Observable versus Latent Factors

Viktor Todorov

Based on joint work with Yuan Liao

Venice Time Series Workshop

April 22, 2026

Motivation

We are interested in observable vs. latent (risk) factors:

- Latent factors
  - generate co-movements in cross-section of asset prices
  - can be recovered (up to rotation) via PCA
  - are the sources of priced risk
- Observable factors
  - object of economic interest
  - guided by economic models
  - examples: consumption growth, market volatility,...

Motivation

How are observable and latent risk factors related?

- observable $=$ latent
  - can do asset pricing with observables
  - but very strong assumption
  - we need to know all sources of priced risk
- observable $\subseteq$ linear span (latent) $+$ "noise"
  - can still perform standard linear asset pricing
  - using latent factors and linear projection of observables

Motivation

Observable $\subseteq$ linear span (latent) $+$ "noise"

- "noise" can capture
  - measurement error
  - risk in observable factors not priced in cross-section of stocks
- linear relation between factors
  - holds exactly in linear models
  - holds approximately for nonlinear models over short time span

Motivation

Goal: Test observable $\subseteq$ linear span (latent) $+$ "noise" hypothesis

- If null is true $\Longrightarrow$ linear asset pricing works
- If null is not true $\Longrightarrow$ linear asset pricing does not work
  - nonlinearity can have pricing effects
- Test conducted for a fixed sampling frequency
  - null should hold for high frequency
  - how high should that be?

Outline

- The Continuous-time Factor Model
- The Null and Alternative Hypotheses
- A Test for Linear Relation between Observable and Latent Factors
- Empirical Application:

  The Market Volatility Factor and the Cross-Section of Stock Prices

The Continuous-time Factor Model

The $p \times 1$ price vector Y_t follows:

$$dY_t = \alpha_t dt + \beta df_t + dJ_t + d\epsilon_t$$

where

- systematic diffusive risk:

$$df_t = \Lambda_t dW_t, \quad W_t \text{ is } K \times 1$$

- idiosyncratic diffusive risk:

$$d\epsilon_{t,j} = \sigma_{t,j} dW_{t,j}, \quad j = 1, ..., p$$

The Continuous-time Factor Model

Comments:

- factors f_t are latent and not observed
- only strong factors are considered (extended later)
- J_t is jump part which is not analyzed/modeled
- factor loadings are assumed constant
  - discrete changes in factor loadings allowed by augmenting factor space

The Continuous-time Factor Model

Above setup nests model with time-varying factor loadings:

$$dY_t = \alpha_t dt + \tilde{\beta}_t d\tilde{f}_t + dJ_t + d\epsilon_t$$

where

$$\tilde{\beta}_t = \tilde{\beta}_0 + \sum_{i=1}^{m} (\tilde{\beta}_{t_i} - \tilde{\beta}_{t_{i-1}}) 1(t < t_i)$$

The Continuous-time Factor Model

We can alternatively write:

$$\tilde{\beta}_t d\tilde{f}_t = \beta df_t, \text{ for } t \in [0, T]$$

where

$$\beta = \left(\ \tilde{\beta}_0, \tilde{\beta}_{t_1} - \tilde{\beta}_{t_0}, \cdots, \tilde{\beta}_{t_m} - \tilde{\beta}_{t_{m-1}}\ \right), \quad f_t = \begin{pmatrix} \tilde{f}_t \\ \tilde{f}_t 1(t \in (t_1, t_2]) \\ \vdots \\ \tilde{f}_t 1(t \in (t_{m-1}, t_m]) \end{pmatrix}$$

The Continuous-time Factor Model

Note:

- static factor model representation is of higher dimension
- static factors have special structure: they are present on part of the interval only
- equivalent to dormant factors (factors active on part of the interval only)
- example: factors in morning trading hours different from those in afternoon trading hours

Observable Factor

We denote observable risk factor with g_t (tradable or non-tradable)

- factor is univariate (this can be generalized)
- our focus on its continuous component g_t^c

Our goal: study relation between g_t and f_t

Observable Factor

We decompose g_t^c:

$$g_t^c = \gamma_{t,0} + \int_0^t \gamma_s^\top df_s + \epsilon_t^g$$

- $\gamma_{t,0}$ is of finite variation
- ϵ_t^g is a martingale orthogonal to f:

$$\langle f, \epsilon^g \rangle_t = 0, \;\; t \in [0, T]$$

Observable Factor

This decomposition can always be made:

$$\gamma_t = (c_t^{ff})^{-1} c_t^{fg}$$

where

$$c_t^{ff} = \frac{d\langle f, f\rangle_t}{dt} \text{ and } c_t^{fg} = \frac{d\langle f, g\rangle_t}{dt}$$

