1}{T}\sum_{t=1}^{T} \gamma(\mathbf{z}_{t+1}, \delta, \widehat{F}\left(\cdot, \delta\right)) \otimes \mathbf{x}_t.$
- Regardless the model is correctly or incorrectly specified, estimate δ by minimizing GMM objective:

$$\widehat{\delta} = \arg\min_{\delta \in \mathcal{D}} \left[\mathbf{g}_T(\delta, \widehat{F}\left(\cdot, \delta\right); \mathbf{y}^T)\right]' \mathbf{W} \left[\mathbf{g}_T(\delta, \widehat{F}\left(\cdot, \delta\right); \mathbf{y}^T)\right]$$

- Examples: $\mathbf{W} = \mathbf{I}$, $\mathbf{W} = \mathbf{G}_T^{-1}$.

EZW Preferences: Second Step GMM Est of δ_o

Unrestricted Dynamics, Distribution-Free Estimation: Chen, Favilukis, Ludvigson '07

- Under correct specification, δ_o satisfies :

$$E\left\{\gamma_i(\mathbf{z}_{t+1},\delta_o,F^*\left(\cdot,\delta_o\right))\otimes\mathbf{x}_t\right\}=0, \qquad i=1,...,N.$$

- Sample moments:
$\mathbf{g}_T(\delta,\widehat{F}\left(\cdot,\delta\right);\mathbf{y}^T)\equiv\frac{1}{T}\sum_{t=1}^T\gamma(\mathbf{z}_{t+1},\delta,\widehat{F}\left(\cdot,\delta\right))\otimes\mathbf{x}_t.$

- Regardless the model is correctly or incorrectly specified, estimate δ by minimizing GMM objective:

$$\widehat{\delta}=\arg\min_{\delta\in\mathcal{D}}\left[\mathbf{g}_T(\delta,\widehat{F}\left(\cdot,\delta\right);\mathbf{y}^T)\right]'\mathbf{W}\left[\mathbf{g}_T(\delta,\widehat{F}\left(\cdot,\delta\right);\mathbf{y}^T)\right]$$

- Examples: $\mathbf{W}=\mathbf{I}$, $\mathbf{W}=\mathbf{G}_T^{-1}$.

- $\widehat{F}\left(\cdot,\delta\right)$ not held fixed in this step: depends on δ!

- Under correct specification, δ_o satisfies :

$$E\left\{\gamma_i(\mathbf{z}_{t+1}, \delta_o, F^*(\cdot, \delta_o)) \otimes \mathbf{x}_t\right\} = 0, \qquad i = 1, ..., N.$$

- Sample moments:
  $\mathbf{g}_T(\delta, \widehat{F}(\cdot, \delta); \mathbf{y}^T) \equiv \frac{1}{T}\sum_{t=1}^{T} \gamma(\mathbf{z}_{t+1}, \delta, \widehat{F}(\cdot, \delta)) \otimes \mathbf{x}_t.$
- Regardless the model is correctly or incorrectly specified, estimate δ by minimizing GMM objective:

$$\widehat{\delta} = \arg\min_{\delta \in \mathcal{D}} \left[\mathbf{g}_T(\delta, \widehat{F}(\cdot, \delta); \mathbf{y}^T)\right]' \mathbf{W} \left[\mathbf{g}_T(\delta, \widehat{F}(\cdot, \delta); \mathbf{y}^T)\right]$$

- Examples: $\mathbf{W} = \mathbf{I}$, $\mathbf{W} = \mathbf{G}_T^{-1}$.
- $\widehat{F}(\cdot, \delta)$ not held fixed in this step: depends on δ!
- Estimator $\widehat{F}(\cdot, \delta)$ obtained using min. dist over a grid of values δ.

