Testing Conditional Factor Models

Andrew Ang   Dennis Kristensen

Columbia Uni and UCL

November 2011

Motivation

Consider a standard asset pricing factor model:

$$E\left[R_{k,t}\right] = \beta_k' \lambda,$$

where

- $R_{k,t}$ is the excess return of stock k at time t.
- β_k is the vector of factor loadings of stock j w.r.t. a set of factors.
- λ is the vector of risk premia.

There are two ways developed in the literature to test the model:

- Time-series tests.
- Cross-sectional tests.

Today's talk will focus on time-series tests.

Time-series factor model

Assume that excess return of a given asset/porfolio k $(k = 1, ..., M)$ at time t is given by:

$$R_{k,t} = \alpha_k + \beta_k' f_t + \varepsilon_{k,t},$$

where

- $R_{k,t}$ - excess return of the k'th asset/portfolio.
- $f_t \in \mathbb{R}^J$ - vector of J tradeable factors (and so observed).
- $\varepsilon_{k,t}$ - idiosyncratic error satisfying $E\left[\varepsilon_{k,t} | f_t\right] = 0$.
- α_k - intercept.
- $\beta_k \in \mathbb{R}^J$ - vector of J factor loadings.

Time-series factor model

- **Asset pricing hypothesis of factor model:**

$$H_0 : \alpha_k = 0, \quad k = 1, \ldots M.$$

Time-series factor model

- **Asset pricing hypothesis of factor model:**

$$H_0 : \alpha_k = 0, \quad k = 1, ...M.$$

- Usual tests are based on OLS estimates; see e.g. Gibbons, Ross and Shanken (1989) [GRS]:

Time-series factor model

- **Asset pricing hypothesis of factor model:**

$$H_0 : \alpha_k = 0, \quad k = 1, \dots M.$$

- Usual tests are based on OLS estimates; see e.g. Gibbons, Ross and Shanken (1989) [GRS]:
  - Permits joint tests across assets ($k = 1, \dots M$).

Time-series factor model

- **Asset pricing hypothesis of factor model:**

$$H_0 : \alpha_k = 0, \quad k = 1, \ldots M.$$

- Usual tests are based on OLS estimates; see e.g. Gibbons, Ross and Shanken (1989) [GRS]:
  - Permits joint tests across assets ($k = 1, \ldots M$).
  - Sampling variation of the alpha estimate is affected by the sampling variation of the beta estimate.

- **Asset pricing hypothesis of factor model:**

$$H_0 : \alpha_k = 0, \quad k = 1, ...M.$$

- Usual tests are based on OLS estimates; see e.g. Gibbons, Ross and Shanken (1989) [GRS]:
  - Permits joint tests across assets ($k = 1, ...M$).
  - Sampling variation of the alpha estimate is affected by the sampling variation of the beta estimate.
  - F-type tests of H_0 follow χ^2-distributions.

Conditional Factor Model

Suppose the factor loadings, β_k, are time-varying.

- Strong empirical evidence that this is indeed the case even at portfolio level - see e.g. Fama and French (1997), Lewellen and Nagel (2006), Ang and Chen (2006).

Conditional Factor Model

Suppose the factor loadings, β_k, are time-varying.

- Strong empirical evidence that this is indeed the case even at portfolio level - see e.g. Fama and French (1997), Lewellen and Nagel (2006), Ang and Chen (2006).
- Time variation in factor loadings distorts standard GRS-type factor model tests.

Conditional Factor Model

Suppose the factor loadings, β_k, are time-varying.

- Strong empirical evidence that this is indeed the case even at portfolio level - see e.g. Fama and French (1997), Lewellen and Nagel (2006), Ang and Chen (2006).
- Time variation in factor loadings distorts standard GRS-type factor model tests.
- As such, traditional statistical inference for the validity of a factor model is in general misleading.

Conditional Factor Model

In the case with time-varying betas, the following approaches have been taken:

- **Instrument the betas** (Shanken, 1990; Ferson and Harvey, 1991, 1993):

  $$\beta_t = a + B' X_t,$$

  for a set of observed instruments X_t.
  Estimated factor loadings are very sensitive to the choice of X_t and many instruments are only available at coarser frequencies.

Conditional Factor Model

- **Latent variable Model** (Ang and Chen, 2006):

$$\beta_t = a + B'\beta_{t-1} + z_t.$$

Relies on correct specification of the dynamics of the betas; computationally and statistically hard to estimate when dim (β_t) "large".

Conditional Factor Model

- **Latent variable Model** (Ang and Chen, 2006):

$$\beta_t = a + B'\beta_{t-1} + z_t.$$

Relies on correct specification of the dynamics of the betas; computationally and statistically hard to estimate when dim (β_t) "large".

- **Rolling-window estimation** (French, Scwert and Stambaugh, 1987; Andersen et al, 2006; Lewellen and Nagel, 2006):
Estimate β_t by OLS over (small) subsamples.

Conditional Factor Model

- **Latent variable Model** (Ang and Chen, 2006):

$$\beta_t = a + B'\beta_{t-1} + z_t.$$

  Relies on correct specification of the dynamics of the betas; computationally and statistically hard to estimate when dim (β_t) "large".

- **Rolling-window estimation** (French, Scwert and Stambaugh, 1987; Andersen et al, 2006; Lewellen and Nagel, 2006):
  Estimate β_t by OLS over (small) subsamples.
  - What is the correct (optimal) choice of the subsample size (aka window width)?

Conditional Factor Model

- **Latent variable Model** (Ang and Chen, 2006):

$$\beta_t = a + B'\beta_{t-1} + z_t.$$

  Relies on correct specification of the dynamics of the betas; computationally and statistically hard to estimate when dim (β_t) "large".

