Measuring the "Dark Matter" in Asset Pricing Models

Hui Chen[1] Winston Dou[2] Leonid Kogan[1]

[1]MIT and NBER

[2]UPenn

MFM Summer Session

June 20, 2017

Outline

Outline

Overview

Fragility Measures

Finite-sample Interpretation

Applications

- Disaster Risk
- Long-Run Risk

Conclusion

“Dark matter”-based asset pricing

- Some asset pricing models rely heavily on “dark matter”:
  (1) parameters are difficult to measure directly in the data;
  (2) results are sensitive to how the “dark matter” is specified.

- Common defense:
  - ↪ Assumed parameter values/model structure are acceptable if not rejected by the data;
  - ↪ The fact that they help “explain” asset prices is viewed as evidence supporting such choices.

"Dark matter"-based asset pricing

- Some asset pricing models rely heavily on "dark matter":
  (1) parameters are difficult to measure directly in the data;
  (2) results are sensitive to how the "dark matter" is specified.

- Common defense:
  ↪ Assumed parameter values/model structure are acceptable if not rejected by the data;
  ↪ The fact that they help "explain" asset prices is viewed as evidence supporting such choices.

- However, (1) + (2) imply model fragility:
  ↪ Implicit degrees of freedom is large;
  ↪ The model tends to over-fit the data in sample;
  ↪ Poor out-of-sample performance.

A traditional approach: sensitivity analysis

- A model is robust if its results are *“not sensitive to small perturbations”* in parameter values.

A traditional approach: sensitivity analysis

- A model is robust if its results are *"not sensitive to small perturbations"* in parameter values.
- How to formalize:
  1. How small is *"small"* — how should we choose the size of perturbation?
  2. What does it mean for the results to be *"not sensitive"* to such perturbations?
  3. Difficult to extend sensitivity analysis to high-dimensional parameter vectors.

A traditional approach: sensitivity analysis

- A model is robust if its results are *"not sensitive to small perturbations"* in parameter values.
- How to formalize:
  1. How small is *"small"* — how should we choose the size of perturbation?
  2. What does it mean for the results to be *"not sensitive"* to such perturbations?
  3. Difficult to extend sensitivity analysis to high-dimensional parameter vectors.
- We formally define a measure of model robustness and establish a convenient way to compute it.
  - ↪ multivariate sensitivity analysis
  - ↪ formal connection to over-fitting tendency

Model fragility: informational approach

- Prices P_t are linked to dividends x_t through a pricing model:

$$P_t = f(x_t, \theta, \varepsilon_t)$$

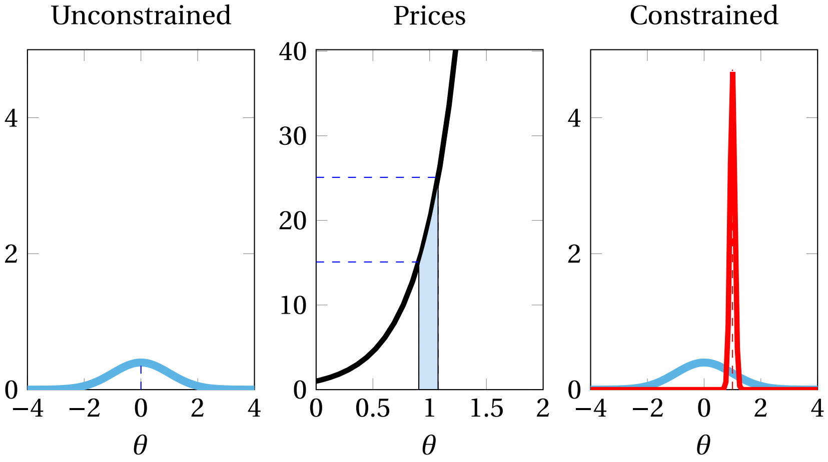
  $\hookrightarrow$ θ: governs the dynamics of x_t and affects stock prices

- We compare information about θ in

  (I) History of x_t

  (II) History of $\{x_t, P_t\}$ with cross-equation restrictions

Illustration



- Prices are sensitive to $\theta \Leftrightarrow$ prices are informative about θ (relative to fundamentals) $\Rightarrow$ sign of fragility

Outline

Overview

Fragility Measures

Finite-sample Interpretation

Applications
- Disaster Risk
- Long-Run Risk

Conclusion

A Generic Structural Model

- Baseline model: specifies dynamics for $\mathbf{x_t}$, with distribution $\mathbb{P}$
  - ↪ indexed by baseline parameters θ: $D_\Theta \times 1$
  - ↪ true baseline parameter value is θ_0

A Generic Structural Model

- Baseline model: specifies dynamics for $\mathbf{x_t}$, with distribution $\mathbb{P}$
  - $\hookrightarrow$ indexed by baseline parameters θ: $D_\Theta \times 1$
  - $\hookrightarrow$ true baseline parameter value is θ_0

