

---

# Cross-Sectional Asset Pricing with Individual Stocks: Betas versus Characteristics

Tarun Chordia, Amit Goyal, and Jay Shanken



# Main question

---

- Are expected returns related to
  - Risk/betas, OR
  - Characteristics
  
- If both, which is more important?

# How to answer?

---

- Use portfolios
  - Helps mitigate EIV problem (Fama and MacBeth, 1973)
  - But,
    - Less efficient (Ang, Liu, and Schwarz, 2010)
    - Method of grouping is important (Lewellen, Nagel, and Shanken, 2010)
- Use individual securities
  - But, EIV problem

# What we do

---

- Use individual securities
- **Correct for EIV bias** (Litzenberger and Ramaswamy, 1979; Shanken, 1992; Kim, 1995)
  - And a few other biases
- **Allow betas to change over time** (Rosenberg and Guy, 1976; Shanken, 1990; Fama and French, 1997; Avramov and Chordia, 2006)
- Quantify relative contribution of risk loadings versus characteristics in explaining the cross-section of returns

# Two-pass procedure

---

- Time-series regression (TSR)
  - Condition betas on
    - Firm characteristics (Size, B/M, six-month return), or
    - Firm characteristics (Size, B/M, six-month return) + Macro variables (Term spread, Default spread)
  - One-, three-, and four-factor models
- Cross-sectional regression (CSR)
- For now, we do not allow for time-variation in risk premia but relax this later on in the third-stage

# TSR

---

- Unconditional TSR

$$R_{it} = B_{0i} + B_i F_t + \varepsilon_{it}$$

- Conditional TSR

- Define  $zts_{it}$  to be a  $p \times 1$  vector of conditioning variables
- Redefine scaled intercept and factors as

$$F_{it}^* = [zts'_{(p-1)it-1}, zts_{i1t-1} F'_t, \dots, zts_{ipt-1} F'_t]'$$

- Then, the TSR is

$$R_{it} = B_{0i}^* + B_i^* F_{it}^* + \varepsilon_{it}$$

# TSR ...

---

- Example of a one-factor model with firm characteristics conditioning variables

$$R_{it} = (B_{i0}^* + B_{i1}^* Sz_{it-1} + B_{i2}^* BM_{it-1} + B_{i3}^* Ret6_{it-2}) \\ + (B_{i0}^{Mkt^*} + B_{i1}^{Mkt^*} Sz_{it-1} + B_{i2}^{Mkt^*} BM_{it-1} + B_{i3}^{Mkt^*} Ret6_{it-2}) Mkt_t \\ + \varepsilon_{it}$$

Characteristics are cross-sectionally demeaned

- Then the implied market beta is given by

$$B_{it-1}^{Mkt} = (B_{i0}^{Mkt^*} + B_{i1}^{Mkt^*} Sz_{it-1}^* + B_{i2}^{Mkt^*} BM_{it-1}^* + B_{i3}^{Mkt^*} Ret6_{it-2}^*)$$

# CSR

---

- Cross-sectional regression (CSR) using OLS

$$R_t = \gamma_{0t} + \hat{B}_{t-1}\gamma_{1t} + Zcs_{t-1}\gamma_{2t} + \varepsilon_t$$

$$\hat{\Gamma}_t \equiv (\hat{\gamma}_{0t}, \hat{\gamma}'_{1t}, \hat{\gamma}'_{2t})'$$

$$\hat{\Gamma}_t = \left( \hat{X}'_t \hat{X}_t \right)^{-1} \hat{X}'_t R_t, \text{ where } \hat{X}_t \equiv \left[ \mathbf{1}_{N_t} : \hat{B}_{t-1} : Zcs_{t-1} \right]$$



# CSR biases

---

$$\begin{aligned}\hat{X}_t &= \left[ \mathbf{1}_{N_t} : \hat{B}_{t-1} : Zcs_{t-1} \right] \\ &= \left[ \mathbf{1}_{N_t} : B_{t-1} : Zcs_{t-1} \right] + \left[ \mathbf{0} : \hat{B}_{t-1} - B_{t-1} : \mathbf{0} \right] \\ &= X_t + \left[ \mathbf{0} : U_t : \mathbf{0} \right]\end{aligned}$$