- γ_t is local linear projection coefficient of g on f
- need generalized inverse to cover dormant factors at time t

Observable Factor

Residual ϵ_t^g does not matter for asset pricing:

- if latent factors capture all priced risk
- pricing kernel depends on f only
- and we have
$$\mathbb{E}_t(f_{t+\Delta}\epsilon^g_{t+\Delta}) = 0, \text{ for } \Delta > 0$$

The Null Hypothesis

We are interested in the following hypothesis

$$\mathbb{H}_0 : \gamma_t = \gamma_0, \; \forall t \in [0, T]$$

Note:

- test is for fixed time interval only
- T can be arbitrary small

Formulation of the Test

How to test null?

- can estimate locally γ_t from shrinking blocks of returns
- and test for temporal variation in estimated γ_t
- but some of the factors might be inactive on parts of $[0, T]$
- and this approach is sensitive to choice of block length

The Null Hypothesis

We formulate null using global linear projection of g on f:

$$g_t^c = \gamma_{t,0} + \int_0^t \gamma^\top df_t + u_t$$

where the global linear projection coefficient γ is

$$\gamma = \left(\int_0^T c_t^{ff} dt \right)^{-1} \int_0^T c_t^{fg} dt$$

The Null Hypothesis

γ is a volatility-weighted average of γ_t:

$$\gamma = \int_0^T \omega_t \gamma_t dt$$

where

$$\omega_t = \left(\int_0^T c_t^{ff} dt \right)^{-1} c_t^{ff}$$

The Null Hypothesis

The residual term u from the global projection:

- satisfies $\langle f, u \rangle_T = 0$
- but we might have $\langle f, u \rangle_t \neq 0$, for $t < T$
- f and u can be locally correlated
- $\Longrightarrow u$ can be priced:

$$\mathbb{E}_t(f_{t+\Delta} u_{t+\Delta}) \neq 0, \text{ for some } \Delta > 0$$

Formulation of the Test

The null hypothesis:

$$\mathbb{H}_0 : \langle f, u \rangle_t = 0, \ \forall t \in [0, T]$$

This means:

$$\mathrm{cov}_t(\Delta f, \Delta u) \approx 0, \quad \forall t \in [0, T]$$

Formulation of the Test

The null hypothesis is equivalent to:

$$\int_0^T \mathcal{W}_\eta(t/T)d\langle f, u\rangle_t = 0$$

where

$$\mathcal{W}_\eta(x) = \cos(\eta x) + \sin(\eta x),$$

for $\eta \in [0, 2\pi]$ and $x \in [0, 1]$

Feasible Implementation

- test is built from high-frequency observations of Y and g
- sampling at times:
$$0, \frac{1}{n}, \frac{2}{n}, .., T$$
- asymptotics is for T fixed, $n \to \infty$ and $p \to \infty$

Feasible Implementation

- Estimate latent factor returns via PCA: $\widehat{f}_t$
- Run a linear regression of g_t on $\widehat{f}_t$
- Denote residual from regression with $\widehat{u}_t$
- Form statistic

$$\widehat{\mathbb{S}}_k(\eta) = \frac{1}{n} \sum_{t=1}^{n} \widehat{f}_{t,k} \widehat{u}_t \mathcal{W}_\eta(t/n), \ k = 1, ..., K,$$

$$\widehat{\mathbb{S}}(\overline{\eta}) = (\widehat{\mathbb{S}}_k(\eta) : k = 1, ..., K), \ \overline{\eta} = (\eta_1, ..., \eta_K)$$

Feasible Implementation

Latent factor returns identified up to rotation:

$$\widehat{f}_t \xrightarrow{\mathbb{P}} \overline{H} \times f_t,$$

where $\overline{H}$ is random and invertible matrix

This has no effect on test:

$$\langle f, u \rangle_t = 0 \;\Leftrightarrow\; \overline{H} \times \langle f, u \rangle_t = 0$$