- Full model: joint dynamics of $(\mathbf{x_t}, \mathbf{y_t})$, with distribution $\mathbb{Q}$
  - $\hookrightarrow$ indexed by θ and nuisance parameter ψ: $D_\Psi \times 1$
  - $\hookrightarrow$ true nuisance parameter value is ψ_0

A Generic Structural Model

- Baseline model: specifies dynamics for $\mathbf{x_t}$, with distribution $\mathbb{P}$
  - ↪ indexed by baseline parameters θ: $D_\Theta \times 1$
  - ↪ true baseline parameter value is θ_0

- Full model: joint dynamics of $(\mathbf{x_t}, \mathbf{y_t})$, with distribution $\mathbb{Q}$
  - ↪ indexed by θ and nuisance parameter ψ: $D_\Psi \times 1$
  - ↪ true nuisance parameter value is ψ_0

- Model performance is summarized by a set of moment conditions (why not MLE?)
  - ↪ baseline sub-model: $\mathbb{E}\left[g_{\mathbb{P}}(\theta_0; \mathbf{x_t})\right] = 0$
  - ↪ full model: $\mathbb{E}\left[g_{\mathbb{Q}}(\theta_0, \psi_0; \mathbf{x_t}, \mathbf{y_t})\right] = 0$
  - ↪ $g_{\mathbb{P}}(\theta_0; \mathbf{x_t})$ is a sub-vector of $g_{\mathbb{Q}}(\theta_0, \psi_0; \mathbf{x_t}, \mathbf{y_t})$

Quick Review of GMM

- Empirical moment conditions for the full model:

$$\widehat{g}_{\mathbb{Q},n}(\theta,\psi) \equiv \frac{1}{n}\sum_{t=1}^{n} g_{\mathbb{Q}}(\theta,\psi;\mathbf{x_t},\mathbf{y_t})$$

- Efficient GMM estimator $(\widehat{\theta}^{\mathbb{Q}},\widehat{\psi}^{\mathbb{Q}})$ minimizes:

$$\widehat{J}_{n,S_{\mathbb{Q}}}(\theta,\psi) \equiv n\widehat{g}_{\mathbb{Q},n}(\theta,\psi)^{T} S_{\mathbb{Q}}^{-1}\widehat{g}_{\mathbb{Q},n}(\theta,\psi)$$

- Spectral density matrix:

$$S_{\mathbb{Q}} = \sum_{\ell=-\infty}^{+\infty} \mathbb{E}_{\mathbb{Q}}\left[g(\theta_0,\psi_0;\mathbf{x_t},\mathbf{y_t})g(\theta_0,\psi_0;\mathbf{x_{t-\ell}},\mathbf{y_{t-\ell}})^{T}\right]$$

- Counterparts for the baseline model: $\widehat{g}_{\mathbb{P},n}(\theta)$, $\widehat{J}_{n,S_{\mathbb{P}}}(\theta)$, $S_{\mathbb{P}}$, $\widehat{\theta}^{\mathbb{P}}$

GMM Fisher Information Matrix

- Baseline model:

$$\mathbf{I}_{\mathbb{P}}(\theta) \equiv G_{\mathbb{P}}(\theta)^T S_{\mathbb{P}}^{-1} G_{\mathbb{P}}(\theta)$$

where $G_{\mathbb{P}}(\theta) \equiv \mathbb{E}\left[\nabla g_{\mathbb{P}}(\theta;\mathbf{x_t})\right]$

- Full model:

$$\mathbf{I}_{\mathbb{Q}}(\theta,\psi) \equiv G_{\mathbb{Q}}(\theta,\psi)^T S_{\mathbb{Q}}^{-1} G_{\mathbb{Q}}(\theta,\psi)$$

where $G_{\mathbb{Q}}(\theta,\psi) \equiv \mathbb{E}\left[\nabla g_{\mathbb{Q}}(\theta,\psi;\mathbf{x_t},\mathbf{y_t})\right]$

Marginal Fisher Information Matrix

- Efficient asymptotic variances of θ in the baseline model:

$$\mathbb{V}_{\mathbb{P}}(\theta) = \mathbf{I}_{\mathbb{P}}(\theta)^{-1}$$

Marginal Fisher Information Matrix

- Efficient asymptotic variances of θ in the baseline model:

$$\mathbb{V}_{\mathbb{P}}(\theta) = \mathbf{I}_{\mathbb{P}}(\theta)^{-1}$$

- Efficient asymptotic variances of θ in the full model:

$$\mathbb{V}_{\mathbb{Q}}(\theta|\psi) = \mathbf{I}_{\mathbb{Q}}(\theta|\psi)^{-1}$$

