Internet Appendix

to

Empirical Tests of Asset Pricing Models with Individual Assets: Resolving the Errors-in-Variables Bias in Risk Premium Estimation

by

Narasimhan Jegadeesh, Joonki Noh,
Kuntara Pukthuanthong, Richard Roll, and Junbo Wang

Appendix E. Proof of Proposition 1

Let $\boldsymbol{\beta}$ be the matrix of true values of betas for all factors and assets. When there are K factors and N assets, it is a $\mathrm{K}\times\mathrm{N}$ matrix. Next, let $\hat{\boldsymbol{\beta}}_{\mathrm{IV}}$ and $\hat{\boldsymbol{\beta}}_{\mathrm{EV}}$ ($\mathrm{K}\times\mathrm{N}$ matrix) be estimated betas, where "IV" subscript indicates the beta instruments and the "EV" subscript denotes the corresponding explanatory variables, respectively. We define IV and EV periods as separate periods of data used to estimate IV and EV betas, respectively, which can be either odd months or even months over a rolling estimation window. For example, if IV period includes odd months, EV period includes even months. The symbol "^" indicates an estimate. Define $\hat{\mathbf{B}}_{\mathrm{IV}} \equiv [\mathbf{1}_{1\times\mathrm{N}}; \hat{\boldsymbol{\beta}}_{\mathrm{IV}}]$ and $\hat{\mathbf{B}}_{\mathrm{EV}} \equiv [\mathbf{1}_{1\times\mathrm{N}}; \hat{\boldsymbol{\beta}}_{\mathrm{EV}}]$, where $\mathbf{1}_{1\times\mathrm{N}}$ denotes a $1\times\mathrm{N}$ vector of ones[1] and the operator ";" stacks the first row vector on top of the second matrix. Hence, $\hat{\mathbf{B}}_{\mathrm{IV}}$ and $\hat{\mathbf{B}}_{\mathrm{EV}}$ are $(\mathrm{K}+1)\times\mathrm{N}$ matrices, each of which contains a vector of ones and K estimated factor loadings for N assets. Similarly, define $\mathbf{B} \equiv [\mathbf{1}_{1\times\mathrm{N}}; \boldsymbol{\beta}]$ as the matrix that contains a vector of ones and the matrix of true betas. Equivalently, we can write this matrix as N column vectors, i.e. $\mathbf{B} = [\mathbf{b}^1, \cdots, \mathbf{b}^{\mathrm{N}}]$, where $\mathbf{b}^{\mathrm{i}} = [1; \beta_1^{\mathrm{i}}; \cdots; \beta_{\mathrm{K}}^{\mathrm{i}}]$ for asset i.[2]

Assuming that we have T+1 months, we use T months to estimate IV and EV betas, and run a cross-sectional regression in time T+1. Define $\mathbf{r}_{\mathrm{t}}$ as the $1\times\mathrm{N}$ vector of excess returns in time

[1] We will use this convention to define vectors or matrices of ones and zeros.
[2] Note that we use superscript i to indicate asset i in the Internet Appendix.

t. Let $\mathbf{f}_t$ denote the factor realization in time t; it is a $1\times K$ vectors. And let $\boldsymbol{\varepsilon}_t = [\varepsilon_t^1, \varepsilon_t^2, \cdots, \varepsilon_t^N]$. It is a $1\times N$ vector of regression residuals in time t for N assets in Eq. (2) in the main text. Assume that factors have zero means (or they are demeaned factors). The first-stage time-series regression (Eq.(2) in the main text) can be written as $\mathbf{r}_t = \boldsymbol{\alpha} + \mathbf{f}_t\boldsymbol{\beta} + \boldsymbol{\varepsilon}_t$ where $\boldsymbol{\alpha} = [\alpha^1, \cdots, \alpha^N]$, which is a $1\times N$ vector.

Using OLS methods, we estimate IV and EV betas over the IV and EV periods, respectively. Assuming that T is even, the estimation error for IV beta can be expressed as follows:

$$\hat{\boldsymbol{\beta}}_{IV} - \boldsymbol{\beta} = \left(\mathbf{F}_{IV}^{d}{}'\mathbf{F}_{IV}^{d}\right)^{-1}\mathbf{F}_{IV}^{d}{}'\boldsymbol{\Omega}_{IV}^{d} \equiv \mathbf{u}_{IV},$$

where $\mathbf{F}_{IV}^d \equiv [\mathbf{f}_1^d; \mathbf{f}_3^d \cdots; \mathbf{f}_{T-1}^d]$ is a $T/2\times K$ matrix when the IV period includes odd months, or $\mathbf{F}_{IV}^d \equiv [\mathbf{f}_2^d; \mathbf{f}_4^d; \cdots; \mathbf{f}_T^d]$ is a $T/2\times K$ matrix when the IV period includes even months. $\boldsymbol{\Omega}_{IV}^d = [\boldsymbol{\varepsilon}_1^d; \boldsymbol{\varepsilon}_3^d; \cdots; \boldsymbol{\varepsilon}_{T-1}^d]$ is a $T/2\times N$ matrix when the IV period includes odd months, or $\boldsymbol{\Omega}_{IV}^d = [\boldsymbol{\varepsilon}_2^d; \boldsymbol{\varepsilon}_4^d; \cdots; \boldsymbol{\varepsilon}_T^d]$ is a $T/2\times N$ matrix when the IV period includes even months. The superscript d indicates a demeaned factor or demeaned residual, where average values of factors or residuals are taken over the corresponding IV period. For example, $\mathbf{f}_1^d \equiv \mathbf{f}_1 - \dfrac{2}{T}(\mathbf{f}_1 + \mathbf{f}_3 + \cdots + \mathbf{f}_{T-1})$ and $\mathbf{f}_2^d \equiv \mathbf{f}_2 - \dfrac{2}{T}(\mathbf{f}_2 + \mathbf{f}_4 + \cdots + \mathbf{f}_T)$. From the above expression for $\hat{\boldsymbol{\beta}}_{IV} - \boldsymbol{\beta}$, we obtain $\hat{\mathbf{B}}_{IV} - \mathbf{B} = [\mathbf{0}_{1\times N}; \left(\mathbf{F}_{IV}^{d}{}'\mathbf{F}_{IV}^{d}\right)^{-1}\mathbf{F}_{IV}^{d}{}'\boldsymbol{\Omega}_{IV}^{d}]$. Similarly, it can be easily shown that the estimation error for EV beta is $\hat{\boldsymbol{\beta}}_{EV} - \boldsymbol{\beta} = \left(\mathbf{F}_{EV}^{d}{}'\mathbf{F}_{EV}^{d}\right)^{-1}\mathbf{F}_{EV}^{d}{}'\boldsymbol{\Omega}_{EV}^{d} \equiv \mathbf{u}_{EV}$, leading to $\hat{\mathbf{B}}_{EV} - \mathbf{B} = [\mathbf{0}_{1\times N}; \left(\mathbf{F}_{EV}^{d}{}'\mathbf{F}_{EV}^{d}\right)^{-1}\mathbf{F}_{EV}^{d}{}'\boldsymbol{\Omega}_{EV}^{d}]$, where $\mathbf{F}_{EV}^d$ and $\boldsymbol{\Omega}_{EV}^d$ are similarly defined as $\mathbf{F}_{IV}^d$ and $\boldsymbol{\Omega}_{IV}^d$, respectively.

The model for expected returns follows Eq. (1) in the main text. If riskless borrowing and lending are allowed, then the zero-beta asset earns the risk-free rate and its excess return is zero, i.e. in Eq. (1), $\gamma_0 = 0$. Thus, the cross-sectional regression model with $\gamma_0 = 0$ and ex-post risk premium (defined as $\mathbf{f}_{T+1}+\boldsymbol{\gamma}$) can be written as $\mathbf{r}_{T+1}=(\mathbf{f}_{T+1}+\boldsymbol{\gamma})\boldsymbol{\beta}+\boldsymbol{\varepsilon}_{T+1}$, where gamma is a 1 x K vector of ex-ante risk premium.[3] We rewrite it to

$$\mathbf{r}_{T+1}=(\mathbf{f}_{T+1}+\boldsymbol{\gamma})\hat{\boldsymbol{\beta}}_{EV}+(\mathbf{f}_{T+1}+\boldsymbol{\gamma})(\boldsymbol{\beta}-\hat{\boldsymbol{\beta}}_{EV})+\boldsymbol{\varepsilon}_{T+1}.$$

Let $\boldsymbol{\xi}_{T+1}^{EV}=(\mathbf{f}_{T+1}+\boldsymbol{\gamma})(\boldsymbol{\beta}-\hat{\boldsymbol{\beta}}_{EV})+\boldsymbol{\varepsilon}_{T+1}=-(\mathbf{f}_{T+1}+\boldsymbol{\gamma})\mathbf{u}_{EV}+\boldsymbol{\varepsilon}_{T+1}$, and let $\boldsymbol{\Gamma}=(0,\boldsymbol{\gamma}+\mathbf{f}_{T+1})$, a $1\times(K+1)$ vector that contains the ex-post risk premiums. The assets' returns in time T+1 can be written as

$$\mathbf{r}_{T+1}=\boldsymbol{\Gamma}\hat{\mathbf{B}}_{EV}+\boldsymbol{\xi}_{T+1}^{EV}.$$

We then propose the following IV estimator for ex-post risk premium in month T+1:

$$\hat{\boldsymbol{\Gamma}}_{T+1}'=(\hat{\mathbf{B}}_{IV}\hat{\mathbf{B}}_{EV}')^{-1}(\hat{\mathbf{B}}_{IV}\mathbf{r}_{T+1}'). \qquad \text{(Eq. (4) in the main text)}$$

In order to show the N-consistency of the IV estimator (which is $\hat{\boldsymbol{\Gamma}}_{T+1}$), we need to make the following regularity assumptions:

(A1) The residual process $\{\boldsymbol{\varepsilon}_t, t=1,\cdots,T\}$, where $\boldsymbol{\varepsilon}_t=[\varepsilon_t^1,\varepsilon_t^2,\cdots,\varepsilon_t^N]$, is stationary with zero mean and finite fourth moments. In addition, ε_t^i and ε_s^j are uncorrelated as long as $s\neq t$. Moreover, for any $s\neq t$, $\frac{1}{\sqrt{N}}(\varepsilon_t^1\varepsilon_s^1+\cdots+\varepsilon_t^N\varepsilon_s^N)$ have finite variances.