Feasible Implementation

$$\text{under } \mathbb{H}_0 : \sup_{\overline{\eta}} |\widehat{\mathbb{S}}(\overline{\eta})| \ \xrightarrow{\mathbb{P}} \ 0$$

versus

$$\text{under } \mathbb{H}_A : \sup_{\overline{\eta}} |\widehat{\mathbb{S}}(\overline{\eta})| \ \xrightarrow{\mathbb{P}} \ K > 0$$

CLT for Test Statistic

Let $n, p \to \infty$ with

$$\sqrt{n}/p \to 0$$

Under $\mathbb{H}_0$ and some assumptions:

$$\sqrt{n}\,\widehat{\mathbb{S}}(\overline{\eta}) \xrightarrow{\mathcal{L}-s} \mathbb{Z}_K(\overline{\eta})$$

uniformly in $\overline{\eta}$ and for some $\mathcal{F}$-conditional zero-mean Gaussian process

Test Statistic under Alternative

Under alternative:

$$\widehat{\mathbb{S}}(\bar{\eta}) \xrightarrow{\mathbb{P}} \mathbb{S}(\bar{\eta})$$

where

$$\mathbb{S}(\bar{\eta}) = \frac{1}{T}\int_0^T \mathrm{diag}\left(w_{\eta_1}(t/T), ..., w_{\eta_K}(t/T)\right)\overline{H}'\, d\langle f, u\rangle_t$$

and if $\langle f, u\rangle_t$ nonzero on $[0, T]$:

$$\sup_{\overline{\eta}} |\mathbb{S}(\overline{\eta})| > K$$

Critical Values

- Limit distribution is non-standard, implement wild bootstrap:

$$g_t^* = \widehat{\gamma}_0 + \widehat{\gamma}^\top \widehat{f}_t + \eta_t^* \widehat{u}_t,$$

  for η_t^* being iid standard normals

- Regress g_t^*on $\widehat{f}_t$ and get $\widehat{u}_t^*$

- Compute critical values from $\mathcal{F}$-conditional distribution of

$$\widehat{\mathbb{S}}_k^*(\eta) = \frac{1}{n}\sum_{t=1}^{n} \widehat{f}_{t,k}\widehat{u}_t^* \mathcal{W}_\eta(t/n)$$

Critical Values

We have:

$$\sqrt{n}\,\widehat{\mathbb{S}}^*(\overline{\eta}) \xrightarrow{\mathcal{L}|\mathcal{F}} \mathbb{Z}_K(\overline{\eta})$$

where $\mathbb{Z}_K(\overline{\eta})$ has same $\mathcal{F}$-conditional law as that of the limit of the test statistic under null

$\Longrightarrow$ quantiles of $\widehat{\mathbb{S}}^*(\overline{\eta})$ can be used as critical values for test

Extension to Weak Factors

We can extend to weak factors:

- Denote

$$B_p = \mathrm{diag}\{p^{\alpha_1/2}, ..., p^{\alpha_K/2}\}, \quad 0 = \alpha_1 \leq ... \leq \alpha_K < 1/2$$

- and assume

$$\left\| B_p \left(\Sigma_f^{1/2} \frac{1}{p} \beta' \beta \Sigma_f^{1/2} \right) B_p - \bar{D} \right\| = o_P(p^{-\alpha_K}), \text{ as } p \to \infty$$

where $\Sigma_f = \langle f, f \rangle_T$

Extension to Weak Factors

- Case of strong factors: $\alpha_1 = ... = \alpha_K = 0$
- Weak factors = weak factor loadings
- Weakest factor corresponds to $\alpha_K = 1$, we restrict $\alpha_K < 1/2$
- We allow for heterogeneous strength of factors
- Weak factors $\Longrightarrow$ slower rate of convergence

Extension to Weak Factors

Let $n, p \to \infty$ with

$$p^{2\alpha_K}/n \to 0 \text{ and } n/p^{2-2\alpha_K} \to 0$$

Under $\mathbb{H}_0$ and some assumptions:

$$\sqrt{n}\, B_p \widehat{\mathbb{S}}(\overline{\eta}) \xrightarrow{\mathcal{L}-s} \mathbb{Z}_K(\overline{\eta})$$

uniformly in $\overline{\eta}$ and for some $\mathcal{F}$-conditional zero-mean Gaussian process

Extension to Weak Factors

Under alternative:

$$B_p \widehat{\mathbb{S}}(\bar{\eta}) \xrightarrow{\mathbb{P}} \mathbb{S}(\bar{\eta})$$

where

$$\mathbb{S}(\bar{\eta}) = \frac{1}{T}\int_0^T \operatorname{diag}\left(w_{\eta_1}(t/T), ..., w_{\eta_K}(t/T)\right)\overline{H}'\, d\langle f, u\rangle_t$$