- **Rolling-window estimation** (French, Scwert and Stambaugh, 1987; Andersen et al, 2006; Lewellen and Nagel, 2006):
  Estimate β_t by OLS over (small) subsamples.
  - What is the correct (optimal) choice of the subsample size (aka window width)?
  - No formal testing procedure of conditional factor model.

Our contributions

- **New class of rolling-window estimators**:
  We develop nonparametric estimators of both conditional alphas and betas given high-frequency data.
  Estimators are on closed form and so simple to implement.
  We also develop estimators of so-called long-run alphas and betas.

Our contributions

- **New class of rolling-window estimators**:
  We develop nonparametric estimators of both conditional alphas and betas given high-frequency data.
  Estimators are on closed form and so simple to implement.
  We also develop estimators of so-called long-run alphas and betas.
- **New tests of asset pricing hypothesis:**
  Given estimators, we propose new tests of H_0 that are robust to time-variation in alphas and betas.
  In the case of constant betas and homoskedasticity, our tests collapse to GRS.

Our contributions

- **New class of rolling-window estimators**:
  We develop nonparametric estimators of both conditional alphas and betas given high-frequency data.
  Estimators are on closed form and so simple to implement.
  We also develop estimators of so-called long-run alphas and betas.
- **New tests of asset pricing hypothesis:**
  Given estimators, we propose new tests of H_0 that are robust to time-variation in alphas and betas.
  In the case of constant betas and homoskedasticity, our tests collapse to GRS.
- **New test for constancy of alphas and betas.**

Our contributions

- **New class of rolling-window estimators**:
  We develop nonparametric estimators of both conditional alphas and betas given high-frequency data.
  Estimators are on closed form and so simple to implement.
  We also develop estimators of so-called long-run alphas and betas.
- **New tests of asset pricing hypothesis:**
  Given estimators, we propose new tests of H_0 that are robust to time-variation in alphas and betas.
  In the case of constant betas and homoskedasticity, our tests collapse to GRS.
- **New test for constancy of alphas and betas.**
- **Inferential Tools:**
  Derive joint distributions of conditional and long-run estimates.
  Derive distributions of test statistics.

Application

- Decile portfolios of stocks sorted on book-to-market ratios and past returns (momentum).

Application

- Decile portfolios of stocks sorted on book-to-market ratios and past returns (momentum).
- Estimate conditional one-factor market model and conditional three-factor Fama-French model.

Application

- Decile portfolios of stocks sorted on book-to-market ratios and past returns (momentum).
- Estimate conditional one-factor market model and conditional three-factor Fama-French model.
- Findings:

Application

- Decile portfolios of stocks sorted on book-to-market ratios and past returns (momentum).
- Estimate conditional one-factor market model and conditional three-factor Fama-French model.
- Findings:
    - Reject that betas are constant.

Application

- Decile portfolios of stocks sorted on book-to-market ratios and past returns (momentum).
- Estimate conditional one-factor market model and conditional three-factor Fama-French model.
- Findings:
  - Reject that betas are constant.
  - Long-run alphas are jointly significantly different from zero for both models and in both sets of portfolios.

Application

- Decile portfolios of stocks sorted on book-to-market ratios and past returns (momentum).
- Estimate conditional one-factor market model and conditional three-factor Fama-French model.
- Findings:
  - Reject that betas are constant.
  - Long-run alphas are jointly significantly different from zero for both models and in both sets of portfolios.
  - Find little evidence that conditional market betas increase during "bad" times.

Discrete-time model

Suppose we have observed assets and factors at n time points in the time interval $[0, T]$,

$$0 < t_1 < t_2 < ... < t_n < T.$$

Data comes from the factor model

$$R_{t_i} = \alpha(t_i) + \beta(t_i)' f_{t_i} + \Omega^{1/2}(t_i) z_{t_i}.$$

- $R_t = (R_{1,t}, ..., R_{M,t})'$ is a vector of M excess returns.
- $f_t = (f_{1,t}, ..., f_{J,t})'$ is a vector of J factors.
- $\alpha(t) = (\alpha_1(t), ..., \alpha_M(t))'$ is a vector of M time-varying intercepts.
- $\beta(t) = (\beta_1(t), ..., \beta_M(t))'$ is a $(J \times M)$-matrix of time-varying factor loadings.
- $\Omega(t)$ is a $(M \times M)$ covariance matrix.
- $z_t = (z_{1,t}, ..., z_{M,t})$ satisfies $E[z_t | f_t] = 0$ and $E[z_t z_t' | f_t] = I_M$.

Hypotheses of interest

- We are interested in testing

$$H_0 : \alpha(t) = 0 \text{ for all } t.$$

Hypotheses of interest

- We are interested in testing

$$H_0 : \alpha(t) = 0 \text{ for all } t.$$

- We are also interested in testing a weaker hypothesis saying that the alphas may be non-zero in the short-run but zero in the long-run.

Hypotheses of interest

- We are interested in testing

$$H_0 : \alpha\left(t\right) = 0 \text{ for all } t.$$

- We are also interested in testing a weaker hypothesis saying that the alphas may be non-zero in the short-run but zero in the long-run.
- Long-run alphas and betas:

$$\alpha_{\mathrm{LR}} \equiv \lim_{n\to\infty} \frac{1}{n} \sum_{i=1}^{n} \alpha(t_i), \quad \beta_{\mathrm{LR}} \equiv \lim_{n\to\infty} \frac{1}{n} \sum_{i=1}^{n} \beta(t_i).$$