[3]Rewriting Eq. (1) (with $\gamma_0=0$) and Eq. (2) in the main text in matrix notations, we have $E(\mathbf{r}_{T+1})=\boldsymbol{\gamma}\boldsymbol{\beta}$ denoted as Eq. (1') and $\mathbf{r}_{T+1}=\boldsymbol{\alpha}+\mathbf{f}_{T+1}\boldsymbol{\beta}+\boldsymbol{\varepsilon}_{T+1}$ denoted as Eq. (2') where $\boldsymbol{\gamma}=[\gamma_1,\cdots,\gamma_K]$, $\boldsymbol{\alpha}=[\alpha^1,\cdots,\alpha^N]$, and $\boldsymbol{\varepsilon}_{T+1}=[\varepsilon_{T+1}^1,\cdots,\varepsilon_{T+1}^N]$. With the assumptions that $E(\mathbf{f}_t)=0$ and $E(\boldsymbol{\varepsilon}_t)=0$, taking the expectation of Eq. (2') produces $E(\mathbf{r}_{T+1})=\boldsymbol{\alpha}$ denoted as Eq.(2''). Combining Eqs. (1'), (2'), and (2") gives us $\mathbf{r}_{T+1}=(\mathbf{f}_{T+1}+\boldsymbol{\gamma})\boldsymbol{\beta}+\boldsymbol{\varepsilon}_{T+1}$.

(A2) The factor process $\{\mathbf{f}_t, t=1,\cdots,T\}$ is stationary with finite fourth moments, and is independent of the residual process $\{\boldsymbol{\varepsilon}_t, t=1,\cdots,T\}$.

(A3) For any i and t, the elements in $\mathbf{b}^i$ and ε_t^i are uncorrelated. For any t, the elements in $\frac{1}{\sqrt{N}}(\mathbf{b}^1\varepsilon_t^1+\cdots+\mathbf{b}^N\varepsilon_t^N)$. Moreover, when N→∞, $\mathbf{BB}'/N$ converges to an invertible matrix $\mathbf{bb}'$ $(=E(\mathbf{b}^i(\mathbf{b}^i)'))$.

Note that the Assumptions (A1) and (A3) are satisfied if in each time t, the residuals in $[\varepsilon_t^1,\cdots,\varepsilon_t^N]$ are asymptotically weakly correlated (Shanken,1992), regression residuals have finite fourth moments, and the maximum values of betas across all stocks are finite.

With the Assumptions (A1) to (A3), N-consistency is presented in Theorem E.

Theorem E (Consistency, Proposition 1 in the main text) Assuming (A1) to (A3), the estimated risk premium $\hat{\boldsymbol{\Gamma}}_{T+1}'\equiv(\frac{1}{N}\hat{\mathbf{B}}_{IV}\hat{\mathbf{B}}_{EV}')^{-1}(\frac{1}{N}\hat{\mathbf{B}}_{IV}\mathbf{r}_{T+1}')$ converges to $\boldsymbol{\Gamma}'$ when N approaches to infinity.

Proof of the consistency: Note that

$$\hat{\boldsymbol{\Gamma}}_{T+1}'-\boldsymbol{\Gamma}'=(\frac{1}{N}\hat{\mathbf{B}}_{IV}\hat{\mathbf{B}}_{EV}')^{-1}(\frac{1}{N}\hat{\mathbf{B}}_{IV}\boldsymbol{\xi}_{T+1}^{EV}').$$

To show the consistency as N goes to infinity, we need to prove the following two convergences:

$$\frac{1}{N}\hat{\mathbf{B}}_{IV}\boldsymbol{\xi}_{T+1}^{EV}'=\frac{1}{N}\sum_{i=1}^{N}(\mathbf{b}^i+[0;\left(\mathbf{F}_{IV}^{d}{}'\mathbf{F}_{IV}^{d}\right)^{-1}\mathbf{F}_{IV}^{d}{}'(\boldsymbol{\varepsilon}_{IV}^{d,i})'])\xi_{T+1}^{EV,i}\to\mathbf{0}_{(K+1)\times 1} \text{ and}$$

$$\frac{1}{N}\hat{\mathbf{B}}_{IV}\hat{\mathbf{B}}_{EV}'=\frac{1}{N}(\mathbf{B}+[\mathbf{0}_{1\times N};\left(\mathbf{F}_{IV}^{d}{}'\mathbf{F}_{IV}^{d}\right)^{-1}\mathbf{F}_{IV}^{d}{}'\boldsymbol{\Omega}_{IV}^{d}])(\mathbf{B}+[\mathbf{0}_{1\times N};\left(\mathbf{F}_{EV}^{d}{}'\mathbf{F}_{EV}^{d}\right)^{-1}\mathbf{F}_{EV}^{d}{}'\boldsymbol{\Omega}_{EV}^{d}])'$$

$$=\frac{1}{N}\sum_{i=1}^{N}((\mathbf{b}^i+[0;\left(\mathbf{F}_{IV}^{d}{}'\mathbf{F}_{IV}^{d}\right)^{-1}\mathbf{F}_{IV}^{d}{}'(\boldsymbol{\varepsilon}_{IV}^{d,i})'])(\mathbf{b}^i+[0;\left(\mathbf{F}_{EV}^{d}{}'\mathbf{F}_{EV}^{d}\right)^{-1}\mathbf{F}_{EV}^{d}{}'(\boldsymbol{\varepsilon}_{EV}^{d,i})'])')\to\mathbf{bb}'.$$

Here, $\boldsymbol{\varepsilon}_{IV}^{d,i}$ is a $1\times T/2$ vector that contains the demeaned residuals for asset i in the IV period, and $\boldsymbol{\varepsilon}_{EV}^{d,i}$ is a $1\times T/2$ vector that contains the demeaned residuals for asset i in the EV period. For example, if the IV period includes odd months, and the EV period includes even months, $\boldsymbol{\varepsilon}_{IV}^{d,i}=[\varepsilon_1^{d,i},\varepsilon_3^{d,i},\cdots,\varepsilon_{T-1}^{d,i}]$ and $\boldsymbol{\varepsilon}_{EV}^{d,i}=[\varepsilon_2^{d,i},\varepsilon_4^{d,i},\cdots,\varepsilon_T^{d,i}]$ where $\varepsilon_t^{d,i}\equiv\varepsilon_t^i-\frac{2}{T}(\varepsilon_1^i+\varepsilon_3^i+\cdots+\varepsilon_{T-1}^i)$

when t is odd, and $\varepsilon_t^{d,i} \equiv \varepsilon_t^i - \frac{2}{T}(\varepsilon_2^i + \varepsilon_4^i + \cdots + \varepsilon_T^i)$ when t is even. Moreover, ξ_{T+1}^{EV}, a $1 \times N$ vector, can be written as $[\xi_{T+1}^{EV,1}, \cdots, \xi_{T+1}^{EV,N}]$.

Under Assumptions (A1) to (A3), we now have the following observations:

I. For any i, $E(\mathbf{b}^i \xi_{T+1}^{EV,i}) = \mathbf{0}_{(K+1)\times 1}$, $E(\mathbf{b}^i [0; \left(\mathbf{F}_{IV}^d{}'\mathbf{F}_{IV}^d\right)^{-1} \mathbf{F}_{IV}^d{}'(\boldsymbol{\varepsilon}_{IV}^{d,i})']') = \mathbf{0}_{(K+1)\times(K+1)}$,

$E(\mathbf{b}^i ([0; \left(\mathbf{F}_{EV}^d{}'\mathbf{F}_{EV}^d\right)^{-1} \mathbf{F}_{EV}^d{}'(\boldsymbol{\varepsilon}_{EV}^{d,i})']') = \mathbf{0}_{(K+1)\times(K+1)}$,

$E([0; \left(\mathbf{F}_{IV}^d{}'\mathbf{F}_{IV}^d\right)^{-1} \mathbf{F}_{IV}^d{}'(\boldsymbol{\varepsilon}_{IV}^{d,i})'][0; \left(\mathbf{F}_{EV}^d{}'\mathbf{F}_{EV}^d\right)^{-1} \mathbf{F}_{EV}^d{}'(\boldsymbol{\varepsilon}_{EV}^{d,i})']') = \mathbf{0}_{(K+1)\times(K+1)}$, and

$E([0; \left(\mathbf{F}_{IV}^d{}'\mathbf{F}_{IV}^d\right)^{-1} \mathbf{F}_{IV}^d{}'(\boldsymbol{\varepsilon}_{IV}^{d,i})']\xi_{\mathbf{T+1}}^{EV,\mathbf{i}}) = \mathbf{0}_{(K+1)\times 1}$.

Since the elements in $\frac{1}{\sqrt{N}}(\mathbf{b}^1\varepsilon_t^1 + \cdots + \mathbf{b}^N\varepsilon_t^N)$ have finite variances for any time t (Assumption (A3)), it is clear that the elements in $\frac{1}{\sqrt{N}}(\mathbf{b}^1\xi_{T+1}^{EV,1} + \cdots + \mathbf{b}^N\xi_{T+1}^{EV,N})$ have finite variances. For the same reason, the elements in

$\frac{1}{\sqrt{N}}(\mathbf{b}^1[0; \left(\mathbf{F}_{IV}^d{}'\mathbf{F}_{IV}^d\right)^{-1} \mathbf{F}_{IV}^d{}'(\boldsymbol{\varepsilon}_{IV}^{d,1})']' + \cdots + \mathbf{b}^N[0; \left(\mathbf{F}_{IV}^d{}'\mathbf{F}_{IV}^d\right)^{-1} \mathbf{F}_{IV}^d{}'(\boldsymbol{\varepsilon}_{IV}^{d,N})']')$ have finite variances, and the elements in $\frac{1}{\sqrt{N}}(\mathbf{b}^1[0; \left(\mathbf{F}_{EV}^d{}'\mathbf{F}_{EV}^d\right)^{-1} \mathbf{F}_{EV}^d{}'(\boldsymbol{\varepsilon}_{EV}^{d,1})']' + \cdots + \mathbf{b}^N[0; \left(\mathbf{F}_{EV}^d{}'\mathbf{F}_{EV}^d\right)^{-1} \mathbf{F}_{EV}^d{}'(\boldsymbol{\varepsilon}_{EV}^{d,N})']')$ have finite variances.