Extension to Weak Factors

We have:

$$\sqrt{n}\, B_p \widehat{\mathbb{S}}^*(\overline{\eta}) \xrightarrow{\mathcal{L}|\mathcal{F}} \mathbb{Z}_K(\overline{\eta})$$

Note:

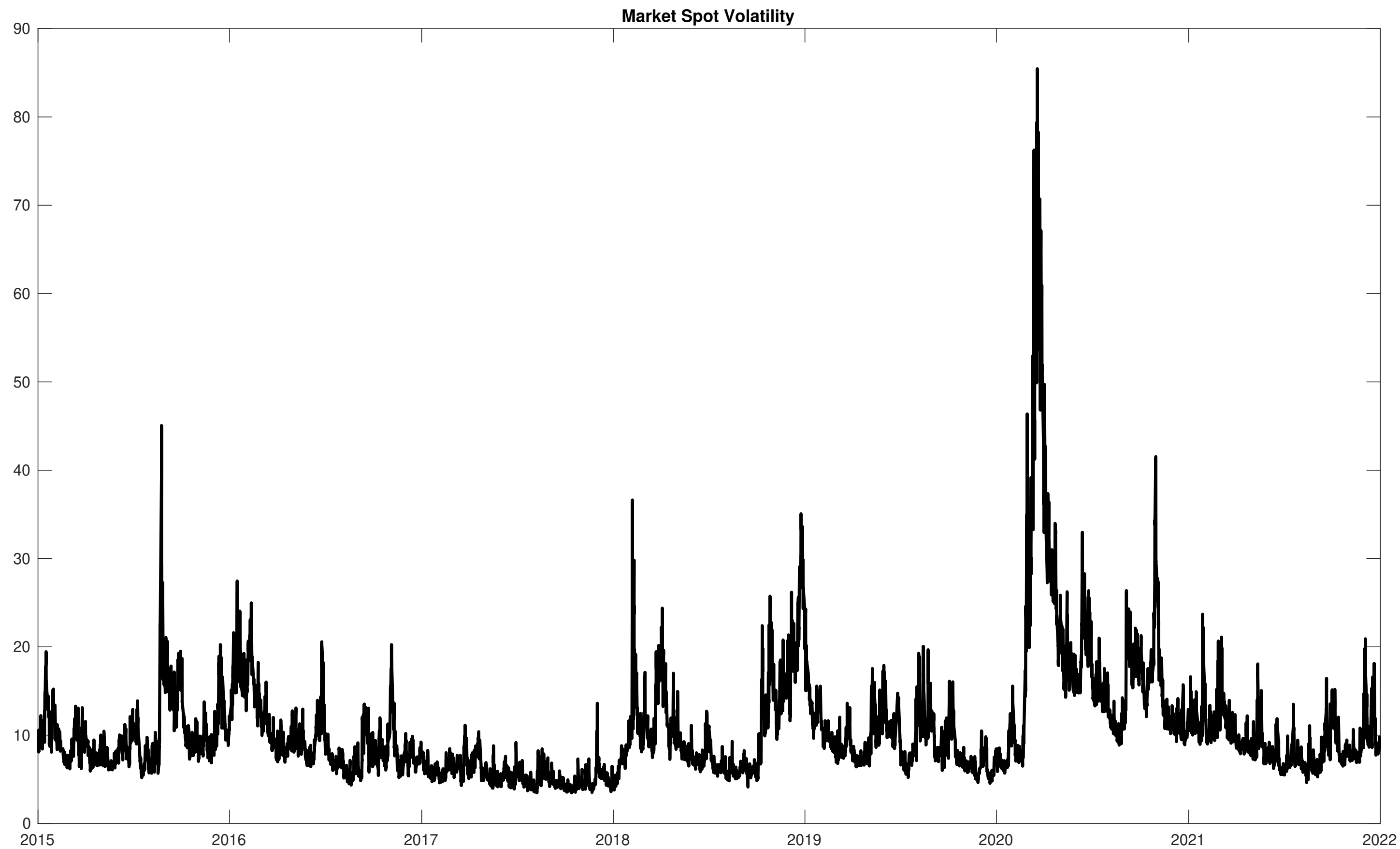
- Econometrician does not know strength of factors (α_k-s)
- For implementation of test, this is not needed
- $\Longrightarrow$ implementation proceeds as with strong factors