Hypotheses of interest

- We are interested in testing

$$H_0 : \alpha(t) = 0 \text{ for all } t.$$

- We are also interested in testing a weaker hypothesis saying that the alphas may be non-zero in the short-run but zero in the long-run.
- Long-run alphas and betas:

$$\alpha_{\mathrm{LR}} \equiv \lim_{n\to\infty} \frac{1}{n} \sum_{i=1}^{n} \alpha(t_i), \quad \beta_{\mathrm{LR}} \equiv \lim_{n\to\infty} \frac{1}{n} \sum_{i=1}^{n} \beta(t_i).$$

- A weaker version of H_0 is then

$$H_{\mathrm{LR}} : \alpha_{\mathrm{LR}} = 0.$$

Conditional estimator

Rewrite model as

$$R_{t_i} = \gamma\left(t_i\right)' X_{t_i} + \Omega^{1/2}\left(t_i\right) z_{t_i},$$

$$\gamma\left(t\right) = \left(\alpha\left(t\right), \beta\left(t\right)\right), \quad X_{t_i} = \left(1, f_{t_i}'\right)'.$$

- **Local OLS:** To obtain a consistent estimator of $\gamma\left(t\right)$ at some given value t, we modify the OLS estimator to only include relevant information.

Conditional estimator

Rewrite model as

$$R_{t_i} = \gamma(t_i)' X_{t_i} + \Omega^{1/2}(t_i) z_{t_i},$$

$$\gamma(t) = (\alpha(t), \beta(t)), \quad X_{t_i} = \left(1, f'_{t_i}\right)'.$$

- **Local OLS:** To obtain a consistent estimator of $\gamma(t)$ at some given value t, we modify the OLS estimator to only include relevant information.
- Suppose that $t \mapsto \gamma(t)$ is slowly varying (continuous). Then observations in a small time window around t will be informative about $\gamma(t)$.

Conditional estimator

Rewrite model as

$$R_{t_i} = \gamma(t_i)' X_{t_i} + \Omega^{1/2}(t_i) z_{t_i},$$
$$\gamma(t) = (\alpha(t), \beta(t)), \quad X_{t_i} = \left(1, f_{t_i}'\right)'.$$

- **Local OLS:** To obtain a consistent estimator of $\gamma(t)$ at some given value t, we modify the OLS estimator to only include relevant information.
- Suppose that $t \mapsto \gamma(t)$ is slowly varying (continuous). Then observations in a small time window around t will be informative about $\gamma(t)$.
- **Kernel-weighted OLS:** For a given time point $t \in [0, T]$,

$$\hat{\gamma}(t) = \left[\sum_{i=1}^{n} K\left(\frac{t_i - t}{hT}\right) X_{t_i} X_{t_i}'\right]^{-1} \left[\sum_{i=1}^{n} K\left(\frac{t_i - t}{hT}\right) X_{t_i} R_{t_i}'\right],$$

where the function K is a *kernel* (density) and $h > 0$ is a *band- or window-width.*

Conditional estimator

$$\hat{\gamma}(t) = \left[\sum_{i=1}^{n} K\left(\frac{t_i - t}{hT} \right) X_{t_i} X_{t_i}' \right]^{-1} \left[\sum_{i=1}^{n} K\left(\frac{t_i - t}{hT} \right) X_{t_i} R_{t_i}' \right].$$

Our estimator is simply a weighted least-squares estimator!

- The kernel K and the bandwidth $h > 0$ jointly determine how much weight should be given to individual observations in the weighted least-squares estimator.

Conditional estimator

$$\hat{\gamma}(t) = \left[\sum_{i=1}^{n} K\left(\frac{t_i - t}{hT}\right) X_{t_i} X_{t_i}'\right]^{-1} \left[\sum_{i=1}^{n} K\left(\frac{t_i - t}{hT}\right) X_{t_i} R_{t_i}'\right].$$

Our estimator is simply a weighted least-squares estimator!

- The kernel K and the bandwidth $h > 0$ jointly determine how much weight should be given to individual observations in the weighted least-squares estimator.
- If K is chosen as the uniform density on $[-1/2, 1/2]$,

$$\hat{\gamma}(t) = \left[\sum_{i:|t_i - t| \leq hT/2} X_{t_i} X_{t_i}'\right]^{-1} \left[\sum_{i:|t_i - t| \leq hT/2} X_{t_i} R_{t_i}'\right].$$

  Thus, our estimator can be seen as a generalization of rolling-window/realized covariance estimators.

Conditional estimator

- **Small bandwidth:** Only observations very close to t are used to estimate $\gamma(t)$. As $h \to 0$,

$$\hat{\gamma}(t) \approx \left[X_t X_t'\right]^{-1} \left[X_t R_t'\right]$$

Conditional estimator

- **Small bandwidth:** Only observations very close to t are used to estimate $\gamma(t)$. As $h \to 0$,

$$\hat{\gamma}(t) \approx \left[X_t X_t'\right]^{-1} \left[X_t R_t'\right]$$

- **Large bandwidth:** All observations are used to estimate $\hat{\gamma}(\tau)$. As $h \to \infty$,

$$\hat{\gamma}(t) \approx \hat{\gamma}_{\text{OLS}}.$$

Long-run estimator

To estimate the long-run alphas and betas, we simply plug in the conditional estimates that we have just proposed:

$$\hat{\alpha}_{\mathrm{LR}} \equiv \frac{1}{n}\sum_{i=1}^{n} \hat{\alpha}(t_i), \quad \hat{\beta}_{\mathrm{LR}} \equiv \frac{1}{n}\sum_{i=1}^{n} \hat{\beta}(t_i).$$

Continuous-time model

For the theoretical analysis of the proposed estimators, we introduce a continuous-time version of the discrete-time factor model:

$$ds(t) = \alpha(t)\,dt + \beta(t)'\,dF(t) + \Sigma^{1/2}(t)\,dB_s(t),$$

$$dF(t) = \mu_F(t)\,dt + \Lambda_{FF}^{1/2}(t)\,dB_F(t).$$

- $s(t)$ - observed M asset prices.
- $F(t)$ - observed J factors.
- $B_s(t)$ and $B_F(t)$ are standard Brownian motions.

