On Estimation of Risk Premia in Linear Factor Models

Kewei Hou   Robert Kimmel

Fisher College of Business
The Ohio State University

Fields Institute
24 April 2010

Linear Factor Models

- Linear factor models have a long tradition in financial economics
  - Well grounded in financial theory (mean-variance analysis)

$$\mathrm{E}\left[R_i\right] = R_f + \beta_{i,1}\gamma_1 + \ldots + \beta_{i,N}\gamma_N$$

- One of the quantities of interest is the *risk premium* γ_j associated with a factor
- There is a bewildering variety of econometric techniques for the estimation of the risk premia and testing of the model
  - One- and two-pass regression methodologies (OLS, GLS, etc.)

$$\mathrm{E}\left[R_i\right] - R_f = \hat{\gamma}_0 + \beta_{i,1}\hat{\gamma}_1 + \ldots + \beta_{i,N}\hat{\gamma}_N + \hat{\eta}_i$$

  - GMM (which moments to match?)

Results

- Estimated risk premia (standard errors in parentheses):

Factor	Cross-sectional Regression OLS	Cross-sectional Regression GLS	Sample Average
$\hat{\gamma}_0$	1.95% (0.32%)	1.45% (0.25%)	—
$\hat{\gamma}_{RMRF}$	-1.23% (0.31%)	-0.76% (0.24%)	0.63% (0.18%)
$\hat{\gamma}_{SMB}$	0.20% (0.05%)	0.32% (0.02%)	0.30% (0.11%)
$\hat{\gamma}_{HML}$	0.44% (0.07%)	0.40% (0.02%)	0.44% (0.12%)

- Different methods roughly agree on $\hat{\gamma}_{SMB}$ and $\hat{\gamma}_{HML}$, but huge differences for $\hat{\gamma}_{RMRF}$

Unspanned Factors

- The Fama-French factors RMRF, SMB, and HML are all traded, but in typical procedures, are treated as untraded
- To measure the size of unspanned components, regress

$$F_j = \alpha_j + \beta_{j,1} R_1 + \ldots + \beta_{j,25} R_{25} + \eta_j$$

- The R-squared statistics in these regressions are

Factor	RMRF	SMB	HML
R^2 (constant)	0.9922	0.9764	0.9613
R^2 (no constant)	0.9923	0.9766	0.9616

- Also, the (in-sample) monthly Sharpe ratio available with the 25 assets is 0.3216; with the three factors added, the Sharpe ratio increases by less than 4% to 0.3338
- The unspanned components are small—do they matter?

Redo the Regressions

- Rerun the regressions, but include the factor portfolios as right-hand variables also

Factor	Original Regression OLS	Original Regression GLS	Augmented Regression OLS	Augmented Regression GLS	Sample Average
$\hat{\gamma}_0$	1.95% (0.32%)	1.45% (0.25%)	0.22% (0.13%)	0.01% (0.01%)	—
$\hat{\gamma}_{RMRF}$	-1.23% (0.31%)	-0.76% (0.24%)	0.44% (0.13%)	0.62% (0.01%)	0.63% (0.18%)
$\hat{\gamma}_{SMB}$	0.20% (0.05%)	0.32% (0.02%)	0.16% (0.07%)	0.29% (0.01%)	0.30% (0.11%)
$\hat{\gamma}_{HML}$	0.44% (0.07%)	0.40% (0.02%)	0.40% (0.09%)	0.42% (0.01%)	0.44% (0.12%)

- Small effects on $\hat{\gamma}_{HML}$ and $\hat{\gamma}_{SMB}$, huge effects on $\hat{\gamma}_{RMRF}$ and $\hat{\gamma}_0$

Model Selection Example

- Same test assets, but consider two different two-factor models (standard errors in parentheses):

Factor	First Model OLS	First Model GLS	Second Model OLS	Second Model GLS
$\hat{\gamma}_0$	0.00% (0.00%)	0.00% (0.00%)	0.00% (0.00%)	0.00% (0.00%)
$\hat{\gamma}_{F_1}$	2.60% (0.00%)	2.60% (0.00%)	1.53% (0.00%)	1.53% (0.00%)
$\hat{\gamma}_{F_2}$	0.93% (0.00%)	0.93% (0.00%)	0.00% (0.00%)	0.00% (0.00%)

- Both models explain all expected returns *perfectly*
  - In first model, both factors have a significant risk premium
  - In the second model, only the first factor does

Drop the Second Factor

- What happens when the second factor is dropped?

Factor	First Model OLS	First Model GLS	Second Model OLS	Second Model GLS
$\hat{\gamma}_0$	0.00% (0.00%)	0.00% (0.00%)	– 0.18% (0.09%)	– 0.15% (0.20%)
$\hat{\gamma}_{F_1}$	2.60% (0.00%)	2.60% (0.00%)	1.23% (0.10%)	1.63% (0.28%)

- First model—second factor had a positive risk premium
  - But, dropping the second factor had *no* effect on the model's ability to fit returns
- Second model—second factor has zero risk premium
  - But the second factor helps to fit expected returns!