II. If regression residuals have finite fouth moments (Assumption (A1)), and $\frac{1}{\sqrt{N}}(\varepsilon_t^1\varepsilon_s^1 + \cdots + \varepsilon_t^N\varepsilon_s^N)$ have finite variances for any $s \neq t$ (Assumption (A1)), then for any i, the elements in $\frac{1}{\sqrt{N}}\sum_{i=1}^{N}([0; \left(\mathbf{F}_{IV}^d{}'\mathbf{F}_{IV}^d\right)^{-1} \mathbf{F}_{IV}^d{}'(\boldsymbol{\varepsilon}_{IV}^{d,i})']\xi_{T+1}^{EV,i})$ have finite variances, and the elements in $\frac{1}{\sqrt{N}}\sum_{i=1}^{N}([0; \left(\mathbf{F}_{IV}^d{}'\mathbf{F}_{IV}^d\right)^{-1} \mathbf{F}_{IV}^d{}'(\boldsymbol{\varepsilon}_{IV}^{d,i})'][0; \left(\mathbf{F}_{EV}^d{}'\mathbf{F}_{EV}^d\right)^{-1} \mathbf{F}_{EV}^d{}'(\boldsymbol{\varepsilon}_{EV}^{d,i})']')$ also have finite variances.

With the observations I, II, and III, apply Markov's Law of Large Numbers,

$$\frac{1}{N}\sum_{i=1}^{N}\mathbf{b}^{i}\xi_{T+1}^{EV,i} \rightarrow \mathbf{0}_{(K+1)\times 1}$$

$$\frac{1}{N}\sum_{i=1}^{N}([0;\left(\mathbf{F}_{IV}^{d}{}'\mathbf{F}_{IV}^{d}\right)^{-1}\mathbf{F}_{IV}^{d}{}'(\boldsymbol{\varepsilon}_{IV}^{d,i})']\xi_{T+1}^{EV,i}) \rightarrow \mathbf{0}_{(K+1)\times 1}$$

$$\frac{1}{N}\sum_{i=1}^{N}([0;\left(\mathbf{F}_{IV}^{d}{}'\mathbf{F}_{IV}^{d}\right)^{-1}\mathbf{F}_{IV}^{d}{}'(\boldsymbol{\varepsilon}_{IV}^{d,i})'](\mathbf{b}^{i})') \rightarrow \mathbf{0}_{(K+1)\times(K+1)}$$

$$\frac{1}{N}\sum_{i=1}^{N}(\mathbf{b}^{i}[0;\left(\mathbf{F}_{EV}^{d}{}'\mathbf{F}_{EV}^{d}\right)^{-1}\mathbf{F}_{EV}^{d}{}'(\boldsymbol{\varepsilon}_{EV}^{d,i})']') \rightarrow \mathbf{0}_{(K+1)\times(K+1)}$$

$$\frac{1}{N}\sum_{i=1}^{N}([0;\left(\mathbf{F}_{IV}^{d}{}'\mathbf{F}_{IV}^{d}\right)^{-1}\mathbf{F}_{IV}^{d}{}'(\boldsymbol{\varepsilon}_{IV}^{d,i})'][0;\left(\mathbf{F}_{EV}^{d}{}'\mathbf{F}_{EV}^{d}\right)^{-1}\mathbf{F}_{EV}^{d}{}'(\boldsymbol{\varepsilon}_{EV}^{d,i})']') \rightarrow \mathbf{0}_{(K+1)\times(K+1)}$$

From Assumption (A3), $\mathbf{BB}'/N$ converges to $\mathbf{bb}'$. Together with the equations above, we have

$$\frac{1}{N}\hat{\mathbf{B}}_{IV}\boldsymbol{\xi}_{T+1}^{EV}{}' = \frac{1}{N}\sum_{i=1}^{N}(\mathbf{b}^{i} + [0;\left(\mathbf{F}_{IV}^{d}{}'\mathbf{F}_{IV}^{d}\right)^{-1}\mathbf{F}_{IV}^{d}{}'(\boldsymbol{\varepsilon}_{IV}^{d,i})'])(\xi_{T+1}^{EV,i}) \rightarrow \mathbf{0}_{(K+1)\times 1}, \text{ and}$$

$$\frac{1}{N}\hat{\mathbf{B}}_{IV}\hat{\mathbf{B}}_{EV}{}' = \frac{1}{N}(\mathbf{B} + [\mathbf{0}_{1\times N};\left(\mathbf{F}_{IV}^{d}{}'\mathbf{F}_{IV}^{d}\right)^{-1}\mathbf{F}_{IV}^{d}{}'\boldsymbol{\Omega}_{IV}^{d}])(\mathbf{B} + [\mathbf{0}_{1\times N};\left(\mathbf{F}_{EV}^{d}{}'\mathbf{F}_{EV}^{d}\right)^{-1}\mathbf{F}_{EV}^{d}{}'\boldsymbol{\Omega}_{EV}^{d}])'$$
$$= \frac{1}{N}\sum_{i=1}^{N}((\mathbf{b}^{i} + [0;\left(\mathbf{F}_{IV}^{d}{}'\mathbf{F}_{IV}^{d}\right)^{-1}\mathbf{F}_{IV}^{d}{}'(\boldsymbol{\varepsilon}_{IV}^{d,i})'])(\mathbf{b}^{i} + [0;\left(\mathbf{F}_{EV}^{d}{}'\mathbf{F}_{EV}^{d}\right)^{-1}\mathbf{F}_{EV}^{d}{}'(\boldsymbol{\varepsilon}_{EV}^{d,i})'])') \rightarrow \mathbf{bb}'. \quad \blacksquare$$

We also derive the conditional and unconditional asymptotic distributions of the IV estimator as N goes to infinity. These theorems and proofs are available from the authors upon request.

Appendix F. Consistency of IV estimator with time-varying betas

Theorem E requires that betas and true risk premiums are constant. In this section, we relax this assumption (but still keep Assumptions (A1) and (A2) in Appendix E). With the assumption that riskless borrowing and lending are allowed, Eq. (1) in the main text with time-varying betas and risk premiums can be written as:

$$E(r_t^i \mid \gamma_{t,1},\cdots,\gamma_{t,K},\beta_{t,1}^i,\cdots,\beta_{t,K}^i)=\sum_{k=1}^{K}\beta_{t,k}^i\times\gamma_{t,k}.$$[4]

Here $\beta_{t,k}^i$ is the beta of factor k for asset i in time t, and $\gamma_{t,k}$ is the risk premium for factor k in time t. Similarly, the first-stage time-series regression can be written as: $r_t^i=\alpha^i+\sum_{k=1}^{K}\beta_{t,k}^i\times f_{k,t}+\varepsilon_t^i$. Let $\beta_{t,k}^i\equiv\beta_k^i+u_{t,k}^i$ where β_k^i is the unconditional mean of the beta of factor k for asset i, and $u_{t,k}^i$ is the shock in beta in time t.

We can rewrite the above equations in vector and matrix notations. Assume that the true risk premium $\boldsymbol{\gamma}_t$, a $1\times K$ vector, is also time-varying and satisfies the following assumption:

Assumption (G): For all t, s and i, $\boldsymbol{\gamma}_t$ is independent of the regression residual ε_s^i.

Let $\boldsymbol{\beta}_t^i$, a $K\times 1$ vector, be the betas of K factors for asset i in time t, and $\boldsymbol{\beta}^i$ be its unconditional mean. Let $\boldsymbol{\alpha}=[\alpha^1,\cdots,\alpha^N]$. In addition, Let $\boldsymbol{\beta}_t^i\equiv\boldsymbol{\beta}^i+\mathbf{u}_t^i$, where the time t shock $\mathbf{u}_t^i$ is a $K\times 1$ vector. Denote $\boldsymbol{\beta}=[\boldsymbol{\beta}^1,\cdots,\boldsymbol{\beta}^N]$, $\boldsymbol{\beta}_t=[\boldsymbol{\beta}_t^1,\cdots,\boldsymbol{\beta}_t^N]$ and $\mathbf{u}_t=[\mathbf{u}_t^1,\cdots,\mathbf{u}_t^N]$. The asset pricing model in Eq. (1) in the main text with time-varying betas and risk premium can be written as $E(\mathbf{r}_{T+1}\mid\boldsymbol{\gamma}_{T+1},\boldsymbol{\beta}_{T+1})=\boldsymbol{\gamma}_{T+1}\boldsymbol{\beta}_{T+1}$ (with $\gamma_0=0$), and the first-stage time-series regression can be written as $\mathbf{r}_t=\boldsymbol{\alpha}+\mathbf{f}_t\boldsymbol{\beta}_t+\boldsymbol{\varepsilon}_t$. With the similar derivation in Appendix E, we can show that the cross-sectional regression model can be written as $\mathbf{r}_{T+1}=(\mathbf{f}_{T+1}+\boldsymbol{\gamma}_{T+1}-E(\mathbf{f}_{T+1}\mid\boldsymbol{\gamma}_{T+1},\boldsymbol{\beta}_{T+1}))\boldsymbol{\beta}+\mathbf{v}_{T+1}$, where $\mathbf{v}_{T+1}=\boldsymbol{\varepsilon}_{T+1}+(\mathbf{f}_{T+1}+\boldsymbol{\gamma}_{T+1}-E(\mathbf{f}_{T+1}\mid\boldsymbol{\gamma}_{T+1},\boldsymbol{\beta}_{T+1}))\mathbf{u}_{T+1}$.