- Under correct specification, δ_o satisfies :

$$E\left\{\gamma_i(\mathbf{z}_{t+1}, \delta_o, F^*\left(\cdot, \delta_o\right)) \otimes \mathbf{x}_t\right\} = 0, \qquad i = 1, ..., N.$$

- Sample moments:
$\mathbf{g}_T(\delta, \widehat{F}\left(\cdot, \delta\right); \mathbf{y}^T) \equiv \frac{1}{T}\sum_{t=1}^{T} \gamma(\mathbf{z}_{t+1}, \delta, \widehat{F}\left(\cdot, \delta\right)) \otimes \mathbf{x}_t.$
- Regardless the model is correctly or incorrectly specified, estimate δ by minimizing GMM objective:

$$\widehat{\delta} = \arg\min_{\delta \in \mathcal{D}} \left[\mathbf{g}_T(\delta, \widehat{F}\left(\cdot, \delta\right); \mathbf{y}^T)\right]' \mathbf{W} \left[\mathbf{g}_T(\delta, \widehat{F}\left(\cdot, \delta\right); \mathbf{y}^T)\right]$$

- Examples: $\mathbf{W} = \mathbf{I}$, $\mathbf{W} = \mathbf{G}_T^{-1}$.
- $\widehat{F}\left(\cdot, \delta\right)$ not held fixed in this step: depends on δ!
- Estimator $\widehat{F}\left(\cdot, \delta\right)$ obtained using min. dist over a grid of values δ.
- Choose the δ and corresponding $\widehat{F}\left(\cdot, \delta\right)$ that minimizes GMM criterion.

- Why estimate in two steps? All params could be estimated in one step by minimizing the SMD criterion.

- Why estimate in two steps? All params could be estimated in one step by minimizing the SMD criterion.

- Less desirable for asset pricing:

- Why estimate in two steps? All params could be estimated in one step by minimizing the SMD criterion.

- Less desirable for asset pricing:

  1. Want estimates of RRA and EIS to reflect values required to match unconditional risk premia. Not possible using SMD which emphasizes *conditional* moments.

- Why estimate in two steps? All params could be estimated in one step by minimizing the SMD criterion.
- Less desirable for asset pricing:
    1. Want estimates of RRA and EIS to reflect values required to match unconditional risk premia. Not possible using SMD which emphasizes *conditional* moments.
    2. SMD procedure effectively changes set of test assets–linear combinations of original portfolio returns. But we may be interested in explaining original returns!

- Why estimate in two steps? All params could be estimated in one step by minimizing the SMD criterion.

- Less desirable for asset pricing:

  1. Want estimates of RRA and EIS to reflect values required to match unconditional risk premia. Not possible using SMD which emphasizes *conditional* moments.

  2. SMD procedure effectively changes set of test assets–linear combinations of original portfolio returns. But we may be interested in explaining original returns!

  3. Linear combinations may imply implausible long and short positions, do not necessarily deliver a large spread in unconditional mean returns.

- Procedure allows for model misspecification:

- Procedure allows for model misspecification:
  - Euler equation need not hold with equality.

- Procedure allows for model misspecification:
  - Euler equation need not hold with equality.
- As before, compare models by relative *magnitude* of misspecification, rather than...

EZW Recursive Preferences: Two Step Estimation

Unrestricted Dynamics, Distribution-Free Estimation: Chen, Favilukis, Ludvigson '07

- Procedure allows for model misspecification:
  - Euler equation need not hold with equality.
- As before, compare models by relative *magnitude* of misspecification, rather than...
- ...asking whether each model individually fits data perfectly (given sampling error).

EZW Recursive Preferences: Two Step Estimation

Unrestricted Dynamics, Distribution-Free Estimation: Chen, Favilukis, Ludvigson '07

- Procedure allows for model misspecification:
  - Euler equation need not hold with equality.
- As before, compare models by relative *magnitude* of misspecification, rather than...
- ...asking whether each model individually fits data perfectly (given sampling error).
  - Use $\mathbf{W} = \mathbf{G}^{-1}$ in second step, compute HJ distance.

- Procedure allows for model misspecification:
  - Euler equation need not hold with equality.
- As before, compare models by relative *magnitude* of misspecification, rather than...
- ...asking whether each model individually fits data perfectly (given sampling error).
  - Use $\mathbf{W} = \mathbf{G}^{-1}$ in second step, compute HJ distance.
- Test whether HJ distances of competing models are statistically different (White reality check–Chen and Ludvigson '09).

- Bansal, Gallant, Tauchen '07: SMM estimation of LRR model: Bansal & Yaron '04.