Empirical Application: Market Volatility and Cross-Section of Stocks

Data:

- sample period: 2015-2021
- cross-section of 500 largest traded US stocks (determined every year)
- observable factor: market volatility
- sampling frequency: 10-minutes

Market Volatility



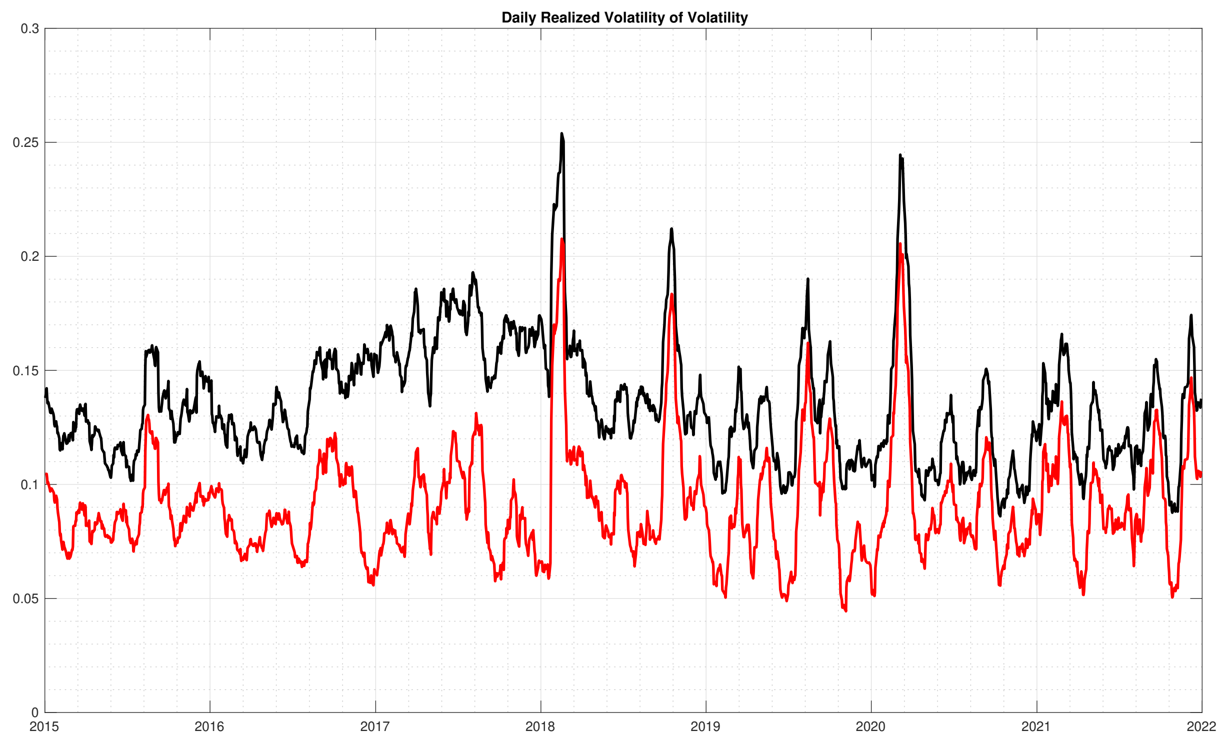
Note: model-free estimate of spot volatility from options, ticker: SPOTVOL

Empirical Test Results for Market Volatility

Time Aggregation	Empirical Rejection Rate of Test	Average R^2 of regression
1 day	0.0956	0.4483
5 days	0.2478	0.4040
10 days	0.3430	0.3994
20 days	0.4643	0.4036
40 days	0.6429	0.4100

The nominal size of the test is 5%.

Volatility of Volatility



Note: black line is total vol-of-vol, red line is part due to projection on $\widehat{f}$

Residual Market Volatility

- Market Volatility strongly correlated with market return (leverage effect)
- Is market volatility "net" of market return a separate risk factor?
- Form residual market volatility:
  - residual from local linear projection of market volatility returns on market returns

Empirical Results for Residual Market Volatility

Time Aggregation	Empirical Rejection Rate of Test	Average R^2 of regression
1 day	0.0845	0.2417
5 days	0.2588	0.1122
10 days	0.3455	0.0841
20 days	0.4286	0.0890
40 days	0.6571	0.0957

The nominal size of the test is 5%.