- Marginal GMM Fisher information matrix:

$$\mathbf{I}_{\mathbb{Q}}(\theta|\psi) \equiv \underbrace{\mathbf{I}_{\mathbb{Q}}^{(1,1)}(\theta,\psi)}_{\text{conditional infomation}} - \underbrace{\mathbf{I}_{\mathbb{Q}}^{(1,2)}(\theta,\psi)\mathbf{I}_{\mathbb{Q}}^{(2,2)}(\theta,\psi)\mathbf{I}_{\mathbb{Q}}^{(1,2)}(\theta,\psi)^{T}}_{\text{information loss due to uncertainty in } \psi}$$

$$\mathbf{I}_{\mathbb{Q}}(\theta,\psi) = \begin{bmatrix} \mathbf{I}_{\mathbb{Q}}^{(1,1)}(\theta,\psi), & \mathbf{I}_{\mathbb{Q}}^{(1,2)}(\theta,\psi) \\ \mathbf{I}_{\mathbb{Q}}^{(2,1)}(\theta,\psi), & \mathbf{I}_{\mathbb{Q}}^{(2,2)}(\theta,\psi) \end{bmatrix}$$

Fisher Fragility Measure

Definition

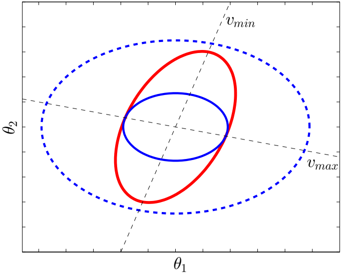
$$\varrho(\theta_0|\psi_0) = \mathbf{tr}\left[\mathbb{V}_{\mathbb{Q}}(\theta_0|\psi_0)^{-1}\mathbb{V}_{\mathbb{P}}(\theta_0)\right]$$

Fisher Fragility Measure

Definition

$$\varrho(\theta_0|\psi_0) = \mathbf{tr}\left[\mathbb{V}_{\mathbb{Q}}(\theta_0|\psi_0)^{-1}\mathbb{V}_{\mathbb{P}}(\theta_0)\right]$$

- Comparing estimation precision under full model and baseline model: $\mathbb{V}_{\mathbb{Q}}(\theta_0|\psi_0)$ vs. $\mathbb{V}_{\mathbb{P}}(\theta_0)$
- Informativeness of cross-equation restrictions
- $\varrho(\theta_0|\psi_0) = \lambda_1 + \lambda_2 + \cdots + \lambda_{D_\Theta}$



$\lambda_1 \geq \cdots \geq \lambda_{D_\Theta}$ are eigenvalues of $\mathbb{V}_{\mathbb{Q}}(\theta_0|\psi_0)^{-1/2}\mathbb{V}_{\mathbb{P}}(\theta_0)\mathbb{V}_{\mathbb{Q}}(\theta_0|\psi_0)^{-1/2}$

THE "WORST-CASE" DIRECTIONS

Definition

For any $1 \leq D \leq D_\Theta$,

$$\varrho^D(\theta_0|\psi_0) = \max_{\mathbf{v}\in\mathbb{R}^{D\times D_\Theta},\mathbf{Rank}(\mathbf{v})=D} \mathbf{tr}\left[\left(\mathbf{v}\mathbb{V}_{\mathbb{Q}}(\theta_0|\psi_0)\mathbf{v}^T\right)^{-1}\left(\mathbf{v}\mathbb{V}_{\mathbb{P}}(\theta_0)\mathbf{v}^T\right)\right]$$

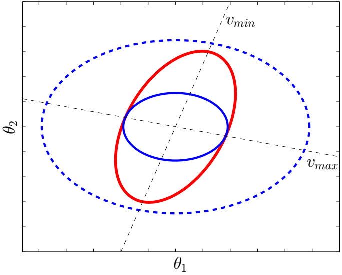
The "Worst-Case" Directions

Definition

For any $1 \leq D \leq D_\Theta$,

$$\varrho^D(\theta_0|\psi_0) = \max_{\mathbf{v}\in\mathbb{R}^{D\times D_\Theta},\mathbf{Rank}(\mathbf{v})=D} \mathbf{tr}\left[\left(\mathbf{v}\mathbb{V}_{\mathbb{Q}}(\theta_0|\psi_0)\mathbf{v}^T\right)^{-1}\left(\mathbf{v}\mathbb{V}_{\mathbb{P}}(\theta_0)\mathbf{v}^T\right)\right]$$

- Search all directions $\mathbf{v}$ for the largest discrepancy between $\mathbb{V}_{\mathbb{Q}}(\theta_0|\psi_0)$, $\mathbb{V}_{\mathbb{P}}(\theta_0)$
- "Effective sample size": extra sample needed to match the variance of $\mathbf{v}\theta$
- $\varrho^D(\theta_0|\psi_0) = \lambda_1 + \cdots + \lambda_D$



WHY IS ϱ A MEASURE OF MODEL FRAGILITY?