- Bansal, Gallant, Tauchen '07: SMM estimation of LRR model: Bansal & Yaron '04.
- Structural estimation of EZW utility, *restricting* to specific law of motion for cash flows ("long-run risk"LRR).

- Bansal, Gallant, Tauchen '07: SMM estimation of LRR model: Bansal & Yaron '04.
- Structural estimation of EZW utility, *restricting* to specific law of motion for cash flows ("long-run risk"LRR).
- Cash flow dynamics in BGT version of LRR model:

$$\begin{aligned}
\Delta c_{t+1} &= \mu_c + x_{c,t} + \sigma_t \varepsilon_{c,t+1} \\
\Delta d_{t+1} &= \mu_d + \phi_x \underbrace{x_{c,t}}_{\text{LR risk}} + \phi_s s_t + \sigma_{\varepsilon_d} \sigma_t \varepsilon_{d,t+1} \\
x_{c,t} &= \phi x_{c,t-1} + \sigma_{\varepsilon_x} \sigma \varepsilon_{xc,t} \\
\sigma_t^2 &= \sigma^2 + \nu(\sigma_{t-1}^2 - \sigma^2) + \sigma_w w_t
\end{aligned}$$

$$s_t = (\mu_d - \mu_c) + d_t - c_t$$

$$\varepsilon_{c,t+1}, \varepsilon_{d,t+1}, \varepsilon_{xc,t}, w_t \sim \mathbf{N.i.i.d}\,(\mathbf{0}, \mathbf{1})$$

- SMM methodology: Gallant & Tauchen '96; Gallant, Hsieh, Tauchen '97, Tauchen '97.

- SMM methodology: Gallant & Tauchen '96; Gallant, Hsieh, Tauchen '97, Tauchen '97.
- Outline of SMM steps

- SMM methodology: Gallant & Tauchen '96; Gallant, Hsieh, Tauchen '97, Tauchen '97.
- Outline of SMM steps
  1. Solve the model over grid of values of deep parameters: $\rho_d = (\beta, \theta, \rho, \phi, \phi_x, \mu_c, \mu_d, \sigma, \sigma_{\epsilon_d}, \sigma_{\epsilon_x} \nu, \phi_s, \sigma_w)'$

- SMM methodology: Gallant & Tauchen '96; Gallant, Hsieh, Tauchen '97, Tauchen '97.
- Outline of SMM steps
  1. Solve the model over grid of values of deep parameters: $\rho_d = (\beta, \theta, \rho, \phi, \phi_x, \mu_c, \mu_d, \sigma, \sigma_{\epsilon_d}, \sigma_{\epsilon_x} \nu, \phi_s, \sigma_w)'$
  2. For each value of ρ_d on the grid, combine solutions with long simulation of **length N** of model.

- SMM methodology: Gallant & Tauchen '96; Gallant, Hsieh, Tauchen '97, Tauchen '97.
- Outline of SMM steps
  1. Solve the model over grid of values of deep parameters: $\rho_d = (\beta, \theta, \rho, \phi, \phi_x, \mu_c, \mu_d, \sigma, \sigma_{\epsilon_d}, \sigma_{\epsilon_x}\nu, \phi_s, \sigma_w)'$
  2. For each value of ρ_d on the grid, combine solutions with long simulation of **length N** of model.
  3. Simulation: Monte Carlo draws from the Normal distribution for primitive shocks.

- SMM methodology: Gallant & Tauchen '96; Gallant, Hsieh, Tauchen '97, Tauchen '97.
- Outline of SMM steps
  1. Solve the model over grid of values of deep parameters: $\rho_d = (\beta, \theta, \rho, \phi, \phi_x, \mu_c, \mu_d, \sigma, \sigma_{\epsilon_d}, \sigma_{\epsilon_x} \nu, \phi_s, \sigma_w)'$
  2. For each value of ρ_d on the grid, combine solutions with long simulation of **length N** of model.
  3. Simulation: Monte Carlo draws from the Normal distribution for primitive shocks.
  4. Form obs eqn for simulated *and* historical data, e.g., $y_t = (d_t - c_t, c_t - c_{t-1}, p_t - d_t, r_{d,t})'$