This is the ANOVA model considered in Andersen et al. (2006) and Mykland and Zhang (2006).

Continuous-time model

- Let $\Delta \equiv t_i - t_{i-1}$ be the (constant) time distance between the individual observations from the continuous-time version.

Continuous-time model

- Let $\Delta \equiv t_i - t_{i-1}$ be the (constant) time distance between the individual observations from the continuous-time version.
- Defining

$$R_{t_i} = \frac{s(t_i) - s(t_{i-1})}{\Delta}, \quad f_{t_i} = \frac{F(t_i) - F(t_{i-1})}{\Delta},$$

the continuous-time model implies that (as $\Delta \to 0$)

$$R_{t_i} \approx \alpha(t_i) + \beta(t_i)' f_{t_i} + \Omega^{1/2}(t_i) z_{t_i},$$

where $z_{t_i} \sim N(0, I_M)$ and $\Omega(t) = \Sigma(t)/\Delta$.

Continuous-time model

- Let $\Delta \equiv t_i - t_{i-1}$ be the (constant) time distance between the individual observations from the continuous-time version.
- Defining

$$R_{t_i} = \frac{s(t_i) - s(t_{i-1})}{\Delta}, \quad f_{t_i} = \frac{F(t_i) - F(t_{i-1})}{\Delta},$$

  the continuous-time model implies that (as $\Delta \to 0$)

$$R_{t_i} \approx \alpha(t_i) + \beta(t_i)' f_{t_i} + \Omega^{1/2}(t_i) z_{t_i},$$

  where $z_{t_i} \sim N(0, I_M)$ and $\Omega(t) = \Sigma(t)/\Delta$.
- Natural estimators of $\alpha(t)$ and $\beta(t)$ therefore take on the same form as the discrete-time estimators.

Continuous-time model

- Let $\Delta \equiv t_i - t_{i-1}$ be the (constant) time distance between the individual observations from the continuous-time version.
- Defining

$$R_{t_i} = \frac{s(t_i) - s(t_{i-1})}{\Delta}, \quad f_{t_i} = \frac{F(t_i) - F(t_{i-1})}{\Delta},$$

  the continuous-time model implies that (as $\Delta \to 0$)

$$R_{t_i} \approx \alpha(t_i) + \beta(t_i)' f_{t_i} + \Omega^{1/2}(t_i) z_{t_i},$$

  where $z_{t_i} \sim N(0, I_M)$ and $\Omega(t) = \Sigma(t)/\Delta$.
- Natural estimators of $\alpha(t)$ and $\beta(t)$ therefore take on the same form as the discrete-time estimators.
- In particular, $\hat{\beta}(t)$ is simply a localized version of the well-known realized beta estimator considered by Andersen et al (2006) and Mykland and Zhang (2006).

Properties of conditional beta estimator

- Extending the arguments in Kristensen (2010), as $\Delta \to 0$:

$$
\begin{aligned}
E[\hat{\beta}(t)] &\simeq \beta(t) + (hT)^2 \beta^{(2)}(t), \\
\text{Var}(\hat{\beta}(t)) &\simeq \frac{1}{nh} \times \kappa_2 \Lambda_{FF}^{-1}(t) \otimes \Sigma(t),
\end{aligned}
$$

where $\beta^{(2)}(t) =$ 2nd derivative of $\beta(t)$ and $\kappa_2 = \int K^2(z)\, dz$.

Properties of conditional beta estimator

- Extending the arguments in Kristensen (2010), as $\Delta \to 0$:

$$
\begin{aligned}
E[\hat{\beta}(t)] &\simeq \beta(t) + (hT)^2 \beta^{(2)}(t), \\
\text{Var}(\hat{\beta}(t)) &\simeq \frac{1}{nh} \times \kappa_2 \Lambda_{FF}^{-1}(t) \otimes \Sigma(t),
\end{aligned}
$$

  where $\beta^{(2)}(t) =$ 2nd derivative of $\beta(t)$ and $\kappa_2 = \int K^2(z)\, dz$.

- In particular, as $nh \to \infty$ and $nT^4 h^5 \to 0$:

$$
\sqrt{nh}\{\hat{\beta}(t) - \beta(t)\} \sim N\left(0, \kappa_2 \Lambda_{FF}^{-1}(t) \otimes \Sigma(t)\right) \quad \text{in large samples.}
$$

Properties of conditional beta estimator

- Extending the arguments in Kristensen (2010), as $\Delta \to 0$:

$$
\begin{aligned}
E[\hat{\beta}(t)] &\simeq \beta(t) + (hT)^2 \beta^{(2)}(t), \\
\mathrm{Var}(\hat{\beta}(t)) &\simeq \frac{1}{nh} \times \kappa_2 \Lambda_{FF}^{-1}(t) \otimes \Sigma(t),
\end{aligned}
$$

  where $\beta^{(2)}(t) =$ 2nd derivative of $\beta(t)$ and $\kappa_2 = \int K^2(z)\,dz$.
- In particular, as $nh \to \infty$ and $nT^4h^5 \to 0$:

$$
\sqrt{nh}\{\hat{\beta}(t) - \beta(t)\} \sim N\left(0, \kappa_2 \Lambda_{FF}^{-1}(t) \otimes \Sigma(t)\right) \quad \text{in large samples.}
$$

- Slower rate of convergence than parametric estimators: $\sqrt{nh}$ versus $\sqrt{n}$. Do not need $T \to \infty$.