Over-fitting tendency of model $\mathbb{Q}$ relative to $\mathbb{P}$:

$$\varrho_o(\theta_0|\psi_0,\mathbf{x^n},\mathbf{y^n}) \equiv \int d_{S_{\mathbb{Q}}}\{\theta;\mathbf{x^n},\mathbf{y^n}\}\pi_{\mathbb{P}}(\theta|\mathbf{x^n})\mathrm{d}\theta,$$

- $d_{S_{\mathbb{Q}}}\{\theta;\mathbf{x^n},\mathbf{y^n}\}$ measures the over-fitting for a particular model θ, based on J-distance (GMM analog of the log likelihood ratio)

$$d_{S_{\mathbb{Q}}}\{\theta;\mathbf{x^n},\mathbf{y^n}\} = \underbrace{\widehat{J}_{n,S_{\mathbb{Q}}}(\theta,\check{\psi}^{\mathbb{Q}}) - \widehat{J}_{n,S_{\mathbb{Q}}}(\hat{\theta}^{\mathbb{Q}},\hat{\psi}^{\mathbb{Q}})}_{\text{J-distance gap} \geq 0}$$

  ↪ $(\hat{\theta}^{\mathbb{Q}},\hat{\psi}^{\mathbb{Q}})$ is the GMM estimator
  ↪ $(\theta,\check{\psi}^{\mathbb{Q}})$ is the constrained GMM estimator with fixed θ

WHY IS ϱ A MEASURE OF MODEL FRAGILITY?

Over-fitting tendency of model $\mathbb{Q}$ relative to $\mathbb{P}$:

$$\varrho_o(\theta_0|\psi_0, \mathbf{x^n}, \mathbf{y^n}) \equiv \int d_{S_\mathbb{Q}}\{\theta; \mathbf{x^n}, \mathbf{y^n}\}\pi_\mathbb{P}(\theta|\mathbf{x^n})\mathrm{d}\theta,$$

- $d_{S_\mathbb{Q}}\{\theta; \mathbf{x^n}, \mathbf{y^n}\}$ measures the over-fitting for a particular model θ, based on J-distance (GMM analog of the log likelihood ratio)

$$d_{S_\mathbb{Q}}\{\theta; \mathbf{x^n}, \mathbf{y^n}\} = \underbrace{\widehat{J}_{n,S_\mathbb{Q}}(\theta, \check{\psi}^\mathbb{Q}) - \widehat{J}_{n,S_\mathbb{Q}}(\hat{\theta}^\mathbb{Q}, \hat{\psi}^\mathbb{Q})}_{\text{J-distance gap} \geq 0}$$

  - ↪ $(\hat{\theta}^\mathbb{Q}, \hat{\psi}^\mathbb{Q})$ is the GMM estimator
  - ↪ $(\theta, \check{\psi}^\mathbb{Q})$ is the constrained GMM estimator with fixed θ

- Averaged over $\pi_\mathbb{P}(\theta|\mathbf{x^n})$, the posterior of θ in the baseline model.

- This measure extends the DIC measure for over-fitting tendency in statistics (Spiegelhalter et al. 2002)

Equivalence Results

Theorem: Under standard regularity conditions,

$$\operatorname*{wlim}_{n\to\infty} \varrho_o(\theta_0|\psi_0, \mathbf{x^n}, \mathbf{y^n}) = \varrho(\theta_0|\psi_0) + \sum_{i=1}^{D_\Theta} [\lambda_i - 1]\, \chi^2_{i,1}$$

Equivalence Results

Theorem: Under standard regularity conditions,

$$\underset{n\to\infty}{\text{wlim}}\, \varrho_o(\theta_0|\psi_0,\mathbf{x^n},\mathbf{y^n}) = \varrho(\theta_0|\psi_0) + \sum_{i=1}^{D_\Theta} [\lambda_i - 1]\, \chi^2_{i,1}$$

- "wlim": a variable with the limiting distribution (in the sense of weak convergence)
- $\chi^2_{1,i}$: i.i.d. chi-squared random variables with 1 degree of freedom

Average Tendency of Over-Fitting

$$\mathbb{E}\left[\operatorname*{wlim}_{n\to\infty} \varrho_o(\theta_0|\psi_0,\mathbf{x^n},\mathbf{y^n})\right] = 2\varrho(\theta_0|\psi_0) - D_\Theta$$

Interpretation:

- $\varrho(\theta_0|\psi_0)$ is the implicit degrees of freedom, while D_Θ is the standard explicit degrees of freedom
- $\varrho(\theta_0|\psi_0)$ captures the average tendency of over-fitting when sample size is large
- When bootstrapping based on a large sample $(\mathbf{x^n},\mathbf{y^n})$ to estimate the average tendency of over-fitting, the result should be close to $2\varrho(\theta_0|\psi_0) - D_\Theta$

Tail distribution for ϱ

Tail probability of the limiting variable converges to zero at an exponential rate:

$$\lim_{x\to\infty} \frac{1}{x} \ln \mathbb{P}\left\{\operatorname*{wlim}_{n\to\infty} \varrho_o(\theta_0|\psi_0, \mathbf{x^n}, \mathbf{y^n}) > x\right\} = -\frac{1}{2\left(\lambda_1 - 1\right)}$$

Interpretation:

- The thickness of the tail distribution is fully characterized by the largest eigenvalue λ_1, which is the worst-case 1-D Fisher fragility measure $\varrho^1(\theta_0|\psi_0)$.
- Given the average level $\varrho(\theta_0|\psi_0)$, the more the distribution of eigenvalues is concentrated on λ_1, the higher the probability of extreme over-fitting.