- SMM methodology: Gallant & Tauchen '96; Gallant, Hsieh, Tauchen '97, Tauchen '97.
- Outline of SMM steps
  1. Solve the model over grid of values of deep parameters: $\rho_d = (\beta, \theta, \rho, \phi, \phi_x, \mu_c, \mu_d, \sigma, \sigma_{\epsilon_d}, \sigma_{\epsilon_x}\nu, \phi_s, \sigma_w)'$
  2. For each value of ρ_d on the grid, combine solutions with long simulation of **length N** of model.
  3. Simulation: Monte Carlo draws from the Normal distribution for primitive shocks.
  4. Form obs eqn for simulated *and* historical data, e.g., $y_t = (d_t - c_t, c_t - c_{t-1}, p_t - d_t, r_{d,t})'$
  5. Choose value ρ_d that most closely "matches" moments between dist of simulated and historical data ("match" made precise below.)

- Let $\{\widehat{y}_t\}_{t=1}^N$ denote simulated data (in obs eqn).

- Let $\{\widehat{y}_t\}_{t=1}^{N}$ denote simulated data (in obs eqn).
- Let $\{\tilde{y}_t\}_{t=1}^{T}$ denote historical data on same variables.

- Let $\{\widehat{y}_t\}_{t=1}^{N}$ denote simulated data (in obs eqn).
- Let $\{\tilde{y}_t\}_{t=1}^{T}$ denote historical data on same variables.
- *Auxiliary model* of hist. data: e.g., VAR, with density $f(y_t | y_{t-L}, ...y_{t-1}, \alpha)$, good LOM for data–$f$**-model**.

- Let $\{\widehat{y}_t\}_{t=1}^{N}$ denote simulated data (in obs eqn).
- Let $\{\tilde{y}_t\}_{t=1}^{T}$ denote historical data on same variables.
- *Auxiliary model* of hist. data: e.g., VAR, with density $f(y_t|y_{t-L}, ...y_{t-1}, \alpha)$, good LOM for data–***f*-model**.
- Score function of f-model:

$$s_f(y_t|y_{t-L}, ...y_{t-1}, \alpha) = \frac{\partial}{\partial \alpha} \ln[f(y_t|y_{t-L}, ..., y_{t-1}, \alpha)]$$

- Let $\{\widehat{y}_t\}_{t=1}^{N}$ denote simulated data (in obs eqn).
- Let $\{\tilde{y}_t\}_{t=1}^{T}$ denote historical data on same variables.
- *Auxiliary model* of hist. data: e.g., VAR, with density $f(y_t|y_{t-L},...y_{t-1},\alpha)$, good LOM for data–**$f$-model**.
- Score function of f-model:

$$s_f(y_t|y_{t-L},...y_{t-1},\alpha) = \frac{\partial}{\partial\alpha}\ln[f(y_t|y_{t-L},...,y_{t-1},\alpha)]$$

- QMLE estimator of auxiliary model on historical data

$$\tilde{\alpha} = \arg\max_{\alpha} \mathcal{L}_T(\alpha, \{\tilde{y}_t\}_{t=1}^{T})$$

$$\mathcal{L}_T(\alpha, \{\tilde{y}_t\}_{t=1}^{T}) = \frac{1}{T}\sum_{t=L+1}^{T}\ln f(\tilde{y}_t|\tilde{y}_{t-L},...,\tilde{y}_{t-1},\alpha)$$

- First-order-condition:

$$\frac{\partial}{\partial \alpha}\mathcal{L}_T(\tilde{\alpha}, \{\tilde{y}_t\}_{t=1}^T) = 0 \quad \text{or,} \quad \frac{1}{T}\sum_{t=L+1}^{T} s_f(\tilde{y}_t | \tilde{y}_{t-L}, ..., \tilde{y}_{t-1}, \tilde{\alpha}) = 0.$$

- First-order-condition:

$$\frac{\partial}{\partial \alpha}\mathcal{L}_T(\tilde{\alpha}, \{\tilde{y}_t\}_{t=1}^T) = 0 \quad \text{or,} \quad \frac{1}{T}\sum_{t=L+1}^{T} s_f(\tilde{y}_t | \tilde{y}_{t-L}, ..., \tilde{y}_{t-1}, \tilde{\alpha}) = 0.$$