Define $\mathbf{B}\equiv[\mathbf{1}_{1\times N};\boldsymbol{\beta}]$, $\hat{\mathbf{B}}_{IV}\equiv[\mathbf{1}_{1\times N};\hat{\boldsymbol{\beta}}_{IV}]$ and $\hat{\mathbf{B}}_{EV}\equiv[\mathbf{1}_{1\times N};\hat{\boldsymbol{\beta}}_{EV}]$ where $\hat{\boldsymbol{\beta}}_{IV}$ and $\hat{\boldsymbol{\beta}}_{EV}$ are the estimated betas in the IV and EV periods, respectively. The first-stage time-series regression can be written as $\mathbf{r}_t=\boldsymbol{\alpha}+\mathbf{f}_t\boldsymbol{\beta}+\mathbf{e}_t$ where $\mathbf{e}_t\equiv\boldsymbol{\varepsilon}_t+\mathbf{f}_t\mathbf{u}_t$. The estimation errors for IV and EV betas are:

[4] Note that we use superscript i to indicate asset i in the Internet Appendix.

$$\hat{\boldsymbol{\beta}}_{IV} - \boldsymbol{\beta} = \left(\mathbf{F}_{IV}^{d}{}'\mathbf{F}_{IV}^{d}\right)^{-1}\mathbf{F}_{IV}^{d}{}'\mathbf{E}_{IV}^{d},$$

$$\hat{\boldsymbol{\beta}}_{EV} - \boldsymbol{\beta} = \left(\mathbf{F}_{EV}^{d}{}'\mathbf{F}_{EV}^{d}\right)^{-1}\mathbf{F}_{EV}^{d}{}'\mathbf{E}_{EV}^{d},$$

respectively, where $\mathbf{E}_{IV}^{d}$ ($\mathbf{E}_{EV}^{d}$) is defined similarly to $\boldsymbol{\Omega}_{IV}^{d}$ ($\boldsymbol{\Omega}_{EV}^{d}$) in Appendix E, by replacing $\boldsymbol{\varepsilon}_t$ with $\mathbf{e}_t$.

Denote $\boldsymbol{\Gamma}_{T+1}$, a $1\times(K+1)$ vector, as the ex-post risk premiums in time T+1 (defined as $\boldsymbol{\Gamma}_{T+1} = (0, \mathbf{f}_{T+1} + \boldsymbol{\gamma}_{T+1} - E(\mathbf{f}_{T+1} \mid \boldsymbol{\gamma}_{T+1}, \boldsymbol{\beta}_{T+1}))$). Following the same derivation in Appendix E, the cross-sectional regression model can be written as

$$\mathbf{r}_{T+1} = \boldsymbol{\Gamma}_{T+1}\hat{\mathbf{B}}_{EV} + \boldsymbol{\xi}_{T+1}^{EV},$$

where $\boldsymbol{\xi}_{T+1}^{EV} = (\mathbf{f}_{T+1} + \boldsymbol{\gamma}_{T+1} - E(\mathbf{f}_{T+1} \mid \boldsymbol{\gamma}_{T+1}, \boldsymbol{\beta}_{T+1}))(\boldsymbol{\beta} - \hat{\boldsymbol{\beta}}_{EV}) + \mathbf{v}_{T+1}$.

The IV estimator is $\hat{\boldsymbol{\Gamma}}_{T+1}' = (\hat{\mathbf{B}}_{IV}\hat{\mathbf{B}}_{EV}')^{-1}(\hat{\mathbf{B}}_{IV}\mathbf{r}_{T+1}')$. We can show two types of consistency of this estimator, with different assumptions on the dynamcs of $\mathbf{u}_t^i$.

Assumption (U1): For all t, s and i, $\mathbf{u}_t^i$ is independent of $\mathbf{f}_s$ (factor in time s), ε_s^i (regression residuals in time s), unconditional mean of beta $\boldsymbol{\beta}^i$ and $\boldsymbol{\gamma}_s$ (risk premium in time s). It also has zero mean and finite fourth moment. In addition, $\mathbf{u}_t^i$ and $\mathbf{u}_s^j$ are uncorrelated as long as $s \neq t$. Moreover, for each asset i, the stochastic process $\{\mathbf{u}_t^i, t=1,\cdots,T\}$ is stationary. Moreover, for any $s \neq t$, elements in $\frac{1}{\sqrt{N}}(\mathbf{u}_t^1(\mathbf{u}_s^1)' + \cdots + \mathbf{u}_t^N(\mathbf{u}_s^N)')$ and $\frac{1}{\sqrt{N}}(\mathbf{u}_t^1\varepsilon_s^1 + \cdots + \mathbf{u}_t^N\varepsilon_s^N)$ have finite variances.

Assumption (U2): For all t, s and i, $\mathbf{u}_t^i$ is independent of $\mathbf{f}_s$ (factor in time s), ε_s^i (regression residuals in time s), unconditional mean of beta $\boldsymbol{\beta}^i$ and $\boldsymbol{\gamma}_s$ (risk premium in time s). It also has zero mean and finite fourth moments. In addition, for each asset i, the stochastic process $\{\mathbf{u}_t^i, t=1,\cdots,T\}$ is stationary and ergodic.

The key difference between those two assumptions is the autocorrelation of $\{\mathbf{u}_t^i, t=1,\cdots,T\}$ process. Assumption (U1) imposes no autocorrelation, while (U2) relaxes this assumption. In addition, since $\mathbf{u}_t^i$ is independent of $\mathbf{f}_{T+1}$, we have $E(\mathbf{f}_{T+1} | \boldsymbol{\gamma}_{T+1}, \boldsymbol{\beta}_{T+1}) = E(\mathbf{f}_{T+1} | \boldsymbol{\gamma}_{T+1})$. Therefore, $\boldsymbol{\xi}_{T+1}^{EV} = (\mathbf{f}_{T+1} + \boldsymbol{\gamma}_{T+1} - E(\mathbf{f}_{T+1} | \boldsymbol{\gamma}_{T+1}))(\boldsymbol{\beta} - \hat{\boldsymbol{\beta}}_{EV}) + \boldsymbol{\varepsilon}_{T+1} + (\mathbf{f}_{T+1} + \boldsymbol{\gamma}_{T+1} - E(\mathbf{f}_{T+1} | \boldsymbol{\gamma}_{T+1}))\mathbf{u}_{T+1}$ and $\boldsymbol{\Gamma}_{T+1} = (0, \mathbf{f}_{T+1} + \boldsymbol{\gamma}_{T+1} - E(\mathbf{f}_{T+1} | \boldsymbol{\gamma}_{T+1}))$. Since $\mathbf{f}_{T+1} + \boldsymbol{\gamma}_{T+1} - E(\mathbf{f}_{T+1} | \boldsymbol{\gamma}_{T+1})$ is independent of $\mathbf{u}_t^i$ and ε_t^i for any t and i (Assumptions (G) and (U1)), we can show that $\boldsymbol{\xi}_{T+1}^{EV}$ has zero mean.

We also impose the following assumption on $\mathbf{u}_t^i$ process, ε_t^i and matrix $\mathbf{B}$ (which can be written as $[\mathbf{b}^1, \cdots, \mathbf{b}^N]$ as in Appendix E).

Assumption (B): The elements in $\frac{1}{\sqrt{N}}(\mathbf{b}^1 (u_t^1)' + \cdots + \mathbf{b}^N (u_t^N)')$ and $\frac{1}{\sqrt{N}}(\mathbf{b}^1 \varepsilon_t^1 + \cdots + \mathbf{b}^N \varepsilon_t^N)$ have finite variances for any t, and when N→∞, $\mathbf{BB}'/N$ converges to an invertible matrix $\mathbf{bb}'$ ($= E(\mathbf{b}^i (\mathbf{b}^i)')$). In addition, for any i and t, the elements in $\mathbf{b}^i$ and ε_t^i are uncorrelated.

We write $\boldsymbol{\xi}_{T+1}^{EV}$ ($1 \times N$ vector), as $[\xi_{T+1}^{EV,1}, \cdots, \xi_{T+1}^{EV,N}]$. Assumption (B) implies that the elements in $\frac{1}{\sqrt{N}}(\mathbf{b}^1 \xi_{T+1}^{EV,1} + \cdots + \mathbf{b}^N \xi_{T+1}^{EV,N})$ have finite variances, since $\xi_{T+1}^{EV,i}$ is a linear combination of $\mathbf{u}_t^i$'s and ε_t^i's. Assumptions (U1) and (B) imply that $\xi_{T+1}^{EV,i}$ and $\mathbf{b}^i$ are uncorrelated, i.e. $E(\mathbf{b}^i \xi_{T+1}^{EV,i}) = \mathbf{0}_{(K+1)\times 1}$.

With the assumptions above, we state the following Theorem F:

Theorem F (1) (N-Consistency) Assuming (A1), (A2), (G), (U1) and (B), the estimated risk premium $\hat{\boldsymbol{\Gamma}}_{T+1}' \equiv (\frac{1}{N}\hat{\mathbf{B}}_{IV}\hat{\mathbf{B}}_{EV}')^{-1}(\frac{1}{N}\hat{\mathbf{B}}_{IV}\mathbf{r}_{T+1}')$ converges to $\boldsymbol{\Gamma}_{T+1}'$ when N approaches to infinity.

(2) (Sequential consistency) Assuming (A1), (A2), (U2) and (B), the estimated risk premium $\hat{\mathbf{\Gamma}}_{T+1}' \equiv (\frac{1}{N}\hat{\mathbf{B}}_{IV}\hat{\mathbf{B}}_{EV}')^{-1}(\frac{1}{N}\hat{\mathbf{B}}_{IV}\mathbf{r}_{T+1}')$ converges to $\mathbf{\Gamma}_{T+1}'$ when we take a probability limit as T approaches to infinity first, and then take a probability limit as N approaches to infinity.

Proof of (1): With the assumptions above, the proof will be exactly the same as the proof for Theorem E, by replacing $\mathbf{\Omega}_{IV}^{d}$ ($\mathbf{\Omega}_{EV}^{d}$) with $\mathbf{E}_{IV}^{d}$ ($\mathbf{E}_{EV}^{d}$), and $\boldsymbol{\varepsilon}_t$ with $\mathbf{e}_t\,(=\boldsymbol{\varepsilon}_t+\mathbf{f}_t\mathbf{u}_t)$.