Outline

Overview

Fragility Measures

Finite-sample Interpretation

Applications
- Disaster Risk
- Long-Run Risk

Conclusion

Informativeness of Economic Restrictions

- In finite sample, we can compare posterior beliefs about θ
  - ↪ Compare the entire posterior distributions
  - ↪ Quantify the additional information in cross-equation restrictions

- Prior and Posteriors
  - ↪ $\pi(\theta)$: the common prior about θ
  - ↪ $\pi_{\mathbb{P}}(\theta|\mathbf{x}^{\mathbf{n}})$: posterior based on baseline model and $\mathbf{x}^{\mathbf{n}}$
  - ↪ $\pi_{\mathbb{Q}}(\theta|\mathbf{x}^{\mathbf{n}}, \mathbf{y}^{\mathbf{n}})$: posterior based on full model and $(\mathbf{x}^{\mathbf{n}}, \mathbf{y}^{\mathbf{n}})$

- Use relative entropy to measure the statistical discrepancy between $\pi_{\mathbb{P}}(\theta|\mathbf{x}^{\mathbf{n}})$ and $\pi_{\mathbb{Q}}(\theta|\mathbf{x}^{\mathbf{n}}, \mathbf{y}^{\mathbf{n}})$:

$$\mathbf{D}_{KL}(\pi_{\mathbb{Q}}(\theta|\mathbf{x}^{\mathbf{n}}, \mathbf{y}^{\mathbf{n}})||\pi_{\mathbb{P}}(\theta|\mathbf{x}^{\mathbf{n}})) = \int \ln\left(\frac{\pi_{\mathbb{Q}}(\theta|\mathbf{x}^{\mathbf{n}}, \mathbf{y}^{\mathbf{n}})}{\pi_{\mathbb{P}}(\theta|\mathbf{x}^{\mathbf{n}})}\right) \pi_{\mathbb{Q}}(\theta|\mathbf{x}^{\mathbf{n}}, \mathbf{y}^{\mathbf{n}})\, d\theta$$

Effective Sample Size

- Relative entropy is difficult to interpret. We normalize it by asking:
  - ↪ How much extra data (m^*) is needed on average to generate the same amount of additional information about θ based only on the baseline model
- The additional information provided by "effective sample" is measured by the mutual information between θ and $\tilde{\mathbf{x}}^{\mathbf{m}}$ conditional observed data $\mathbf{x}^{\mathbf{n}}$:

$$I(\tilde{\mathbf{x}}^{\mathbf{m}};\theta|\mathbf{x}^{\mathbf{n}}) = \mathbb{E}^{\tilde{\mathbf{x}}^{\mathbf{m}}|\mathbf{x}^{\mathbf{n}}}\left[\mathbf{D}_{KL}(\pi_{\mathbb{P}}(\theta'|\tilde{\mathbf{x}}^{\mathbf{m}},\mathbf{x}^{\mathbf{n}})||\pi_{\mathbb{P}}(\theta'|\mathbf{x}^{\mathbf{n}}))\right]$$

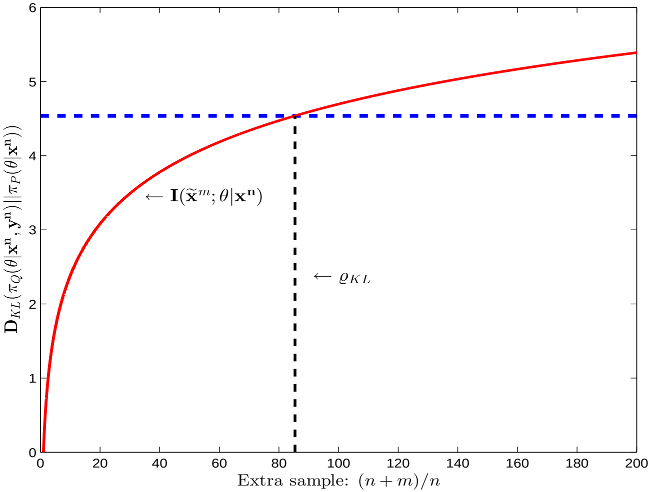
- Effective sample size ratio:

$$\varrho_{KL}(\mathbf{x}^{\mathbf{n}},\mathbf{y}^{\mathbf{n}}) = \frac{n+m^*}{n}$$

such that

$$\mathbf{D}_{KL}(\pi_{\mathbb{Q}}(\theta|\mathbf{x}^{\mathbf{n}},\mathbf{y}^{\mathbf{n}})||\pi_{\mathbb{P}}(\theta|\mathbf{x}^{\mathbf{n}})) = I(\tilde{\mathbf{x}}^{\mathbf{m}^*};\theta|\mathbf{x}^{\mathbf{n}})$$