- Idea: since above, good estimator for ρ_d is one that sets

$$\frac{1}{N}\sum_{t=L+1}^{N} s_f(\widehat{y}_t(\rho_d) | \widehat{y}_{t-L}(\rho_d), ..., \widehat{y}_{t-1}(\rho_d), \tilde{\boldsymbol{\alpha}}) \approx 0.$$

- First-order-condition:

$$\frac{\partial}{\partial \alpha}\mathcal{L}_T(\tilde{\alpha}, \{\tilde{y}_t\}_{t=1}^T) = 0 \quad \text{or,} \quad \frac{1}{T}\sum_{t=L+1}^{T} s_f(\tilde{y}_t | \tilde{y}_{t-L}, ..., \tilde{y}_{t-1}, \tilde{\alpha}) = 0.$$

- Idea: since above, good estimator for ρ_d is one that sets

$$\frac{1}{N}\sum_{t=L+1}^{N} s_f(\widehat{y}_t(\rho_d) | \widehat{y}_{t-L}(\rho_d), ..., \widehat{y}_{t-1}(\rho_d), \tilde{\alpha}) \approx 0.$$

- If $\dim(\alpha) > \dim(\rho_d)$, use GMM:

$$\underbrace{\widehat{m}_T(\rho_d, \alpha)}_{\dim(\alpha)\times 1} = \frac{1}{N}\sum_{t=L+1}^{N} s_f(\widehat{y}_t(\rho_d) | \widehat{y}_{t-L}(\rho_d), ..., \widehat{y}_{t-1}(\rho_d), \tilde{\alpha})$$

- First-order-condition:

$$\frac{\partial}{\partial \alpha}\mathcal{L}_T(\tilde{\alpha}, \{\tilde{y}_t\}_{t=1}^T) = 0 \quad \text{or,} \quad \frac{1}{T}\sum_{t=L+1}^{T} s_f(\tilde{y}_t|\tilde{y}_{t-L}, ..., \tilde{y}_{t-1}, \tilde{\alpha}) = 0.$$

- Idea: since above, good estimator for ρ_d is one that sets

$$\frac{1}{N}\sum_{t=L+1}^{N} s_f(\widehat{y}_t(\rho_d)|\widehat{y}_{t-L}(\rho_d), ..., \widehat{y}_{t-1}(\rho_d), \tilde{\alpha}) \approx 0.$$

- If $\dim(\alpha) > \dim(\rho_d)$, use GMM:

$$\underbrace{\widehat{m}_T(\rho_d, \alpha)}_{\dim(\alpha)\times 1} = \frac{1}{N}\sum_{t=L+1}^{N} s_f(\widehat{y}_t(\rho_d)|\widehat{y}_{t-L}(\rho_d), ..., \widehat{y}_{t-1}(\rho_d), \tilde{\alpha})$$

- The GMM estimator is

$$\widehat{\rho_d} = \arg\min_{\rho_d}\{\widehat{m}_T(\rho_d, \tilde{\alpha})'\tilde{\mathcal{I}}^{-1}\widehat{m}_T(\rho_d, \tilde{\alpha})$$

- GMM: $\widehat{\rho}_d = \underset{\rho_d}{\arg\min}\{\widehat{m}_T(\rho_d, \tilde{\alpha})' \tilde{\mathcal{I}}^{-1} \widehat{m}_T(\rho_d, \tilde{\alpha})\}$.
- $\tilde{\mathcal{I}}^{-1}$ is inv. of var. of score, data determined from f-model

$$\tilde{\mathcal{I}} = \sum_{t=1}^{T} \left\{ \frac{\partial}{\partial \tilde{\alpha}} \ln[f(\tilde{y}_t | \tilde{y}_{t-L}, ..., \tilde{y}_{t-1}, \tilde{\alpha})] \right\} \left\{ \frac{\partial}{\partial \tilde{\alpha}} \ln[f(\tilde{y}_t | \tilde{y}_{t-L}, ..., \tilde{y}_{t-1}, \tilde{\alpha})] \right\}'$$