- Estimation error in betas,  $U$ , is the cause of all the trouble

# CSR biases – EIV

---

- $\hat{X}'_t \hat{X}_t$  contains a term  $U'_t U_t$ , which is the (cross-sectional sum of the) of the estimation error variance in betas
- This increases the “denominator” and causes the classic EIV problem
- Fortunately, we know the estimation error in betas from TSR

$$\hat{\Gamma}_t^{\text{EIV}} = \left( \hat{X}'_t \hat{X}_t - \sum_{i=1}^{N_t} M' \hat{\Sigma}_{\hat{B}_{it-1}} M \right)^{-1} \hat{X}'_t R_t$$

$$\hat{\Sigma}_{\hat{B}_{it-1}} = Zts'_{it-1} \hat{\Sigma}_{\hat{B}_i^*}^{\text{White}} Zts_{it-1}, \text{ and } M = \begin{bmatrix} \mathbf{0}_{k \times 1} & I_{k \times k} & \mathbf{0}_{k \times k_2} \end{bmatrix}$$

# Intuition ...

---

- For a one-factor model with no additional characteristics

$$\gamma^{OLS} = \frac{\text{cov}_{cs}(\hat{\beta}_{it}, R_{it})}{\text{var}_{cs}(\hat{\beta}_{it})} = \frac{\text{cov}_{cs}(\beta_{it}, R_{it})}{\text{var}_{cs}(\hat{\beta}_{it})}$$

- The denominator is equal to the cross-sectional variation of betas (this is what we want) plus the cross-sectional variation of beta estimation error (this is what we want to remove).

Fortunately

$$\begin{aligned}\text{var}_{cs}(\hat{\beta}_{it}) &= \text{var}_{cs}(\beta_{it}) + \text{var}_{cs}(\hat{\beta}_{it} - \beta_{it}) \\ &= \text{var}_{cs}(\beta_{it}) + \text{var}_{cs}(\sigma_{ei}^2 \Sigma_F^{-1})\end{aligned}$$

- This means

$$\gamma^{EIV} = \frac{\text{cov}_{cs}(\beta_{it}, R_{it})}{\text{var}_{cs}(\hat{\beta}_{it}) - \text{var}_{cs}(\sigma_{ei}^2 \Sigma_F^{-1})}$$

# CSR biases – 1-month bias

---

- The “numerator” contains  $U_t' \varepsilon_t$ . This should cancel out in balanced panels and homosekadastic case
  - In one-factor case, this term is equal to the sum of
$$\left(Rm_t - \overline{Rm}\right) \varepsilon_{it}^2$$
- In heteroskedastic case (as observed in the data), this term is not zero and leads to the second correction

$$bias_t = \begin{pmatrix} 0 \\ \sum_{i=1}^{N_t} Zts'_{it-1} (F_{di}^{*'} F_{di}^*)^{-1} F_{dit}^{*'} e_{it}^2 \\ 0_{k_2 \times 1} \end{pmatrix}$$

# CSR biases – another bias

---

- Both the denominator and the numerator contain term such as  $U'_t Zcs_{t-1}$
- Since  $Zcs$  contains price-related variables (size, B/M, six-month return), this term is also not zero
- This necessitates a third correction
  - We assume for simplicity an AR(1) process for size and B/M to implement this correction

# Final formula

---

$$\hat{\Gamma}_t^{\text{final}} = \left( \hat{X}_t' \hat{X}_t - \sum_{i=1}^{N_t} M' \hat{\Sigma}_{B_{it-1}} M - \text{another bias}_t \right)^{-1} \left( \hat{X}_t' R_t - \text{bias}_t \right)$$

# Contribution measures

---

- Using average CSR estimates

$$E_{t-1} [R_t] = \overline{\hat{\gamma}_0} + E_{t-1}^{\text{beta}} [R_t] + E_{t-1}^{\text{char}} [R_t], \text{ where}$$

$$E_{t-1}^{\text{beta}} [R_t] = \hat{B}_{t-1} \overline{\hat{\gamma}_1}, \text{ and } E_{t-1}^{\text{char}} [R_t] = ZCS_{t-1} \overline{\hat{\gamma}_2}$$

- Using cross-sectional variation at time  $t$

$$\%(\text{Betas}) = \text{var}_{cs} \left( E_{t-1}^{\text{beta}} [R_t] \right) / \text{var}_{cs} \left( E_{t-1} [R_t] \right)$$

$$\%(\text{Chars}) = \text{var}_{cs} \left( E_{t-1}^{\text{char}} [R_t] \right) / \text{var}_{cs} \left( E_{t-1} [R_t] \right)$$