Risk Premia

- The risk premia in a correctly specified model are unique (subject to technical restrictions, no multicollinearity, etc.):

$$\gamma = \Sigma_{FF} \left(\Sigma_{FZ} \Sigma_{ZZ}^{-1} \Sigma_{ZF} \right)^{-1} \Sigma_{FZ} \Sigma_{ZZ}^{-1} \mu_{Z}$$

- Basic Results on Spanned Factors in Correct Models:
  - The risk premium of a spanned factor is the (excess) return of the factor mimicking portfolio
  - The risk premium does not depend on the other factors in the model
  - The risk premium does not depend on which test assets are used (as long as they span the factor)

- Assume spanned factors. From the time-series regressions:
$$\bar{Z} = \hat{\alpha} + \hat{\beta}^T \bar{F}$$
- In a cross-sectional regression, the "X" variables are the β coefficients from the time-series regression

$$\begin{aligned}\hat{\gamma} &= \left(\hat{\beta}\Omega^{-1}\hat{\beta}^T\right)^{-1}\hat{\beta}\Omega^{-1}\bar{Z} \\ &= \underbrace{\bar{F}}_{\text{Risk premia of factor portfolios}} + \underbrace{\left(\hat{\beta}\Omega^{-1}\hat{\beta}^T\right)^{-1}\hat{\beta}\Omega^{-1}\hat{\alpha}}_{\text{Misattribution component}}\end{aligned}$$

- $\hat{\alpha}$ is due to sampling variation and/or misspecification
- The cross-sectional regression produces $\hat{\gamma}$ different from $\bar{F}$ if the sampling variation and/or model misspecification is cross-sectionally correlated with the β coefficients
    - Overfitting by construction

Decomposition of Unspanned Factors

- Expected returns predicted by a model are:

$$\mu_Z = \beta_F^T \gamma_F$$

- Write a set of unspanned factors in terms of spanned and unspanned components

$$F = P + \eta$$

- Consider a model with the spanned components in place of the unspanned factors

$$\mu_Z = \beta_P^T \gamma_P$$

- Note that:

$$\Sigma_{FF} = \Sigma_{PP} + \Sigma_{\eta\eta} \qquad \Sigma_{ZF} = \Sigma_{ZP}$$

- How are the two models to make the same predictions? (Good idea?)

Extrapolation

- Set expected returns of the two models equal to each other:

$$
\begin{aligned}
\beta_F^T \gamma_F &= \beta_P^T \gamma_P \\
\Sigma_{ZF}\Sigma_{FF}^{-1}\gamma_F &= \Sigma_{ZP}\Sigma_{PP}^{-1}\gamma_P \\
\Sigma_{FF}^{-1}\gamma_F &= \Sigma_{PP}^{-1}\gamma_P \\
\gamma_F &= \Sigma_{FF}\Sigma_{PP}^{-1}\gamma_P \\
\gamma_F &= \gamma_P + \underbrace{\Sigma_{\eta\eta}\Sigma_{PP}^{-1}}_{\text{Extrapolation Matrix}} \gamma_P
\end{aligned}
$$

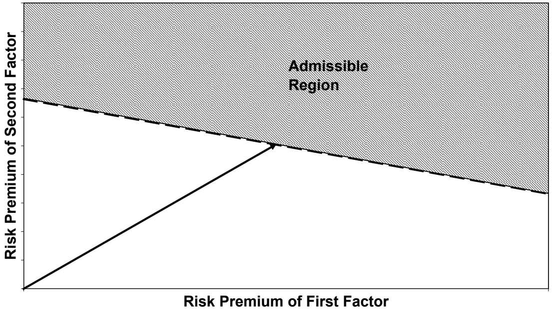
- The risk premia of unspanned factors are the risk premia of their factor mimicking portfolios, plus an extrapolation component that depends on the "noise"
  - The extrapolation component can be extreme

How Bad Can It Be?

- Consider a model with spanned factors
  - The risk premia are the excess returns of the factor mimicking portfolios
- Is it possible, by adding noise to the factors, to change the risk premia to any desired values?
  - No, there isn't that much flexibility
  - But there is a lot of flexibility
  - By adding noise, the risk premia can occupy any point within a half space
- If a model has unspanned factors and is slightly misspecified, what happens when the model is completed?
  - Suppose we find the missing factor and include it in the model
  - The missing factor is not unique—there are many choices that complete the model
  - Depending on the choice, the risk premia of the unspanned factors already in the model can take *any* value at all

Two Factor Example

Admissible Region of Risk Premia Vectors for Equivalence Class of Linear Factor Models



Example—Fama-French Factors

- Regress each factor on the others (standard errors in parentheses):

Regression of	on			
	α	RMRF	SMB	HML
RMRF	0.3472 (0.1629)	—	0.5202 (0.0485)	0.2923 (0.0449)
SMB	0.1545 (0.1037)	0.2102 (0.0196)	—	0.0277 (0.0292)
HML	0.3336 (0.1157)	0.1480 (0.0228)	0.0347 (0.0366)	—

- It is SMB, not RMRF, for which there is no evidence of an independent risk premium (at 95% confidence level)

Conclusions

- The risk premium of a spanned factor is just the excess return of the factor mimicking portfolio
  - A procedure that produces something different than this result is likely engaging in overfitting
- The risk premium of an unspanned factor is not a well-defined concept
  - Result of extrapolation of spanned components to unspanned components
  - Risk premium of spanned component is invariant to changes in other factors, test assets, etc.
  - Useful to report risk premium of spanned component, instead of/in addition to risk premium of factor itself
- Model selection
  - Significance of risk premium and importance of a factor are two different concepts
  - Check significance of α in regression of a factor on the other factors