- GMM: $\widehat{\rho}_d = \arg\min_{\rho_d}\{\widehat{m}_T(\rho_d, \tilde{\alpha})'\tilde{\mathcal{I}}^{-1}\widehat{m}_T(\rho_d, \tilde{\alpha})\}$.
- $\tilde{\mathcal{I}}^{-1}$ is inv. of var. of score, data determined from f-model

$$\tilde{\mathcal{I}} = \sum_{t=1}^{T}\left\{\frac{\partial}{\partial\tilde{\alpha}}\ln[f(\tilde{y}_t|\tilde{y}_{t-L}, ..., \tilde{y}_{t-1}, \tilde{\alpha})]\right\}\left\{\frac{\partial}{\partial\tilde{\alpha}}\ln[f(\tilde{y}_t|\tilde{y}_{t-L}, ..., \tilde{y}_{t-1}, \tilde{\alpha})]\right\}'$$

- Sims $\{\widehat{y}\}_{t=1}^{N}$ follow stationary dens. $p(y_{t-L}, ..., y_t|\rho_d)$. Note: no closed-form for $p(\cdot|\rho_d)$.

- GMM: $\widehat{\rho}_d = \arg\min_{\rho_d}\{\widehat{m}_T(\rho_d, \tilde{\alpha})'\tilde{\mathcal{I}}^{-1}\widehat{m}_T(\rho_d, \tilde{\alpha})\}$.

- $\tilde{\mathcal{I}}^{-1}$ is inv. of var. of score, data determined from f-model

$$\tilde{\mathcal{I}} = \sum_{t=1}^{T}\left\{\frac{\partial}{\partial\tilde{\alpha}}\ln[f(\tilde{y}_t|\tilde{y}_{t-L}, ..., \tilde{y}_{t-1}, \tilde{\alpha})]\right\}\left\{\frac{\partial}{\partial\tilde{\alpha}}\ln[f(\tilde{y}_t|\tilde{y}_{t-L}, ..., \tilde{y}_{t-1}, \tilde{\alpha})]\right\}'$$

- Sims $\{\widehat{y}\}_{t=1}^{N}$ follow stationary dens. $p(y_{t-L}, ..., y_t|\rho_d)$. Note: no closed-form for $p(\cdot|\rho_d)$.

- Intuition: $\widehat{m}_T(\rho_d, \alpha) \xrightarrow{as} m(\rho_d, \alpha)$ as $N \rightarrow \infty$, where

$$m(\rho_d, \alpha) = \int \cdots \int s(y_{t-L}, ..., y_t, \alpha)p(y_{t-L}, ..., y_t|\rho_d)dy_{t-L} \cdots dy_t$$

- GMM: $\widehat{\rho}_d = \underset{\rho_d}{\arg\min}\{\widehat{m}_T(\rho_d, \tilde{\alpha})'\tilde{\mathcal{I}}^{-1}\widehat{m}_T(\rho_d, \tilde{\alpha})\}$.
- $\tilde{\mathcal{I}}^{-1}$ is inv. of var. of score, data determined from f-model

$$\tilde{\mathcal{I}} = \sum_{t=1}^{T} \left\{ \frac{\partial}{\partial \tilde{\alpha}} \ln[f(\tilde{y}_t | \tilde{y}_{t-L}, ..., \tilde{y}_{t-1}, \tilde{\alpha})] \right\} \left\{ \frac{\partial}{\partial \tilde{\alpha}} \ln[f(\tilde{y}_t | \tilde{y}_{t-L}, ..., \tilde{y}_{t-1}, \tilde{\alpha})] \right\}'$$

- Sims $\{\widehat{y}\}_{t=1}^{N}$ follow stationary dens. $p(y_{t-L}, ..., y_t | \rho_d)$. Note: no closed-form for $p(\cdot|\rho_d)$.
- Intuition: $\widehat{m}_T(\rho_d, \alpha) \xrightarrow{as} m(\rho_d, \alpha)$ as $N \to \infty$, where

$$m(\rho_d, \alpha) = \int \cdots \int s(y_{t-L}, ..., y_t, \alpha) p(y_{t-L}, ..., y_t | \rho_d) dy_{t-L} \cdots dy_t$$

- $\Rightarrow$ use Monte Carlo compute expect. of $s(\cdot)$ under $p(\cdot|\rho_d)$.