Properties of conditional beta estimator

- Extending the arguments in Kristensen (2010), as $\Delta \to 0$:

$$E[\hat{\beta}(t)] \simeq \beta(t) + (hT)^2 \beta^{(2)}(t),$$
$$\text{Var}(\hat{\beta}(t)) \simeq \frac{1}{nh} \times \kappa_2 \Lambda_{FF}^{-1}(t) \otimes \Sigma(t),$$

  where $\beta^{(2)}(t) =$ 2nd derivative of $\beta(t)$ and $\kappa_2 = \int K^2(z)\, dz$.
- In particular, as $nh \to \infty$ and $nT^4h^5 \to 0$:

$$\sqrt{nh}\{\hat{\beta}(t) - \beta(t)\} \sim N\left(0, \kappa_2 \Lambda_{FF}^{-1}(t) \otimes \Sigma(t)\right) \quad \text{in large samples.}$$

- Slower rate of convergence than parametric estimators: $\sqrt{nh}$ versus $\sqrt{n}$. Do not need $T \to \infty$.
- Properties are similar to those of other nonparametric estimators in diffusion models; see e.g. Bandi and Phillips (2003), Kanaya and Kristensen (2010), and Kristensen (2010).

Properties of conditional alpha estimator

- We show that as $\Delta \to 0$:

$$E\left[\hat{\alpha}(t)\right] \simeq \alpha(t) + (Th)^2 \alpha^{(2)}(t), \quad \text{Var}\left(\hat{\alpha}(t)\right) \simeq \frac{1}{Th} \times \kappa_2 \Sigma(t).$$

Properties of conditional alpha estimator

- We show that as $\Delta \to 0$:

$$E\left[\hat{\alpha}\left(t\right)\right] \simeq \alpha\left(t\right) + \left(Th\right)^{2}\alpha^{(2)}\left(t\right), \quad \mathrm{Var}\left(\hat{\alpha}\left(t\right)\right) \simeq \frac{1}{Th} \times \kappa_{2}\Sigma\left(t\right).$$

- Bias is of same order as for $\hat{\beta}(t)$, but variance vanishes slower, $1/(Th)$ versus $1/(nh)$.

Properties of conditional alpha estimator

- We show that as $\Delta \to 0$:

$$E\left[\hat{\alpha}\left(t\right)\right] \simeq \alpha\left(t\right) + \left(Th\right)^{2} \alpha^{(2)}\left(t\right), \quad \mathrm{Var}\left(\hat{\alpha}\left(t\right)\right) \simeq \frac{1}{Th} \times \kappa_2 \Sigma\left(t\right).$$

- Bias is of same order as for $\hat{\beta}(t)$, but variance vanishes slower, $1/(Th)$ versus $1/(nh)$.
- The slower rate of convergence of $\mathrm{Var}(\hat{\alpha}(t))$ is a well-known feature of nonparametric drift estimators in diffusion models, as in Bandi and Phillips (2003), and is due to the smaller amount of information regarding the drift relative to the volatility found in data.

Properties of conditional alpha estimator

- Bias and variance of $\hat{\alpha}(t)$ are perfectly balanced. To remove the bias, we have to let $Th \rightarrow 0$, but with this bandwidth choice the variance explodes.

Properties of conditional alpha estimator

- Bias and variance of $\hat{\alpha}(t)$ are perfectly balanced. To remove the bias, we have to let $Th \to 0$, but with this bandwidth choice the variance explodes.
- Consequence: Not possible to state formal results regarding the asymptotic distribution of $\hat{\alpha}(t)$. However, informally, with h chosen "small enough" such that the bias is negiglible, we have

$$\sqrt{Th}\{\hat{\alpha}(t) - \alpha(t)\} \sim N(0, \kappa_2 \Sigma(t)) \quad \text{in large samples.}$$

Properties of conditional alpha estimator

- Bias and variance of $\hat{\alpha}(t)$ are perfectly balanced. To remove the bias, we have to let $Th \to 0$, but with this bandwidth choice the variance explodes.
- Consequence: Not possible to state formal results regarding the asymptotic distribution of $\hat{\alpha}(t)$. However, informally, with h chosen "small enough" such that the bias is negiglible, we have

$$\sqrt{Th}\{\hat{\alpha}(t) - \alpha(t)\} \sim N(0, \kappa_2 \Sigma(t)) \quad \text{in large samples.}$$

- To formalize the above statement, one can impose that

$$\alpha(t) = a(t/T) \quad \text{and} \quad \Sigma(t) = S(t/T).$$

This is similar to the time normalization used in the analysis of break-point estimators.

Properties of estimators of LR versions

- While it is in general not possible to consistently estimate conditional (short-run) alphas, we can still estimate the long-run (LR) versions without any time normalization.

Properties of estimators of LR versions

- While it is in general not possible to consistently estimate conditional (short-run) alphas, we can still estimate the long-run (LR) versions without any time normalization.
- We show that, as $h \to 0$ at a suitable rate:

$$\sqrt{T}(\hat{\alpha}_{\mathrm{LR}} - \alpha_{\mathrm{LR}}) \sim N\left(0, \Sigma_{\mathrm{LR},\alpha\alpha}\right), \quad \sqrt{n}(\hat{\beta}_{\mathrm{LR}} - \beta_{\mathrm{LR}}) \sim N\left(0, \Sigma_{\mathrm{LR},\beta\beta}\right).$$

Properties of estimators of LR versions

- While it is in general not possible to consistently estimate conditional (short-run) alphas, we can still estimate the long-run (LR) versions without any time normalization.
- We show that, as $h \to 0$ at a suitable rate:

$$\sqrt{T}(\hat{\alpha}_{\mathrm{LR}} - \alpha_{\mathrm{LR}}) \sim N\left(0, \Sigma_{\mathrm{LR},\alpha\alpha}\right), \quad \sqrt{n}(\hat{\beta}_{\mathrm{LR}} - \beta_{\mathrm{LR}}) \sim N\left(0, \Sigma_{\mathrm{LR},\beta\beta}\right).$$

- LR estimators converge at standard parametric rates, $\sqrt{n}$ and $\sqrt{T}$ respectively. This is due to the additional smoothing taking place when we average over the preliminary short-run estimates.