# Contribution measures ...

---

$$\text{var}_{cs} \left( E_{t-1}^{\text{beta}} [R_t] \right) = \overline{\hat{\gamma}'_1} \text{var}_{cs} \left( \hat{B}_{t-1} \right) \overline{\hat{\gamma}_1}$$

- Have to again correct the cross-sectional variance of estimated betas using the same trick as in the regular EIV correction
- Standard errors for contributions
  - Resample gammas using their standard errors
  - Recalculate contribution numbers
  - Repeat 1,000 times to obtain empirical distribution
    - Use Efron's procedure to account for non-linearity



# Data

---

- All common stocks on NYSE, AMEX, and NASDAQ
- Sample: 1951 to 2011
- Price greater than \$1 (for CSR)
  
- At least five years of data
  - Could lead to survivorship bias
  - We do not include stocks in CSR in the first five years of their life

# One-factor model

<i>zts</i> →	Panel A: All Stocks			
	Bias uncorrected		Bias corrected	
	Firm	Firm + Macro	Firm	Firm + Macro
Cnst	0.686 (4.02)	0.825 (4.68)	0.838 (4.96)	1.060 (6.11)
B <sub>Mkt</sub>	0.331 (2.13)	0.235 (1.56)	0.158 (0.91)	-0.053 (-0.31)
Sz	-0.043 (-1.56)	-0.048 (-1.72)	-0.045 (-1.67)	-0.050 (-1.82)
B/M	0.282 (6.04)	0.267 (5.79)	0.274 (5.92)	0.269 (5.68)
Ret6	1.248 (8.93)	1.294 (9.06)	1.097 (6.86)	1.065 (6.64)
Nstocks	2,025	2,025	2,025	2,025
% Betas	20.7	14.0	3.6	0.5
% Chars	82.7	88.4	97.6	99.5
% Diff	62.0 (6.5,98.6)	74.4 (10.7,99.9)	93.9 (49.3,100.0)	99.0 (88.3,100.0)

# Three-factor model

<i>zts</i> →	Panel A: All Stocks			
	Bias uncorrected		Bias corrected	
	Firm	Firm + Macro	Firm	Firm + Macro
Cnst	0.406 (3.39)	0.537 (4.70)	0.532 (3.52)	0.677 (4.79)
B <sub>Mkt</sub>	0.254 (1.64)	0.186 (1.24)	0.072 (0.43)	0.000 (-0.00)
B <sub>SMB</sub>	0.241 (2.30)	0.206 (2.03)	0.233 (1.96)	0.191 (1.73)
B <sub>HML</sub>	-0.006 (-0.06)	0.020 (0.21)	0.018 (0.17)	0.029 (0.30)
Sz	0.003 (0.15)	-0.008 (-0.48)	0.001 (0.04)	-0.007 (-0.36)
B/M	0.269 (9.40)	0.242 (8.65)	0.273 (6.93)	0.276 (6.98)
Ret6	1.257 (10.40)	1.288 (10.91)	1.115 (7.68)	1.103 (7.90)
Nstocks	2,025	2,025	2,025	2,025
% Betas	38.8	36.3	23.5	19.9
% Chars	61.4	61.8	71.4	72.0
% Diff	22.6	25.6	47.8	52.1
	(-28.4,86.1)	(-22.0,89.7)	(-10.2,99.4)	(2.9,99.5)

# Four-factor model

<i>zts</i> →	Panel A: All Stocks			
	Bias uncorrected		Bias corrected	
	Firm	Firm + Macro	Firm	Firm + Macro
Cnst	0.398 (3.47)	0.510 (4.67)	0.566 (3.79)	0.663 (4.81)
B <sub>Mkt</sub>	0.257 (1.65)	0.198 (1.31)	0.095 (0.46)	0.084 (0.53)
B <sub>SMB</sub>	0.239 (2.26)	0.213 (2.08)	0.197 (1.67)	0.171 (1.54)
B <sub>HML</sub>	0.008 (0.08)	0.034 (0.36)	0.054 (0.52)	0.045 (0.45)
B <sub>MOM</sub>	0.186 (1.28)	0.202 (1.43)	0.428 (2.48)	0.404 (2.59)
Sz	0.005 (0.28)	-0.005 (-0.30)	-0.003 (-0.14)	-0.014 (-0.73)
B/M	0.276 (10.33)	0.251 (9.66)	0.272 (6.96)	0.291 (7.48)
Ret6	1.265 (12.72)	1.307 (13.40)	1.062 (8.21)	0.999 (7.69)
Nstocks	2,025	2,025	2,025	2,025
% Betas	40.9	45.0	31.9	38.5
% Chars	54.4	49.3	55.4	50.8
% Diff	13.5	4.3	23.5	12.3
	(-23.7,75.6)	(-30.8,68.6)	(-15.5,74.3)	(-22.9,73.1)