Proof of (2): Similarly to the proof for Theorem E,

$$\hat{\mathbf{\Gamma}}_{T+1}'-\mathbf{\Gamma}_{T+1}' = (\frac{1}{N}\hat{\mathbf{B}}_{IV}\hat{\mathbf{B}}_{EV}')^{-1}(\frac{1}{N}\hat{\mathbf{B}}_{IV}\boldsymbol{\xi}_{T+1}^{EV}{}').$$

We have shown that

$$\hat{\boldsymbol{\beta}}_{IV}-\boldsymbol{\beta}=\left(\mathbf{F}_{IV}^{d}{}'\mathbf{F}_{IV}^{d}\right)^{-1}\mathbf{F}_{IV}^{d}{}'\mathbf{E}_{IV}^{d},$$

$$\hat{\boldsymbol{\beta}}_{EV}-\boldsymbol{\beta}=\left(\mathbf{F}_{EV}^{d}{}'\mathbf{F}_{EV}^{d}\right)^{-1}\mathbf{F}_{EV}^{d}{}'\mathbf{E}_{EV}^{d}.$$

Let $\mathbf{e}_t=[e_t^1,\cdots,e_t^N]$. Since e_t^i is a linear combination of $\mathbf{u}_t^i$ and ε_t^i, from Assumptions (A1), (A2) and (U2), it is clear that the stochastic process $\{e_t^i, t=1,\cdots,T\}$ is stationary and ergodic with zero mean and finite fourth moment, and for any i, e_t^i is independent of factors. Taking a probability limit as T approaches to infinity, $\hat{\mathbf{B}}_{IV}\to\mathbf{B}$, $\hat{\mathbf{B}}_{IV}\to\mathbf{B}$ and $\boldsymbol{\xi}_{T+1}^{EV}\to\mathbf{v}_{T+1}$ following the Markov's law of large number. Hence,

$$(\frac{1}{N}\hat{\mathbf{B}}_{IV}\hat{\mathbf{B}}_{EV}')^{-1}(\frac{1}{N}\hat{\mathbf{B}}_{IV}\boldsymbol{\xi}_{T+1}^{EV}{}')\to(\frac{1}{N}\mathbf{BB}')^{-1}(\frac{1}{N}\mathbf{B}\mathbf{v}_{T+1}').$$

Next, since $\mathbf{v}_{T+1}$ is a linear combination of $\mathbf{u}_t$ and $\boldsymbol{\varepsilon}_t$, from Assumptions (B) and (U2), $E(\mathbf{B}\mathbf{v}_{T+1}')=\mathbf{0}_{(k+1)\times 1}$, and the elements in $\frac{1}{\sqrt{N}}(\mathbf{b}^1 v_{T+1}^1+\cdots+\mathbf{b}^N v_{T+1}^N)$ have finite variances when N→∞. In addition, from Assumption (B), $(\frac{1}{N}\mathbf{BB}')\to\mathbf{bb}'$ as N appraoches to infinity. Apply

Markov's law of large numbers, $(\frac{1}{N}\mathbf{BB}')^{-1}(\frac{1}{N}\mathbf{B}\boldsymbol{\nu}_{T+1}') \to \mathbf{0}_{(k+1)\times 1}$ when N→∞. Therefore, $(\frac{1}{N}\hat{\mathbf{B}}_{IV}\hat{\mathbf{B}}_{EV}')^{-1}(\frac{1}{N}\hat{\mathbf{B}}_{IV}\boldsymbol{\xi}_{T+1}^{EV}{}')\to \mathbf{0}_{(k+1)\times 1}$ (i.e. $\hat{\boldsymbol{\Gamma}}_{T+1}'$ converges to $\boldsymbol{\Gamma}_{T+1}'$) when we take a probability limit as T approaches to infinity first, and then take a probability limit as N approaches to infinity. ■

Appendix G. Time-varying characteristics

In this section, we incorpate stock characteristics into the cross-sectional regression: i.e. in the second-stage regression, the independent variables are estimated betas as well as characteristics of stocks. We also assume that both estimated betas and characteristics are proxies for the true factor loading (true betas), and they are correlated cross-sectionally. Thus, characteristics are used both as instruments for beta estimates and as control variables. We propose a new IV estimator: the IV mean-estimator, and prove its convergence to the ex-post risk premium as the dimensions of cross-section and time-series grow indefinitely. The estimator in Proposition 2 in the main text is a special case of the IV mean-estimator proposed in this Appendix.

The dependent variable of the IV mean-estimator in the second-stage cross-sectional regression is the average return $\bar{\mathbf{r}}_{DV} = \frac{1}{T_m}\sum_{t\in DV}\mathbf{r}_t$ over the months not used to construct $\hat{\boldsymbol{\beta}}_{IV}$ and $\hat{\boldsymbol{\beta}}_{EV}$ (we call them the dependent variable period or the DV period, and assume that the DV period has T_m months). Without loss of generality, we assume that IV and EV betas are constructed using observations from months 1 to T, and the DV period has observations from months $T+1$ to $T+T_m$.

Following the similar derivations in Appendices E and F, we can show that for any t in the DV period, the asset return with time-varying beta and true risk premium can be written as $\mathbf{r}_t = (\mathbf{f}_t + \boldsymbol{\gamma}_t - E(\mathbf{f}_t \mid \boldsymbol{\gamma}_t, \boldsymbol{\beta}_t))\boldsymbol{\beta}_t + \boldsymbol{\varepsilon}_t$. From Assumption (U2), we have $E(\mathbf{f}_t \mid \boldsymbol{\gamma}_t, \boldsymbol{\beta}_t) = E(\mathbf{f}_t \mid \boldsymbol{\gamma}_t)$, leading to $\mathbf{r}_t = (\mathbf{f}_t + \boldsymbol{\gamma}_t - E(\mathbf{f}_t \mid \boldsymbol{\gamma}_t))\boldsymbol{\beta}_t + \boldsymbol{\varepsilon}_t$.

The cross-sectional regression model of regressing the average return over the DV period on the estimated beta over the EV period can be written as $\bar{\mathbf{r}}_{\mathrm{DV}} = \boldsymbol{\Gamma}\hat{\mathbf{B}}_{\mathrm{EV}} + \boldsymbol{\xi}_{\mathrm{DV}}$, where $\boldsymbol{\Gamma}$ is defined as $(0, \frac{1}{\mathrm{T_m}} \sum_{\mathrm{t}\in\mathrm{DV}} (\mathbf{f}_\mathrm{t} + \boldsymbol{\gamma}_\mathrm{t} - E(\mathbf{f}_\mathrm{t} \mid \boldsymbol{\gamma}_\mathrm{t})))$, and the residual $\boldsymbol{\xi}_{\mathrm{DV}}$ $(1\times \mathrm{N})$ takes the following form:

$$\boldsymbol{\xi}_{\mathrm{DV}} = \left(\frac{1}{\mathrm{T_m}} \sum_{\mathrm{t}\in\mathrm{DV}} (\mathbf{f}_\mathrm{t} + \boldsymbol{\gamma}_\mathrm{t} - E(\mathbf{f}_\mathrm{t} \mid \boldsymbol{\gamma}_\mathrm{t})) \right) (\bar{\boldsymbol{\beta}}_{\mathrm{DV}} - \hat{\boldsymbol{\beta}}_{\mathrm{EV}}) + \frac{1}{\mathrm{T_m}} \sum_{\mathrm{t}\in\mathrm{DV}} \Big((\mathbf{f}_\mathrm{t} + \boldsymbol{\gamma}_\mathrm{t} - E(\mathbf{f}_\mathrm{t} \mid \boldsymbol{\gamma}_\mathrm{t}))(\boldsymbol{\beta}_\mathrm{t} - \bar{\boldsymbol{\beta}}_{\mathrm{DV}}) \Big) + \bar{\boldsymbol{\varepsilon}}_{\mathrm{DV}}$$

where $\bar{\boldsymbol{\beta}}_{\mathrm{DV}} = \frac{1}{\mathrm{T_m}} \sum_{\mathrm{t}\in\mathrm{DV}} \boldsymbol{\beta}_\mathrm{t}$ and $\bar{\boldsymbol{\varepsilon}}_{\mathrm{DV}} = \frac{1}{\mathrm{T_m}} \sum_{\mathrm{t}\in\mathrm{DV}} \boldsymbol{\varepsilon}_\mathrm{t}$.

Recall that $\boldsymbol{\beta}_\mathrm{t} \equiv \boldsymbol{\beta} + \mathbf{u}_\mathrm{t}$, where the shock in beta is $\mathbf{u}_\mathrm{t} = [\mathbf{u}_\mathrm{t}^1, \cdots, \mathbf{u}_\mathrm{t}^\mathrm{N}]$.[5] Moreover, let $\mathbf{u}_\mathrm{t}^{\mathrm{d,i}} = \mathbf{u}_\mathrm{t}^\mathrm{i} - \frac{1}{\mathrm{T_m}} \sum_{\mathrm{s}\in\mathrm{DV}} \mathbf{u}_\mathrm{s}^\mathrm{i}$, and $\mathbf{u}_\mathrm{t}^\mathrm{d} = [\mathbf{u}_\mathrm{t}^{\mathrm{d},1}, \cdots, \mathbf{u}_\mathrm{t}^{\mathrm{d,N}}]$, which is demeaned residual in time t, and a $\mathrm{K}\times\mathrm{N}$ matrix. Decomposing $\boldsymbol{\beta}_\mathrm{t}$ into its two components, i.e., $\boldsymbol{\beta}$ and $\mathbf{u}_\mathrm{t}$, the above regression residual $\boldsymbol{\xi}_{\mathrm{DV}}$ can be re-written as

$$\begin{aligned}\boldsymbol{\xi}_{\mathrm{DV}} &= \left(\frac{1}{\mathrm{T_m}} \sum_{\mathrm{t}\in\mathrm{DV}} (\mathbf{f}_\mathrm{t} + \boldsymbol{\gamma}_\mathrm{t} - E(\mathbf{f}_\mathrm{t} \mid \boldsymbol{\gamma}_\mathrm{t})) \right) (\boldsymbol{\beta} - \hat{\boldsymbol{\beta}}_{\mathrm{EV}}) + \left(\frac{1}{\mathrm{T_m}} \sum_{\mathrm{t}\in\mathrm{DV}} (\mathbf{f}_\mathrm{t} + \boldsymbol{\gamma}_\mathrm{t} - E(\mathbf{f}_\mathrm{t} \mid \boldsymbol{\gamma}_\mathrm{t})) \right) \left(\frac{1}{\mathrm{T_m}} \sum_{\mathrm{t}\in\mathrm{DV}} \mathbf{u}_\mathrm{t} \right) \\ &+ \frac{1}{\mathrm{T_m}} \sum_{\mathrm{t}\in\mathrm{DV}} \Big((\mathbf{f}_\mathrm{t} + \boldsymbol{\gamma}_\mathrm{t} - E(\mathbf{f}_\mathrm{t} \mid \boldsymbol{\gamma}_\mathrm{t})) \mathbf{u}_\mathrm{t}^\mathrm{d} \Big) + \bar{\boldsymbol{\varepsilon}}_{\mathrm{DV}}\end{aligned},$$

Denote $\mathbf{c}_\mathrm{t}^\mathrm{i}$ as a vector of characteristics for asset i in time t. Assume that there are L characteristics, so $\mathbf{c}_\mathrm{t}^\mathrm{i}$ is a $\mathrm{L}\times 1$ vector. Similarly to time-varying betas, we assume that the characteristics can be decomposed into the two parts: $\mathbf{c}_\mathrm{t}^\mathrm{i} = \mathbf{c}^\mathrm{i} + \boldsymbol{\upsilon}_\mathrm{t}^\mathrm{i}$, where $\mathbf{c}^\mathrm{i}$ is the unconditional expected value of characteristic for asset i, and $\boldsymbol{\upsilon}_\mathrm{t}^\mathrm{i}$ is the shock in characteristic in time t. We make the following assumption on $\boldsymbol{\upsilon}_\mathrm{t}^\mathrm{i}$.