Illustration: Solving for ϱ_{KL}



Asymptotic Equivalence

Under standard regularity conditions,

$$\mathbb{E}\left[\underset{n\to\infty}{\text{wlim}} \ln \varrho_{KL}(\mathbf{x^n}, \mathbf{y^n})\right] = \ln \varrho(\theta_0|\psi_0)$$

Interpretation:

- $\varrho(\theta_0|\psi_0)$ captures the additional effective sample size under a rigorous information-theoretic framework
- It further justifies the economic interpretation: informativeness of cross-equation restrictions implied by theories

Outline

Disaster risk model

- Disaster: $z_t \in (0,1)$, $\Pr(z_t = 1) = p$
- Baseline model for (log) consumption growth g_t:
  - ↪ $z_t = 0$: $g_t \sim N(\mu, \sigma^2)$
  - ↪ $z_t = 1$: $g_t = -v_t$, where $v_t \sim \mathbf{1}\{v_t > \underline{v}\}\xi e^{-\xi(v_t - \underline{v})}$
- Full model: adds excess log return on the market portfolio r_t
  - ↪ $z_t = 0$: r_t and g_t are jointly normal,

$$r_t = \eta + \rho\frac{\tau}{\sigma}(g_t - \mu) + \sqrt{1-\rho^2}\,\tau\varepsilon_{0,t},$$

  - ↪ $z_t = 1$:

$$r_t = bg_t + \varsigma\varepsilon_{1,t}$$

  - ↪ Representative agent with power utility: γ
- “Dark matter” parameters: p, ξ

Cross-equation restriction

- Consumption Euler equation for excess log returns

$$1 = \mathbb{E}_t \left[\frac{m_{t+1}}{\mathbb{E}_t[m_{t+1}]} e^{r_{t+1}} \right]$$

- The log equity premium is then given by

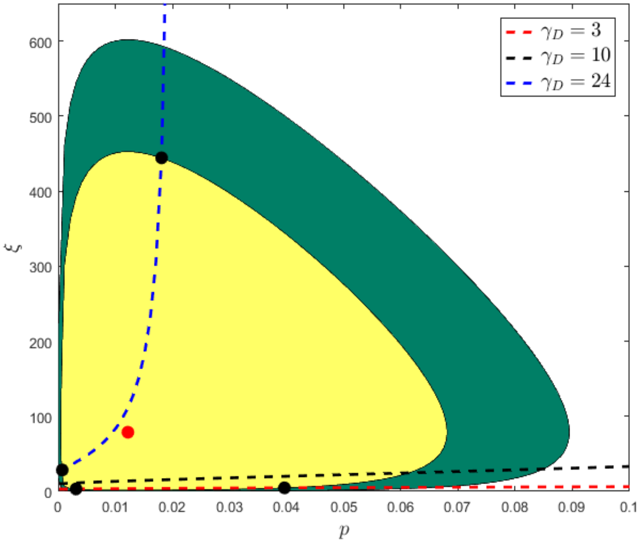
$$\mathbb{E}[r_t] = (1-p)\eta - pb\left(\underline{v} + 1/\xi\right),$$

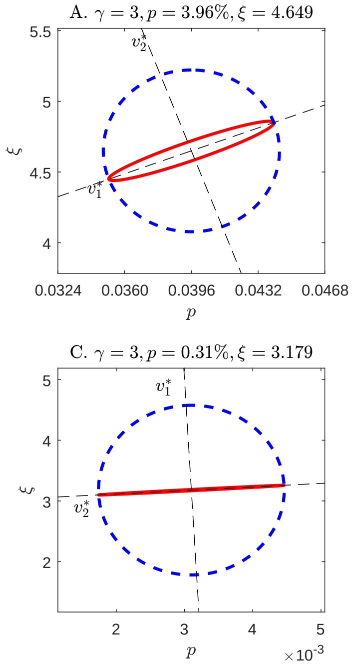
$$\eta \approx \gamma\rho\sigma\tau - \frac{\tau^2}{2} + \underbrace{e^{\gamma\mu - \frac{\gamma^2\sigma^2}{2}} \xi \left(\frac{e^{\gamma\underline{v}}}{\xi - \gamma} - \frac{e^{\frac{\varsigma^2}{2} + (\gamma - b)\underline{v}}}{\xi + b - \gamma} \right) \frac{p}{1-p}}_{\text{disaster risk premium}}$$

- Key properties:
  1. η is unbounded as $\xi \to \gamma$ (no matter how small p is)
  2. Large disaster size (small ξ) $\Longrightarrow$ high sensitivity of η to p

Disaster risk model: An Irrefutable Model?