Properties of estimators of LR versions

- While it is in general not possible to consistently estimate conditional (short-run) alphas, we can still estimate the long-run (LR) versions without any time normalization.
- We show that, as $h \to 0$ at a suitable rate:

$$\sqrt{T}(\hat{\alpha}_{\mathrm{LR}} - \alpha_{\mathrm{LR}}) \sim N\left(0, \Sigma_{\mathrm{LR},\alpha\alpha}\right), \quad \sqrt{n}(\hat{\beta}_{\mathrm{LR}} - \beta_{\mathrm{LR}}) \sim N\left(0, \Sigma_{\mathrm{LR},\beta\beta}\right).$$

- LR estimators converge at standard parametric rates, $\sqrt{n}$ and $\sqrt{T}$ respectively. This is due to the additional smoothing taking place when we average over the preliminary short-run estimates.
- We can test $H_{\mathrm{LR}} : \alpha_{\mathrm{LR}} = 0$ by the following Wald-type statistic:

$$W_{\mathrm{LR}} = T\hat{\alpha}'_{\mathrm{LR}}\hat{\Sigma}^{-1}_{\mathrm{LR},\alpha\alpha}\hat{\alpha}_{\mathrm{LR}} \sim \chi^2_M \quad \text{in large samples.}$$

Testing for constant alphas and betas

$$H_k(\alpha) \quad : \quad \alpha_k(t) = \alpha_k \in \mathbb{R}, \quad \text{for all } t \in [0, T],$$
$$H_k(\beta) \quad : \quad \beta_k(t) = \beta_k \in \mathbb{R}^J, \quad \text{for all } t \in [0, T].$$

- Under either hypothesis, the corresponding LR estimator is a consistent.

Testing for constant alphas and betas

$$H_k(\alpha) \;:\; \alpha_k(t) = \alpha_k \in \mathbb{R}, \quad \text{for all } t \in [0, T],$$
$$H_k(\beta) \;:\; \beta_k(t) = \beta_k \in \mathbb{R}^J, \quad \text{for all } t \in [0, T].$$

- Under either hypothesis, the corresponding LR estimator is a consistent.
- So a natural way to test the two hypotheses is by comparing the LR and SR estimators:

$$W_k(\alpha) \equiv \frac{1}{n}\sum_{i=1}^{n} \hat{\sigma}_{kk}^{-2}(t_i)\left[\hat{\alpha}_k(t_i) - \hat{\alpha}_{\mathrm{LR},k}\right]^2,$$
$$W_k(\beta) \equiv \frac{1}{n}\sum_{i=1}^{n} \hat{\sigma}_{kk}^{-2}(t_i)\left[\hat{\beta}_k(t_i) - \hat{\beta}_{\mathrm{LR},k}\right]' \hat{\Lambda}_{FF}(t_i)\left[\beta_k(t_i) - \hat{\beta}_{\mathrm{LR},k}\right].$$

Testing for constant alphas and betas

$$H_k(\alpha) \;:\; \alpha_k(t) = \alpha_k \in \mathbb{R}, \quad \text{for all } t \in [0, T],$$
$$H_k(\beta) \;:\; \beta_k(t) = \beta_k \in \mathbb{R}^J, \quad \text{for all } t \in [0, T].$$

- Under either hypothesis, the corresponding LR estimator is a consistent.
- So a natural way to test the two hypotheses is by comparing the LR and SR estimators:

$$W_k(\alpha) \equiv \frac{1}{n}\sum_{i=1}^{n} \hat{\sigma}_{kk}^{-2}(t_i)\left[\hat{\alpha}_k(t_i) - \hat{\alpha}_{\mathrm{LR},k}\right]^2,$$
$$W_k(\beta) \equiv \frac{1}{n}\sum_{i=1}^{n} \hat{\sigma}_{kk}^{-2}(t_i)\left[\hat{\beta}_k(t_i) - \hat{\beta}_{\mathrm{LR},k}\right]' \hat{\Lambda}_{FF}(t_i)\left[\beta_k(t_i) - \hat{\beta}_{\mathrm{LR},k}\right].$$

- For suitable location and scale parameters (given in paper):

$$\frac{W_k(\alpha) - m(\alpha)}{v(\alpha)} \sim N(0,1), \quad \frac{W_k(\beta) - m(\beta)}{v(\beta)} \sim N(0,1).$$

Testing factor model

- The factor model hypothesis,

$$H_0 : \alpha(t) = 0 \in \mathbb{R}, \quad \text{for all } t \in [0, T],$$

is nested within the hypothesis of constant alphas. Thus, we can test H_0 by

$$W_0 \equiv \frac{1}{n} \sum_{i=1}^{n} [\hat{\alpha}(t_i) - \hat{\alpha}_{\mathrm{LR}}]' \hat{\Sigma}^{-1}(t_i) [\hat{\alpha}(t_i) - \hat{\alpha}_{\mathrm{LR}}]$$

Testing factor model

- The factor model hypothesis,

$$H_0 : \alpha(t) = 0 \in \mathbb{R}, \quad \text{for all } t \in [0, T],$$

is nested within the hypothesis of constant alphas. Thus, we can test H_0 by

$$W_0 \equiv \frac{1}{n} \sum_{i=1}^{n} \left[\hat{\alpha}(t_i) - \hat{\alpha}_{\mathrm{LR}}\right]' \hat{\Sigma}^{-1}(t_i) \left[\hat{\alpha}(t_i) - \hat{\alpha}_{\mathrm{LR}}\right]$$

- It then follows that (with m_0 and v_0 given in the paper):

$$\frac{W_0 - m_0}{v_0} \sim N(0, 1).$$

Choice of kernel and bandwidth

- We use a two-sided symmetric kernels because the bias is smaller than for one-sided filters as used by Andersen et al. (2006) and Lewellen and Nagel (2006).