[5] Note that we use superscript i to indicate asset i in the Internet Appendix.

Assumption (C): For each asset i, the process $\{\boldsymbol{\upsilon}_t^i, t=1,\cdots,T\}$ is stationary and ergodic with zero mean and finite fourth moments.

In addition, denote $\mathbf{C}=[\mathbf{c}^1,\cdots,\mathbf{c}^N]$, an $L\times N$ matrix, as the unconditional expected value of characteristic. Take the average of characteristic from $T-T_c+1$ to T: $\overline{\mathbf{C}}\equiv\frac{1}{T_c}\sum_{t=0}^{T_c-1}\mathbf{C}_{T\text{-}t}$, with $\mathbf{C}_t=[\mathbf{c}_t^1,\cdots,\mathbf{c}_t^N]$ an $L\times N$ matrix.[6] With characteristics as control variables and additional instruments for beta estimates, we run the following cross-sectional regression: $\bar{\mathbf{r}}_{DV}=\boldsymbol{\Gamma}\hat{\mathbf{B}}_{EV}+\boldsymbol{\kappa}\overline{\mathbf{C}}+\boldsymbol{\xi}_{DV}^{\kappa}$, where the slope coefficient of characteristics, denoted by $\boldsymbol{\kappa}$, is an $1\times L$ vector. When characteristics play roles of proxies for the true factor loadings, i.e. they do not affect the cross-section of expected returns by themselves, the true value of $\boldsymbol{\kappa}$ is zero if the beta estimates are included in the regression. Under this null hypothesis, in above regression, $\boldsymbol{\xi}_{DV}^{\kappa}=\boldsymbol{\xi}_{DV}$, where $\boldsymbol{\xi}_{DV}$ is the error in the regression without characteristics. The estimated slope coefficents of our cross-sectional IV regression are given as

$$[\hat{\boldsymbol{\Gamma}},\hat{\boldsymbol{\kappa}}]'=\left(\frac{1}{N}[\hat{\mathbf{B}}_{IV};\overline{\mathbf{C}}][\hat{\mathbf{B}}_{EV};\overline{\mathbf{C}}]'\right)^{-1}\left(\frac{1}{N}[\hat{\mathbf{B}}_{IV};\overline{\mathbf{C}}]\frac{1}{T_m}\sum_{t\in DV}\mathbf{r}_t'\right),$$

where $[\hat{\mathbf{B}}_{IV};\overline{\mathbf{C}}]$ (which is a $(K+L+1)\times N$ matrix) stacks $\hat{\mathbf{B}}_{IV}$ over $\overline{\mathbf{C}}$.

In order to show the convergences of the estimated slope coefficients above, we need to specify the regularity assumptions. To simplify the notations in the assumptions, we define the following two variables:

$$\delta_s^i=-\left(\frac{1}{T_m}\sum_{t\in DV}(\mathbf{f}_t+\boldsymbol{\gamma}_t-E(\mathbf{f}_t\mid\boldsymbol{\gamma}_t))\right)\left(\left(\frac{2}{T}\mathbf{F}_{EV}^d{}'\mathbf{F}_{EV}^d\right)^{-1}(\mathbf{f}_s^d{}'e_s^{d,i})\right),^{7}$$

[6] In proposition 2, we assume that $T_c=T$, but here we relax this assumption.

[7] Assuming that T is even, $\mathbf{F}_{EV}^d\equiv[\mathbf{f}_1^d;\mathbf{f}_3^d\cdots;\mathbf{f}_{T\text{-}1}^d]$ is a $T/2\times K$ matrix when the EV period includes odd months, or $\mathbf{F}_{EV}^d\equiv[\mathbf{f}_2^d;\mathbf{f}_4^d;\cdots;\mathbf{f}_T^d]$ is a $T/2\times K$ matrix when the EV period includes even months. The superscript d

for any s in the EV period, where $e_s^i = \varepsilon_s^i + \mathbf{f}_s \mathbf{u}_s^i$, and $e_s^{d,i} = e_s^i - \frac{2}{T}\sum_{t\in EV} e_t^i$. Hence

$$[\frac{2}{T}\sum_{s\in EV}\delta_s^1, \cdots, \frac{2}{T}\sum_{s\in EV}\delta_s^N] = \left(\frac{1}{T_m}\sum_{t\in DV}(\mathbf{f}_t + \boldsymbol{\gamma}_t - E(\mathbf{f}_t \mid \boldsymbol{\gamma}_t))\right)(\boldsymbol{\beta} - \hat{\boldsymbol{\beta}}_{EV}).$$

$$\pi_s^i = \left(\frac{1}{T_m}\sum_{t\in DV}(\mathbf{f}_t + \boldsymbol{\gamma}_t - E(\mathbf{f}_t \mid \boldsymbol{\gamma}_t))\right)\mathbf{u}_s^i + (\mathbf{f}_s + \boldsymbol{\gamma}_s - E(\mathbf{f}_s \mid \boldsymbol{\gamma}_s))\mathbf{u}_s^{d,i} + \varepsilon_s^i,$$

for any s in the DV period, leading to

$$\begin{aligned}&[\frac{1}{T_m}\sum_{s=T+1}^{T+T_m}\pi_s^1, \cdots, \frac{1}{T_m}\sum_{s=T+1}^{T+T_m}\pi_s^N] = \\ &\left(\frac{1}{T_m}\sum_{t\in DV}(\mathbf{f}_t + \boldsymbol{\gamma}_t - E(\mathbf{f}_t \mid \boldsymbol{\gamma}_t))\right)\left(\frac{1}{T_m}\sum_{s\in DV}\mathbf{u}_s\right) + \frac{1}{T_m}\sum_{s\in DV}\Big((\mathbf{f}_s + \boldsymbol{\gamma}_s - E(\mathbf{f}_s \mid \boldsymbol{\gamma}_s))\mathbf{u}_s^d\Big) + \bar{\boldsymbol{\varepsilon}}_{DV}\end{aligned}.$$

Therefore, these two variables δ_s^i and π_s^i can be used to decompose the regression residual $\boldsymbol{\xi}_{DV}^{\kappa}$ in the convergence proofs below. More specifically, define $\boldsymbol{\xi}_{DV}^{\kappa} = [\xi_{DV}^{\kappa,1}, \cdots, \xi_{DV}^{\kappa,N}]$, then $\xi_{DV}^{\kappa,i} = \frac{2}{T}\sum_{s\in EV}\delta_s^i + \frac{1}{T_m}\sum_{s=T+1}^{T+T_m}\pi_s^i$. From Assumptions (A1), (A2), and (U2), we have $E(\delta_s^i) = 0$ for any s in the EV period, and $E(\pi_s^i) = 0$ for any s in the DV period. Next we define

$$\varsigma_s^i = \left[0; \left(\frac{2}{T}\mathbf{F}_{SP}^d{}'\mathbf{F}_{SP}^d\right)^{-1}(\mathbf{f}_s^d{}' e_s^{d,i})\right],$$

---

indicates a demeaned factor or demeaned residual, where average values of factors or residuals are taken over the corresponding EV period. For example, $\mathbf{f}_1^d \equiv \mathbf{f}_1 - \frac{2}{T}(\mathbf{f}_1 + \mathbf{f}_3 + \cdots + \mathbf{f}_{T-1})$, $\mathbf{f}_2^d \equiv \mathbf{f}_2 - \frac{2}{T}(\mathbf{f}_2 + \mathbf{f}_4 + \cdots + \mathbf{f}_T)$, and $\varepsilon_t^{d,i} \equiv \varepsilon_t^i - \frac{2}{T}(\varepsilon_1^i + \varepsilon_3^i + \cdots + \varepsilon_{T-1}^i)$.

a $(K+1)\times 1$ vector, where the sample period (SP) is either IV or EV period and s belongs to SP, and $e_s^{d,i} = e_s^i - \frac{2}{T}\sum_{t\in SP} e_t^i$. This variable is used to decompose the estimation errors in $\hat{\mathbf{B}}_{IV}$ and $\hat{\mathbf{B}}_{EV}$ in the convergence proofs. For example, $\hat{\mathbf{B}}_{IV} - \mathbf{B} = [\frac{2}{T}\sum_{t\in IV}\varsigma_t^1, \cdots, \frac{2}{T}\sum_{t\in IV}\varsigma_t^N]$, where $\mathbf{B} = [\mathbf{b}^1, \cdots, \mathbf{b}^N]$ is a $(K+1)\times N$ matrix with a vector of ones and unconditional expected value of beta. Similarly, we have $\hat{\mathbf{B}}_{EV} - \mathbf{B} = [\frac{2}{T}\sum_{t\in EV}\varsigma_t^1, \cdots, \frac{2}{T}\sum_{t\in EV}\varsigma_t^N]$. From Assumptions (A1), (A2), and (U2), we have $E(\varsigma_s^i) = \mathbf{0}_{(K+1)\times 1}$.