"Dark matter" parameters: p, ξ





A. $\gamma = 3, p = 3.96\%, \xi = 4.649$
B. $\gamma = 24, p = 1.81\%, \xi = 446.36$
C. $\gamma = 3, p = 0.31\%, \xi = 3.179$
D. $\gamma = 24, p = 0.0699\%, \xi = 28.43$
v_1^*
v_2^*
ξ
p
5.5
5
4.5
4
0.0324
0.0360
0.0396
0.0432
0.0468
550
500
450
400
350
0.014
0.016
0.018
0.02
0.022
2
3
4
5
$\times 10^{-3}$
40
35
30
25
20
6
7
8
9
$\times 10^{-4}$

Long-Run Risk Model

- Consumption processes:

$$\begin{aligned}\Delta c_{t+1} &= \mu_c + x_t + \sigma_t \epsilon_{c,t+1} \\ x_{t+1} &= \rho x_t + \varphi_x \sigma_t \epsilon_{x,t+1} \\ \tilde{\sigma}^2_{t+1} &= \overline{\sigma}^2 + \nu(\tilde{\sigma}^2_t - \overline{\sigma}^2) + \sigma_w \epsilon_{\sigma,t+1} \\ \sigma^2_{t+1} &= \max\left(\underline{\sigma}^2, \tilde{\sigma}^2_{t+1}\right)\end{aligned}$$

- Dividend process:

$$\Delta d_{t+1} = \mu_d + \phi_d x_t + \varphi_{d,c} \sigma_t \epsilon_{c,t+1} + \varphi_{d,d} \sigma_t \epsilon_{d,t+1}$$

- Representative agent with EZW preference: δ_{L}, γ_{L} and ψ_{L}
- From the consumption Euler equation, derive a linear approximation of the stochastic discount factor

$$m_{t+1} = \Gamma_0 + \Gamma_1 x_t + \Gamma_2 \sigma^2_t - \lambda_c \sigma_t \epsilon_{c,t+1} - \lambda_x \varphi_x \sigma_t \epsilon_{x,t+1} - \lambda_\sigma \sigma_w \epsilon_{\sigma,t+1}$$

Cross-Equation Restrictions

- In equilibrium, the excess (log) market return can be written as

$$r^e_{m,t+1} = \mu^e_{r,t} + \underbrace{\beta_\eta \sigma_t \eta_{t+1} + \beta_e \sigma_t e_{t+1} + \beta_w \sigma_w w_{t+1} + \varphi_{d,d} \sigma_t u_{d,t+1}}_{\text{Loadings on shocks}}$$

- The conditional expected excess (log) return $\mu^e_{r,t} = \mathbb{E}_t \left[r^e_{m,t+1} \right]$ is

$$\mu^e_{r,t} = \lambda_\eta \beta_\eta \sigma^2_{p,t} + \lambda_e \beta_e \sigma^2_{p,t} + \lambda_w \beta_w \sigma^2_w - \frac{1}{2} \left(\beta^2_\eta \sigma^2_t + \beta^2_e \sigma^2_t + \beta^2_w \sigma^2_w + \varphi^2_{d,d} \sigma$$

- β's and λ's are determined as fixed points in the equilibrium.

Baseline Model Choice

- Baseline model: $\mathbf{x_t} = (\Delta c_{t+1}, x_t, \sigma_t^2, \Delta d_{t+1})$
  - ↪ baseline parameters $\theta = (\mu_c, \rho, \varphi_x, \overline{\sigma}^2, \nu, \sigma_w, \mu_d, \phi_d, \varphi_{d,c}, \varphi_{d,d})$
- Full model: asset pricing model with $\mathbf{y_t} = r^e_{m,t+1}$
  - ↪ nuisance parameters $\psi = (\gamma_{\mathrm{L}}, \psi_{\mathrm{L}})$

Model I

- Bansal-Kiku-Yaron (2012) parameters

Preferences	δ_{L}	γ_{L}	ψ_{L}			
	0.9989	**10**	1.5			
Consumption	μ_c	ρ	φ_e	$\overline{\sigma}$	ν	σ_w
	0.0015	0.975	0.038	0.0072	**0.999**	$2.8e-6$
Dividends	μ_d	ψ	π	$\varphi_{d,d}$		
	0.0015	2.5	2.6	5.96		
Other	$\underline{\sigma}$					
	0.0001					

Does a good job matching asset pricing moments

- Simulated and sample moments for the benchmark calibration:

	Data	Model		
Moment	Estimate	5%	Median	95%
$\mathbb{E}\left[r_M - r_f\right]$	7.09	2.33	5.88	10.58
$\mathbb{E}\left[r_M\right]$	7.66	2.91	6.66	11.20
$\sigma\left(r_M\right)$	20.28	12.10	20.99	29.11
$\mathbb{E}\left[r_f\right]$	0.57	-0.20	0.77	1.45
$\sigma\left(r_f\right)$	2.86	0.64	1.07	1.62
$\mathbb{E}\left[p-d\right]$	3.36	2.69	2.99	3.30