- Intuition: $\widehat{m}_T(\rho_d, \alpha) \overset{as}{\rightarrow} m(\rho_d, \alpha)$ as $N \rightarrow \infty$, where

$$m(\rho_d, \alpha) = \int \cdots \int s(y_{t-L}, ..., y_t, \alpha) p(y_{t-L}, ..., y_t | \rho_d) dy_{t-L} \cdots dy_t$$

- Intuition: $\widehat{m}_T(\rho_d, \alpha) \xrightarrow{as} m(\rho_d, \alpha)$ as $N \rightarrow \infty$, where

$$m(\rho_d, \alpha) = \int \cdots \int s(y_{t-L}, ..., y_t, \alpha) p(y_{t-L}, ..., y_t | \rho_d) dy_{t-L} \cdots dy_t$$

- If $f = p$ above is mean of scores of likelihood. Should be zero, given f.o.c for MLE estimator.

- Intuition: $\widehat{m}_T(\rho_d, \alpha) \xrightarrow{as} m(\rho_d, \alpha)$ as $N \to \infty$, where

$$m(\rho_d, \alpha) = \int \cdots \int s(y_{t-L}, ..., y_t, \alpha) p(y_{t-L}, ..., y_t | \rho_d) dy_{t-L} \cdots dy_t$$

- If $f = p$ above is mean of scores of likelihood. Should be zero, given f.o.c for MLE estimator.
- Thus, if data *do* follow the structural model $p(\cdot|\rho_d)$, then $m(\rho_d^o, \alpha^o) = 0$, forms basis of a specification test.

- Intuition: $\widehat{m}_T(\rho_d, \alpha) \stackrel{as}{\rightarrow} m(\rho_d, \alpha)$ as $N \rightarrow \infty$, where

$$m(\rho_d, \alpha) = \int \cdots \int s(y_{t-L}, ..., y_t, \alpha) p(y_{t-L}, ..., y_t | \rho_d) dy_{t-L} \cdots dy_t$$

- If $f = p$ above is mean of scores of likelihood. Should be zero, given f.o.c for MLE estimator.
- Thus, if data *do* follow the structural model $p(\cdot|\rho_d)$, then $m(\rho_d^o, \alpha^o) = 0$, forms basis of a specification test.
- Summary: solve model for many values of ρ_d, store long simulations of model each time, do one-time estimation of auxiliary f-model. Choose ρ_d to minimize GMM criterion above.

- Advantages of using score functions as moments:

- Advantages of using score functions as moments:
  1. Computational: one-time estimation of structural model; useful if f-model is nonlinear.

- Advantages of using score functions as moments:
  1. Computational: one-time estimation of structural model; useful if f-model is nonlinear.
  2. If f-model good description of data, under null, MLE efficiency is obtained.

- Advantages of using score functions as moments:
  1. Computational: one-time estimation of structural model; useful if f-model is nonlinear.
  2. If f-model good description of data, under null, MLE efficiency is obtained.
- If $\dim(\alpha) > \dim(\rho_d)$, score-based SMM is consistent, asymptotically normal, *assuming:*

- Advantages of using score functions as moments:
  1. Computational: one-time estimation of structural model; useful if f-model is nonlinear.
  2. If f-model good description of data, under null, MLE efficiency is obtained.
- If $\dim(\alpha) > \dim(\rho_d)$, score-based SMM is consistent, asymptotically normal, *assuming:*
- That the auxiliary model is rich enough to identify non-linear structural model. Sufficient conditions for identification unknown.

- Advantages of using score functions as moments:
  1. Computational: one-time estimation of structural model; useful if f-model is nonlinear.
  2. If f-model good description of data, under null, MLE efficiency is obtained.
- If $\dim(\alpha) > \dim(\rho_d)$, score-based SMM is consistent, asymptotically normal, *assuming:*
- That the auxiliary model is rich enough to identify non-linear structural model. Sufficient conditions for identification unknown.
- Big issue: are these the economically interesting moments? Regards both choice of moments, and weighting function.

Consumption-Based Asset Pricing: Final Thoughts

- Little work linking financial markets to macroeconomic risks, given by primitives in the IMRS over consumption.