With these new variables, we describe regularity assumptions as follows:

(A4) For any i and t, the elements in both $\mathbf{b}^i$ and $\mathbf{c}^i$ are uncorrelated with δ_t^i. The elements in $\frac{1}{\sqrt{TN}}\sum_{i=1}^{N}\sum_{t\in EV}\delta_t^i$, $\frac{1}{\sqrt{TN}}\sum_{i=1}^{N}\sum_{t\in EV}(\mathbf{b}^i\delta_t^i)$ and $\frac{1}{\sqrt{TN}}\sum_{i=1}^{N}\sum_{t\in EV}(\mathbf{c}^i\delta_t^i)$ have finite variances. Given that $E(\delta_t^i) = 0$, apply Markov's Law of Large Numbers , $\frac{2}{TN}\sum_{i=1}^{N}\sum_{t\in EV}\delta_t^i \to 0$, $\frac{2}{TN}\sum_{i=1}^{N}\sum_{t\in EV}(\mathbf{b}^i\delta_t^i) \to \mathbf{0}_{(K+1)\times 1}$ and $\frac{2}{TN}\sum_{i=1}^{N}\sum_{t\in EV}(\mathbf{c}^i\delta_t^i) \to \mathbf{0}_{L\times 1}$ when both T and N approach to infinity.

(A5) For any i and t, the elements in both $\mathbf{b}^i$ and $\mathbf{c}^i$ are uncorrelated with ς_t^i. The elements in $\frac{1}{\sqrt{TN}}\sum_{i=1}^{N}\sum_{t\in EV}\varsigma_t^i$, $\frac{1}{\sqrt{TN}}\sum_{i=1}^{N}\sum_{t\in IV}\varsigma_t^i$, $\frac{1}{\sqrt{TN}}\sum_{i=1}^{N}\sum_{t\in EV}(\mathbf{b}^i(\varsigma_t^i)')$, $\frac{1}{\sqrt{TN}}\sum_{i=1}^{N}\sum_{t\in IV}(\mathbf{b}^i(\varsigma_t^i)')$, $\frac{1}{\sqrt{TN}}\sum_{i=1}^{N}\sum_{t\in EV}(\mathbf{c}^i(\varsigma_t^i)')$, and $\frac{1}{\sqrt{TN}}\sum_{i=1}^{N}\sum_{t\in IV}(\mathbf{c}^i(\varsigma_t^i)')$ have finite variances. Given that $E(\varsigma_t^i) = \mathbf{0}_{(K+1)\times 1}$, by Markov's Law of Large Numbers, $\frac{2}{TN}\sum_{i=1}^{N}\sum_{t\in EV}\varsigma_t^i \to \mathbf{0}_{(K+1)\times 1}$, $\frac{2}{TN}\sum_{i=1}^{N}\sum_{t\in IV}\varsigma_t^i \to \mathbf{0}_{(K+1)\times 1}$, $\frac{2}{TN}\sum_{i=1}^{N}\sum_{t\in EV}(\mathbf{b}^i(\varsigma_t^i)') \to \mathbf{0}_{(K+1)\times(K+1)}$, $\frac{2}{TN}\sum_{i=1}^{N}\sum_{t\in IV}(\mathbf{b}^i(\varsigma_t^i)') \to \mathbf{0}_{(K+1)\times(K+1)}$,

$\frac{2}{TN}\sum_{i=1}^{N}\sum_{t\in EV}(\mathbf{c}^{i}(\varsigma_{t}^{i})')\rightarrow\mathbf{0}_{L\times(K+1)}$ and $\frac{2}{TN}\sum_{i=1}^{N}\sum_{t\in IV}(\mathbf{c}^{i}(\varsigma_{t}^{i})')\rightarrow\mathbf{0}_{L\times(K+1)}$ when both T and N approach to infinity.

(A6) For any i and t, the elements in both $\mathbf{b}^{i}$ and $\mathbf{c}^{i}$ are uncorrelated with the elements in $\boldsymbol{\upsilon}_{t}^{i}$. The elements in $\frac{1}{\sqrt{T_c N}}\sum_{i=1}^{N}\sum_{t=T-T_c+1}^{T}\boldsymbol{\upsilon}_{t}^{i}$, $\frac{1}{\sqrt{T_c N}}\sum_{i=1}^{N}\sum_{t=T-T_c+1}^{T}(\mathbf{b}^{i}(\boldsymbol{\upsilon}_{t}^{i})')$, and $\frac{1}{\sqrt{T_c N}}\sum_{i=1}^{N}\sum_{t=T-T_c+1}^{T}(\mathbf{c}^{i}(\boldsymbol{\upsilon}_{t}^{i})')$ have finite variances. Given that $E(\boldsymbol{\upsilon}_{t}^{i})=\mathbf{0}_{L\times1}$ (Assumption (C)), by Markov's Law of Large Numbers, $\frac{1}{T_c N}\sum_{i=1}^{N}\sum_{t=T-T_c+1}^{T}\boldsymbol{\upsilon}_{t}^{i}\rightarrow\mathbf{0}_{L\times1}$, $\frac{1}{T_c N}\sum_{i=1}^{N}\sum_{t=T-T_c+1}^{T}(\mathbf{b}^{i}(\boldsymbol{\upsilon}_{t}^{i})')\rightarrow\mathbf{0}_{(K+1)\times L}$ and $\frac{1}{T_c N}\sum_{i=1}^{N}\sum_{t=T-T_c+1}^{T}(\mathbf{c}^{i}(\boldsymbol{\upsilon}_{t}^{i})')\rightarrow\mathbf{0}_{L\times L}$ when both T_c and N approach to infinity.

(A7) For any i and t, the elements in both $\mathbf{b}^{i}$ and $\mathbf{c}^{i}$ are uncorrelated with π_{t}^{i}. The elements in $\frac{1}{\sqrt{NT_m}}\sum_{i=1}^{N}\sum_{t=T+1}^{T+T_m}\pi_{t}^{i}$, $\frac{1}{\sqrt{NT_m}}\sum_{i=1}^{N}\sum_{t=T+1}^{T+T_m}(\mathbf{b}^{i}\pi_{t}^{i})$ and $\frac{1}{\sqrt{NT_m}}\sum_{i=1}^{N}\sum_{t=T+1}^{T+T_m}(\mathbf{c}^{i}\pi_{t}^{i})$ have finite variances. Given that $E(\pi_{t}^{i})=0$, by Markov's Law of Large Numbers, $\frac{1}{NT_m}\sum_{i=1}^{N}\sum_{t=T+1}^{T+T_m}\pi_{t}^{i}\rightarrow 0$, $\frac{1}{NT_m}\sum_{i=1}^{N}\sum_{t=T+1}^{T+T_m}(\mathbf{b}^{i}\pi_{t}^{i})\rightarrow\mathbf{0}_{(K+1)\times1}$, and $\frac{1}{NT_m}\sum_{i=1}^{N}\sum_{t=T+1}^{T+T_m}(\mathbf{c}^{i}\pi_{t}^{i})\rightarrow\mathbf{0}_{L\times1}$ when both T_m and N approach to infinity.

(A8) When N→∞, $\frac{1}{N}[\mathbf{B};\mathbf{C}][\mathbf{B};\mathbf{C}]'$ converges to an invertible matrix, denoted by $\begin{bmatrix}\mathbf{bb'} & \mathbf{bc'}\\ \mathbf{cb'} & \mathbf{cc'}\end{bmatrix}$, which is $E\left[\begin{bmatrix}\mathbf{b}^{i}(\mathbf{b}^{i})' & \mathbf{b}^{i}(\mathbf{c}^{i})'\\ \mathbf{c}^{i}(\mathbf{b}^{i})' & \mathbf{c}^{i}(\mathbf{c}^{i})'\end{bmatrix}\right]$.

With these assumptions, the convergence of the IV mean-estimator is established in the following Theorem.

Theorem G Suppose that Assumptions (A1), (A2), (A4)-(A8), (U2) and (C) in Appendices E, F, and G hold, Then $[\hat{\boldsymbol{\Gamma}},\hat{\boldsymbol{\kappa}}]'$ converges to $(\boldsymbol{\Gamma},\mathbf{0}_{1\times L})'$ when N, $\mathrm{T_m}$, $\mathrm{T_c}$, and T approach to infinity■

Proof: Note that

$$[\hat{\boldsymbol{\Gamma}},\hat{\boldsymbol{\kappa}}]'-(\boldsymbol{\Gamma},\mathbf{0}_{1\times L})'=\left(\frac{1}{N}[\hat{\mathbf{B}}_{IV};\overline{\mathbf{C}}][\hat{\mathbf{B}}_{EV};\overline{\mathbf{C}}]'\right)^{-1}\left(\frac{1}{N}[\hat{\mathbf{B}}_{IV};\overline{\mathbf{C}}](\xi_{DV}^{\kappa})'\right),$$

where $\xi_{DV}^{\kappa}=[\xi_{DV}^{\kappa,1},\cdots,\xi_{DV}^{\kappa,N}]$ and $\xi_{DV}^{\kappa,i}=\frac{2}{T}\sum_{s\in EV}\delta_s^i+\frac{1}{T_m}\sum_{s=T+1}^{T+T_m}\pi_s^i$.

We want to show that the above equation converges to $\mathbf{0}_{(K+L+1)\times 1}$, which requires to show the following three convergences:

$$\frac{1}{N}\hat{\mathbf{B}}_{IV}(\xi_{DV}^{\kappa})'=\frac{1}{NT_m}\sum_{i=1}^{N}\sum_{t=T+1}^{T+T_m}\left((\mathbf{b}^i+\frac{2}{T}\sum_{s\in IV}\varsigma_s^i)\pi_t^i\right)+\frac{2}{NT}\sum_{i=1}^{N}\sum_{t\in EV}\left((\mathbf{b}^i+\frac{2}{T}\sum_{s\in IV}\varsigma_s^i)\delta_t^i\right)\to\mathbf{0}_{(K+1)\times 1},^{8}$$

$$\frac{1}{N}\overline{\mathbf{C}}(\xi_{DV}^{\kappa})'=\frac{1}{NT_m}\sum_{i=1}^{N}\sum_{t=T+1}^{T+T_m}\left((\mathbf{c}^i+\frac{1}{T_c}\sum_{s=T-T_c+1}^{T}\boldsymbol{\upsilon}_s^i)\pi_t^i\right)+\frac{2}{NT}\sum_{i=1}^{N}\sum_{t\in EV}\left((\mathbf{c}^i+\frac{1}{T_c}\sum_{s=T-T_c+1}^{T}\boldsymbol{\upsilon}_s^i)\delta_t^i\right)\to\mathbf{0}_{L\times 1},$$

and $(\frac{1}{N}[\hat{\mathbf{B}}_{IV};\overline{\mathbf{C}}][\hat{\mathbf{B}}_{EV};\overline{\mathbf{C}}]')^{-1}$ converges to $\begin{bmatrix}\mathbf{bb'} & \mathbf{bc'}\\ \mathbf{cb'} & \mathbf{cc'}\end{bmatrix}^{-1}$.