Model II (Alternative Calibration)

Preferences	δ_{L}	γ_{L}	ψ_{L}			
	0.9989	**27**	1.5			
Consumption	μ_c	ρ	φ_e	$\overline{\sigma}$	ν	σ_w
	0.0015	0.975	0.038	0.0072	**0.98**	$2.8e-6$
Dividends	μ_d	ψ	π	φ		
	0.0015	2.5	2.6	5.96		
Other	$\underline{\sigma}$					
	0.0001					

Also matches asset pricing moments

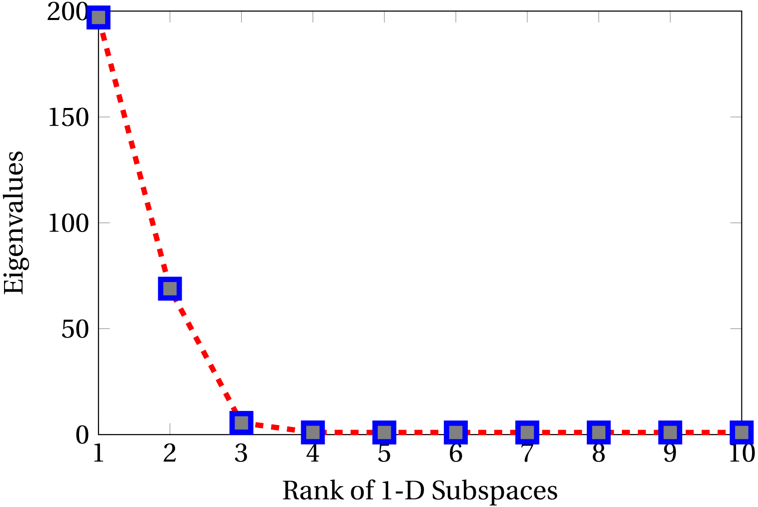
- Simulated and sample moments for alternative calibration:

	Data	Model		
Moment	Estimate	5%	Median	95%
$\mathbb{E}\left[r_M - r_f\right]$	7.09	3.65	6.78	10.05
$\mathbb{E}\left[r_M\right]$	7.66	4.42	7.75	11.20
$\sigma\left(r_M\right)$	20.28	15.01	17.55	20.33
$\mathbb{E}\left[r_f\right]$	0.57	0.47	0.96	1.46
$\sigma\left(r_f\right)$	2.86	0.73	0.94	1.23
$\mathbb{E}\left[p - d\right]$	3.36	2.77	2.81	2.85

But they are very different in terms of fragility

Model	ϱ	ϱ^1	ϱ^v									
			μ_c	ρ	φ_x	$\overline{\sigma}^2$	ν	σ_w	μ_d	ϕ_d	$\varphi_{d,c}$	$\varphi_{d,d}$
I. Nuisance parameter vector ψ: $(\gamma_{\mathrm{L}}, \psi_{\mathrm{L}})$												
(M1)	276.3	196.3	1.0	1.1	1.0	48.9	97.8	1.0	1.0	3.4	1.0	1.0
(M2)	34.0	21.1	1.0	1.1	1.0	1.0	3.4	1.0	1.4	4.2	1.0	1.0
II. Nuisance parameter vector ψ: empty												
(M1)	$3.58 \cdot 10^5$	$3.57 \cdot 10^5$	1.0	2.1	1.1	115.6	117.5	1.3	1.1	7.1	1.0	1.0
(M2)	323.3	287.7	1.0	2.5	1.0	1.0	6.3	1.0	1.9	31.3	1.0	1.0

Diagnosing the sources of fragility



Outline

Overview

Fragility Measures

Finite-sample Interpretation

Applications
- Disaster Risk
- Long-Run Risk

Conclusion

Informativeness: How to Interpret?

What does it mean when the cross-equation restrictions in a model appear highly informative?

- Powerful identification if economic restrictions are valid.
- Sign of fragility
    - ↪ Why should we be concerned: prone to over-fitting in sample, poor out-of-sample performance, problematic implications for counter-factual analysis ...

How to Use the Fragility Measure?

- To improve model selection and inference
  - ↪ Model selection: measure the fragility of a class of models
  - ↪ Structural estimation: guard against mis-specification and false confidence in inference

- Model diagnostics: To identify aspects of the model that require more data or theory support.
  - ↪ International evidence? Survey data?
  - ↪ If it is about agents' beliefs, can they be micro-founded, e.g. via preferences, information aggregation, market frictions?
  - ↪ Explicitly analyze agents' uncertainties inside the model (Hansen-Sargent)

- Method can be applied to structural models in many fields (structural corporate, IO, macro, ...)