# Time variation in risk premia

---

- Allow predictability in risk premia with dividend-price ratio, term spread, and default spread as predictive variables  $x$

$$\hat{\gamma}_t = c_0 + c_1'x_{t-1} + v_t$$

- Calculate fitted values

$$\hat{\gamma}_{t-1}^{fit} = \hat{c}_0 + \hat{c}_1'x_{t-1}$$

- Recalculate the contribution numbers as

$$E_{t-1} [R_t] = \hat{\gamma}_0 + E_{t-1}^{\text{beta}} [R_t] + E_{t-1}^{\text{char}} [R_t], \text{ where}$$

$$E_{t-1}^{\text{beta}} [R_t] = \hat{B}_{t-1} \hat{\gamma}_{1t-1}^{fit}, \text{ and } E_{t-1}^{\text{char}} [R_t] = Zcs_{t-1} \hat{\gamma}_{2t-1}^{fit}$$

# Four-factor

	$B_{Mkt}$	$B_{SMB}$	$B_{HML}$	$B_{MOM}$	Sz	B/M	Ret6
Panel C.1: Conditional betas ( $zts = Firm$ )							
Cnst	-1.245 (-1.45)	0.555 (1.16)	0.718 (1.49)	1.569 (1.62)	-0.063 (-0.70)	0.323 (2.24)	1.907 (3.20)
Payout	21.717 (1.18)	-21.518 (-2.02)	-11.963 (-1.11)	-18.608 (-1.13)	4.558 (2.13)	-10.237 (-3.09)	0.961 (0.07)
Def	0.324 (0.58)	0.696 (2.31)	-0.094 (-0.27)	-0.479 (-0.80)	-0.110 (-1.84)	0.304 (2.83)	-1.029 (-2.73)
Term	4.425 (0.35)	-5.374 (-0.69)	-2.838 (-0.35)	7.374 (0.67)	-1.739 (-1.12)	5.547 (2.25)	5.520 (0.66)
Adj- $R^2$	0.12	0.33	0.07	0.32	0.54	1.94	1.21

% Betas = 42.3, % Chars = 47.4, % Diff = 5.1 (-9.6, 49.7)

	$B_{Mkt}$	$B_{SMB}$	$B_{HML}$	$B_{MOM}$	Sz	B/M	Ret6
Panel C.2: Conditional betas ( $zts = Firm + Macro$ )							
Cnst	-1.128 (-1.62)	0.373 (0.82)	0.308 (0.68)	1.625 (1.93)	-0.158 (-2.08)	0.533 (3.77)	1.473 (2.59)
Payout	20.282 (1.24)	-19.560 (-1.96)	-10.058 (-0.97)	-15.151 (-0.95)	5.671 (3.08)	-8.887 (-2.61)	7.233 (0.55)
Def	0.209 (0.44)	0.813 (2.94)	0.224 (0.74)	-0.660 (-1.32)	-0.062 (-1.29)	0.020 (0.20)	-1.022 (-2.54)
Term	7.086 (0.60)	-7.554 (-1.02)	-2.270 (-0.30)	3.793 (0.33)	-2.303 (-1.70)	6.831 (2.75)	11.201 (1.23)
Adj- $R^2$	0.29	0.59	-0.23	0.65	1.37	1.44	0.92

% Betas = 50.2, % Chars = 43.4, % Diff = -6.8 (-23.8, 35.7)

# Conclusion

---

- Risk premium on SMB strong but that on HML weak
- Reject all factor models
  - Rejection not news
- Both betas and characteristics matter
- Characteristics often more important than betas
  - However, four-factor model conditional bias-corrected betas explain more of returns than characteristics with time-varying risk premia

# Next steps

---

## Potential problems

- Use of full-sample betas
- Data restriction of minimum 5 years

## Limitations

- Conclusions valid only for the factor models and the characteristics analyzed here