Those three convergences can be shown as follows: From Assumption (A4), $\frac{2}{TN}\sum_{i=1}^{N}\sum_{t\in EV}(\mathbf{b}^i\delta_t^i)\to\mathbf{0}_{(K+1)\times 1}$ and $\frac{2}{TN}\sum_{i=1}^{N}\sum_{t\in EV}\delta_t^i\to 0$, when N and T approach to infinity. From Assumptions (A1) and (A2) in Appendix E, and Assumption (U2) in Appendix F, $\frac{2}{T}\sum_{s\in IV}\varsigma_s^i\to\mathbf{0}_{(K+1)\times 1}$ for any asset i, as T approaches to infinity. Together with $\frac{2}{TN}\sum_{i=1}^{N}\sum_{t\in EV}\delta_t^i\to 0$,

[8] Recall that $\hat{\mathbf{B}}_{IV}-\mathbf{B}=[\frac{2}{T}\sum_{t\in IV}\varsigma_t^1,\cdots,\frac{2}{T}\sum_{t\in IV}\varsigma_t^N]$, and $\mathbf{B}=[\mathbf{b}^1,\cdots,\mathbf{b}^N]$.

this implies that when N and T approach to infinity at the same time, $\frac{2}{NT}\sum_{i=1}^{N}\sum_{t\in EV}\left((\frac{2}{T}\sum_{s\in IV}\varsigma_s^i)\delta_t^i\right)\to\mathbf{0}_{(K+1)\times 1}$. Therefore, when T and N approach to infinity,

$$\frac{2}{NT}\sum_{i=1}^{N}\sum_{t\in EV}\left((\mathbf{b}^i+\frac{2}{T}\sum_{s\in IV}\varsigma_s^i)\delta_t^i\right)=\frac{2}{NT}\sum_{i=1}^{N}\sum_{t\in EV}(\mathbf{b}^i\delta_t^i)+\frac{2}{NT}\sum_{i=1}^{N}\sum_{t\in EV}\left((\frac{2}{T}\sum_{s\in IV}\varsigma_s^i)\delta_t^i\right)\to\mathbf{0}_{(K+1)\times 1}.$$

Similarly, from Assumption (A7), $\frac{1}{NT_m}\sum_{i=1}^{N}\sum_{t=T+1}^{T+T_m}(\mathbf{b}^i\pi_t^i)\to\mathbf{0}_{(K+1)\times 1}$ and $\frac{1}{NT_m}\sum_{i=1}^{N}\sum_{t=T+1}^{T+T_m}\pi_t^i\to 0$ when both T_m and N approach to infinity, and $\frac{2}{T}\sum_{s\in IV}\varsigma_s^i\to\mathbf{0}_{(K+1)\times 1}$ for any asset i, as T approaches to infinity. Together with $\frac{1}{NT_m}\sum_{i=1}^{N}\sum_{t=T+1}^{T+T_m}\pi_t^i\to 0$, this implies that when N, T_m, and T approach to infinity at the same time, $\frac{1}{NT_m}\sum_{i=1}^{N}\sum_{t=T+1}^{T+T_m}\left((\frac{2}{T}\sum_{s\in IV}\varsigma_s^i)\pi_t^i\right)\to\mathbf{0}_{(K+1)\times 1}$. Thus, when T , T_m and N approach to infinity,

$$\frac{1}{NT_m}\sum_{i=1}^{N}\sum_{t=T+1}^{T+T_m}\left((\mathbf{b}^i+\frac{2}{T}\sum_{s\in IV}\varsigma_s^i)\pi_t^i\right)=\frac{1}{NT_m}\sum_{i=1}^{N}\sum_{t=T+1}^{T+T_m}\left(\mathbf{b}^i\pi_t^i\right)+\frac{1}{NT_m}\sum_{i=1}^{N}\sum_{t=T+1}^{T+T_m}\left((\frac{2}{T}\sum_{s\in IV}\varsigma_s^i)\pi_t^i\right)\to\mathbf{0}_{(K+1)\times 1}.$$

Hence, $\frac{1}{N}\hat{\mathbf{B}}_{IV}(\xi_{DV}^{\kappa})'\to\mathbf{0}_{(K+1)\times 1}$

Similarly, we can show that $\frac{1}{N}\overline{\mathbf{C}}(\xi_{DV}^{\kappa})'\to\mathbf{0}_{L\times 1}$ with Assumptions (A4), (A7), (C), as well as $E(\pi_t^i)=0$ and $E(\delta_t^i)=0$.

Moreover,

$$\frac{1}{N}\hat{\mathbf{B}}_{IV}\hat{\mathbf{B}}_{EV}'=\frac{1}{N}\mathbf{BB}'+\frac{2}{TN}\sum_{i=1}^{N}\sum_{t\in EV}(\mathbf{b}^i(\varsigma_t^i)')+\frac{2}{TN}\sum_{i=1}^{N}\sum_{t\in IV}(\mathbf{b}^i(\varsigma_t^i)')'+\frac{1}{N}\sum_{i=1}^{N}\left(\frac{2}{T}\sum_{t\in IV}(\varsigma_t^i)\frac{2}{T}\sum_{s\in EV}(\varsigma_s^i)'\right).$$

From Assumption (A5), $\frac{2}{TN}\sum_{i=1}^{N}\sum_{t\in EV}(\mathbf{b}^{i}(\varsigma_t^{i})')\to\mathbf{0}_{(K+1)\times(K+1)}$ and $\frac{2}{TN}\sum_{i=1}^{N}\sum_{t\in IV}(\mathbf{b}^{i}(\varsigma_t^{i})')\to\mathbf{0}_{(K+1)\times(K+1)}$ when both T and N approach to infinity. From Assumptions (A1), (A2) and (U2), $\frac{2}{T}\sum_{t\in IV}\varsigma_t^{i}\to\mathbf{0}_{(K+1)\times 1}$ for any asset i, when T approaches to infinity. Also from Assumption (A5), $\frac{2}{TN}\sum_{i=1}^{N}\sum_{s\in EV}(\varsigma_s^{i})'\to\mathbf{0}_{1\times(K+1)}$ when both T and N approach to infinity. The above two convergences imply that when both T and N approach to infinity, $\frac{1}{N}\sum_{i=1}^{N}\left(\frac{2}{T}\sum_{t\in IV}(\varsigma_t^{i})\frac{2}{T}\sum_{s\in EV}(\varsigma_s^{i})'\right)\to\mathbf{0}_{(K+1)\times(K+1)}$. Therefore, with Assumption (A8), it is clear that $\frac{1}{N}\hat{\mathbf{B}}_{IV}\hat{\mathbf{B}}_{EV}'\to\mathbf{bb'}$ when both T and N approach to infinity.

Similarly, $\frac{1}{N}\hat{\mathbf{B}}_{IV}\overline{\mathbf{C}}'\to\mathbf{bc'}$, $\frac{1}{N}\overline{\mathbf{C}}\hat{\mathbf{B}}_{EV}'\to\mathbf{cb'}$, and $\frac{1}{N}\overline{\mathbf{C}}\overline{\mathbf{C}}'\to\mathbf{cc'}$ when both T and N approach to infinity, with Assumptions (A5), (A6), (A8), (U2) and (C); hence, $(\frac{1}{N}[\hat{\mathbf{B}}_{IV};\overline{\mathbf{C}}][\hat{\mathbf{B}}_{EV};\overline{\mathbf{C}}]')^{-1}$ converges to $\begin{bmatrix}\mathbf{bb'} & \mathbf{bc'}\\ \mathbf{cb'} & \mathbf{cc'}\end{bmatrix}^{-1}$. Together with $\frac{1}{N}\hat{\mathbf{B}}_{IV}(\xi_{DV}^{\kappa})'\to\mathbf{0}_{(K+1)\times 1}$ and $\frac{1}{N}\overline{\mathbf{C}}(\xi_{DV}^{\kappa})'\to\mathbf{0}_{L\times 1}$, we establish $[\hat{\mathbf{\Gamma}},\hat{\boldsymbol{\kappa}}]'-(\mathbf{\Gamma},\mathbf{0}_{1\times L})'\to\mathbf{0}_{(K+L+1)\times 1}$ (i.e. $[\hat{\mathbf{\Gamma}},\hat{\boldsymbol{\kappa}}]'$ converges to $(\mathbf{\Gamma},\mathbf{0}_{1\times L})'$) when N, T_m, T_c, and T approach to infinity. ■

In Theorem G, we assume that betas and characteristics can be any stationary and ergodic processes; hence, Proposition 2 in the main text is a special case under AR(1) processes. The regularity Assumptions (A4)-(A8) are satisfied when (a) Processes $\{\varepsilon_t^{i},t=1,\cdots,T\}$, $\{\boldsymbol{\upsilon}_t^{i},t=1,\cdots,T\}$ and $\{\mathbf{u}_t^{i},t=1,\cdots,T\}$ for each asset i are stationary and ergodic, (b) For all t, s and i, $\varepsilon_t^{i},\boldsymbol{\upsilon}_t^{i}$ and $\mathbf{u}_t^{i}$ are independent of $\mathbf{f}_s$ (factor in time s), ε_s^{i} (regression residuals in time s), unconditional means of beta $\boldsymbol{\beta}^{i}$ and characteristic $\mathbf{c}^{i}$, and $\boldsymbol{\gamma}_s$ (risk premium in time s), and the maximum values for unconditional mean of beta and characteristic of all assets are finite. (c) In each time t, residuals in

$[\varepsilon_t^1,\cdots,\varepsilon_t^N]$, $[\mathbf{v}_t^1,\cdots,\mathbf{v}_t^N]$ and $[\mathbf{u}_t^1,\cdots,\mathbf{u}_t^N]$ are asymptotically weakly correlated cross-sectionally (Shanken,1992).