Consumption-Based Asset Pricing: Final Thoughts

- Little work linking financial markets to macroeconomic risks, given by primitives in the IMRS over consumption.
- No model that relates returns to other returns can *explain* asset prices in terms of primitive economic shocks. Such models of SDF only *describe* asset prices.

Consumption-Based Asset Pricing: Final Thoughts

- Little work linking financial markets to macroeconomic risks, given by primitives in the IMRS over consumption.
- No model that relates returns to other returns can *explain* asset prices in terms of primitive economic shocks. Such models of SDF only *describe* asset prices.
- So far many consumption-based models have been evaluated using calibration exercises $\Rightarrow$

Consumption-Based Asset Pricing: Final Thoughts

- Little work linking financial markets to macroeconomic risks, given by primitives in the IMRS over consumption.
- No model that relates returns to other returns can *explain* asset prices in terms of primitive economic shocks. Such models of SDF only *describe* asset prices.
- So far many consumption-based models have been evaluated using calibration exercises $\Rightarrow$
- A crucial next step in evaluating consumption-based models is structural econometric estimation. But...

Consumption-Based Asset Pricing: Final Thoughts

- Little work linking financial markets to macroeconomic risks, given by primitives in the IMRS over consumption.

- No model that relates returns to other returns can *explain* asset prices in terms of primitive economic shocks. Such models of SDF only *describe* asset prices.

- So far many consumption-based models have been evaluated using calibration exercises $\Rightarrow$

- A crucial next step in evaluating consumption-based models is structural econometric estimation. But...

  - ...models are imperfect and will never fit data infallibly.

  - Argue here for need to move away from testing if models are *true*, towards comparison of models based on *magnitude of misspecification*.

- Example: scaled consumption models.

- Example: scaled consumption models.
- Rather than ask whether scaled models are true...

Consumption-Based Asset Pricing: Final Thoughts

- Example: scaled consumption models.
- Rather than ask whether scaled models are true...
- ...ask whether allowing for *state-dependence* of SDF on consumption growth *reduces misspecification* over the analogous non-state-dependent model.

Consumption-Based Asset Pricing: Final Thoughts

- Macroeconomic data, unlike financial measured with error.

Consumption-Based Asset Pricing: Final Thoughts

- Macroeconomic data, unlike financial measured with error.

- $\Rightarrow$ Can't expect such models to perform as well as financial factor models of SDF.

Consumption-Based Asset Pricing: Final Thoughts

- Macroeconomic data, unlike financial measured with error.

- $\Rightarrow$ Can't expect such models to perform as well as financial factor models of SDF.

- True systematic risk factors are macroeconomic in nature; asset prices derived endogenously from these.

Consumption-Based Asset Pricing: Final Thoughts

- Macroeconomic data, unlike financial measured with error.
- $\Rightarrow$ Can't expect such models to perform as well as financial factor models of SDF.
- True systematic risk factors are macroeconomic in nature; asset prices derived endogenously from these.
- Financial factors could represent projection of true M_t on portfolios (i.e., mimicking portfolios).

Consumption-Based Asset Pricing: Final Thoughts

- Macroeconomic data, unlike financial measured with error.

- $\Rightarrow$ Can't expect such models to perform as well as financial factor models of SDF.

- True systematic risk factors are macroeconomic in nature; asset prices derived endogenously from these.

- Financial factors could represent projection of true M_t on portfolios (i.e., mimicking portfolios).

- In which case, they will always perform at least as well, *or better than*, mismeasured macro factors from true M_t $\Rightarrow$

Consumption-Based Asset Pricing: Final Thoughts

- Macroeconomic data, unlike financial measured with error.

- $\Rightarrow$ Can't expect such models to perform as well as financial factor models of SDF.

- True systematic risk factors are macroeconomic in nature; asset prices derived endogenously from these.

- Financial factors could represent projection of true M_t on portfolios (i.e., mimicking portfolios).

- In which case, they will always perform at least as well, *or better than*, mismeasured macro factors from true M_t $\Rightarrow$

- Not sensible to run horse races between financial factor models and macro models.

- Goal: not to find better factors, but rather to *explain* financial factors from deeper economic models.