Choice of kernel and bandwidth

- We use a two-sided symmetric kernels because the bias is smaller than for one-sided filters as used by Andersen et al. (2006) and Lewellen and Nagel (2006).
- Two different bandwidths needed for conditional and long-run estimates as the two converge at different rates.

Choice of kernel and bandwidth

- We use a two-sided symmetric kernels because the bias is smaller than for one-sided filters as used by Andersen et al. (2006) and Lewellen and Nagel (2006).
- Two different bandwidths needed for conditional and long-run estimates as the two converge at different rates.
- Bandwidth for conditional estimates selected using a two-step plug-in method which minimizes RMSE. Prior is that betas for portfolios vary slowly and the plug-in method accommodates this prior information.

Choice of kernel and bandwidth

- We use a two-sided symmetric kernels because the bias is smaller than for one-sided filters as used by Andersen et al. (2006) and Lewellen and Nagel (2006).
- Two different bandwidths needed for conditional and long-run estimates as the two converge at different rates.
- Bandwidth for conditional estimates selected using a two-step plug-in method which minimizes RMSE. Prior is that betas for portfolios vary slowly and the plug-in method accommodates this prior information.
- Bandwidth for long-run estimates scales down the conditional bandwidth by $T^{-1/3}$ since minimizing RMSE for long-run estimates requires a bandwidth of order $O\left(T^{-1/3}\right)$.

Application - data

Data is at the daily frequency from July 1963 to December 2007.

Returns: Two types of portfolios constructed by Kenneth French.

- Decile portfolios sorted by book-to-market ratios.

Application - data

Data is at the daily frequency from July 1963 to December 2007.

Returns: Two types of portfolios constructed by Kenneth French.

- Decile portfolios sorted by book-to-market ratios.
- Decile portfolios sorted on past returns (momentum).

Application - data

Data is at the daily frequency from July 1963 to December 2007.

Returns: Two types of portfolios constructed by Kenneth French.

- Decile portfolios sorted by book-to-market ratios.
- Decile portfolios sorted on past returns (momentum).
- "book-to-market strategy" - 10-1 decile portfolio that goes long value stocks and shorts growth stocks.

Application - data

Factors: Fama and French (1993) factors,

- *MKT* - excess return of market portfolio.

We only present empirical results for the CAPM version with $f_t = MKT$. See paper for results on the three-factor model.

Application - data

Factors: Fama and French (1993) factors,

- *MKT* - excess return of market portfolio.
- *SMB* - small-big return spread.

We only present empirical results for the CAPM version with $f_t = MKT$. See paper for results on the three-factor model.

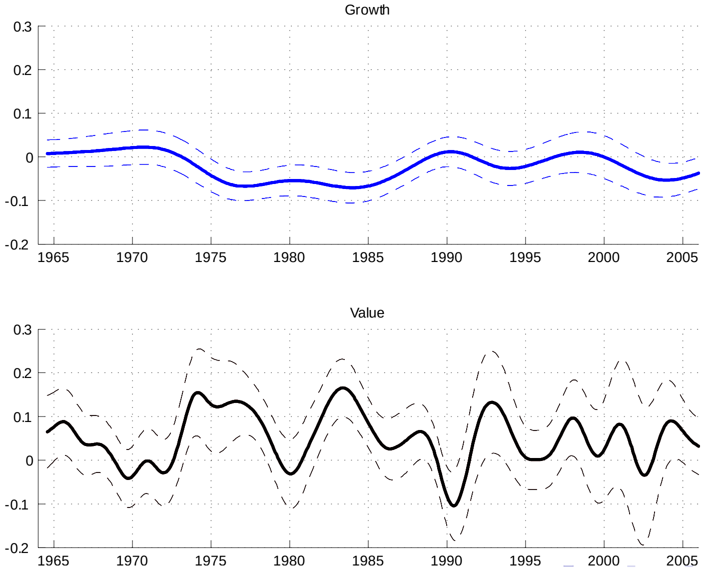
Application - data

Factors: Fama and French (1993) factors,

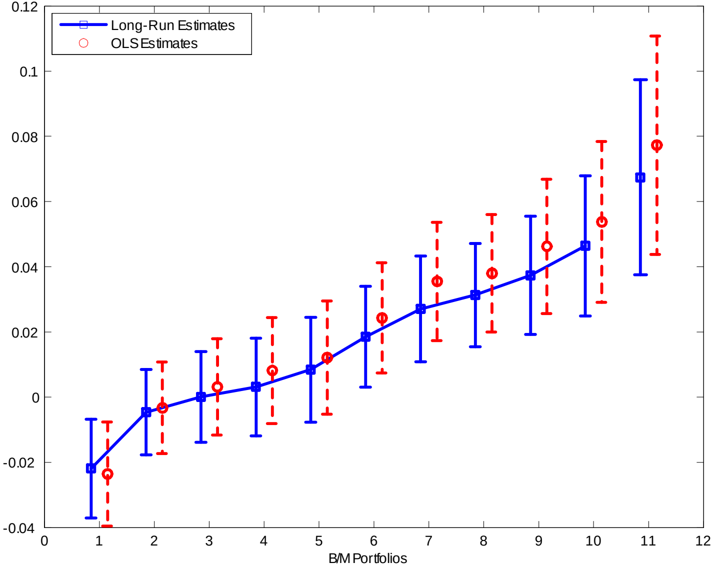
- MKT - excess return of market portfolio.
- SMB - small-big return spread.
- HML - high-low return spread.

We only present empirical results for the CAPM version with $f_t = MKT$. See paper for results on the three-factor model.

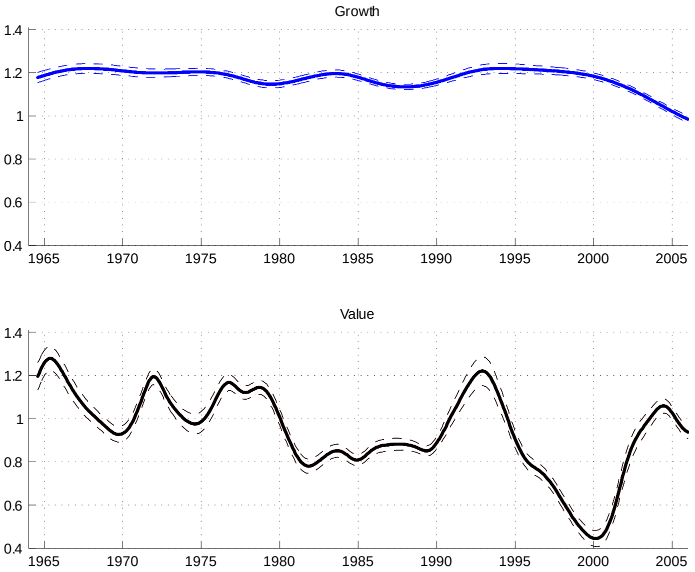
Short-run (conditional) alpha estimates



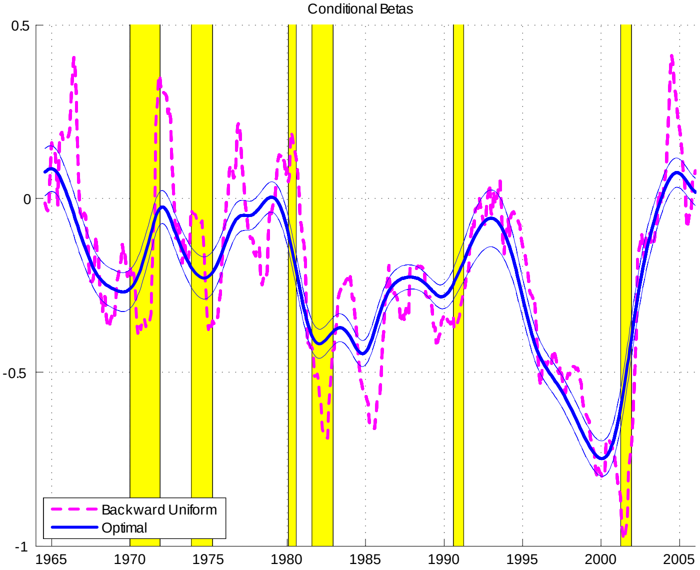
Long-Run alpha estimates



Short-Run (conditional) Betas



Short-run betas and recessions



Characterizing Conditional Value-Growth Betas

	I	II	III	IV	V
Dividend yield	4.55*		16.5**		
Default spread			-1.86		
Industrial production			0.18		
Short rate			-7.33**		
Term spread			-3.96		
Market volatility		-1.38**	-0.96*		
cay			-0.74		
NBER Recession				-0.07*	
Market risk premium					0.37*
Adjusted R^2	0.06	0.15	0.55	0.01	0.06

Market risk premium = fitted predictive regression following Petkova and Zhang (2005)

Conclusion

- **New methodology for estimation and testing of regression models with time-varying coefficients.**

Conclusion

- **New methodology for estimation and testing of regression models with time-varying coefficients.**
- Straightforward implementation and interpretation.

Conclusion

- **New methodology for estimation and testing of regression models with time-varying coefficients.**
- Straightforward implementation and interpretation.
- Formalises rolling window estimation procedures. In particular, we develop a theory regarding choice of window width with corresponding data-driven window width selection.

Conclusion

- **New methodology for estimation and testing of regression models with time-varying coefficients.**
- Straightforward implementation and interpretation.
- Formalises rolling window estimation procedures. In particular, we develop a theory regarding choice of window width with corresponding data-driven window width selection.
- **Application: Testing asset pricing models with time-varying factor loadings**.

Conclusion

- **New methodology for estimation and testing of regression models with time-varying coefficients.**
- Straightforward implementation and interpretation.
- Formalises rolling window estimation procedures. In particular, we develop a theory regarding choice of window width with corresponding data-driven window width selection.
- **Application: Testing asset pricing models with time-varying factor loadings**.
- Use data set constructed by Kenneth French.

Conclusion

- **New methodology for estimation and testing of regression models with time-varying coefficients.**
- Straightforward implementation and interpretation.
- Formalises rolling window estimation procedures. In particular, we develop a theory regarding choice of window width with corresponding data-driven window width selection.
- **Application: Testing asset pricing models with time-varying factor loadings**.
- Use data set constructed by Kenneth French.
- Substantial time variation is found.

Conclusion

- **New methodology for estimation and testing of regression models with time-varying coefficients.**
- Straightforward implementation and interpretation.
- Formalises rolling window estimation procedures. In particular, we develop a theory regarding choice of window width with corresponding data-driven window width selection.
- **Application: Testing asset pricing models with time-varying factor loadings**.
- Use data set constructed by Kenneth French.
- Substantial time variation is found.
- The APT is rejected in both the short and long run.

Future Work

- Extend methods to nonlinear dynamic models.

Future Work

- Extend methods to nonlinear dynamic models.
- Forecasting.

Future Work

- Extend methods to nonlinear dynamic models.
- Forecasting.
- Develop more rigorous bandwidth selection procedures