Online Appendix for
The impact of uncertainty on investment: Empirical challenges and a new estimator

The online Appendix contains five sections. Section A provides details on how to construct orthogonal moment conditions to treat measurement error in q. Section B gives the assumptions and limiting results for our proposed estimators. Appendix C explains how to construct nonparametric model specification test statistics. Appendix D describes how to construct the 10 instrumental variables for the endogenous individual stock volatility variable. Section E shows more robust results. Section F provides the mathematical proofs of the theorems in Section B.

Appendix A. Treating measurement error

In this section, we provide details on how to construct orthogonal moment conditions to treat measurement error in q. For ease in reading, we introduce our notation here. (i) $\mathbf{i}_T$ denotes a $T \times 1$ vector of ones, 0_n denotes an $n \times 1$ vector of zeros, $0_{m \times n}$ is an $m \times n$ matrix of zeros, I_T is a $T \times T$ identity matrix, and $J_T = I_T - T^{-1}\mathbf{i}_T\mathbf{i}_T'$; (ii) $\tilde{\mathbf{a}}_i = J_T\mathbf{a}_i$ denotes the demeaned data for any $T \times 1$ vector $\mathbf{a}_i$; (iii) M, M_1, M_2,... are constants that can take different values at different locations. Denoting a $T \times (3k_{nT})$ matrix $\mathbf{E}_i = [\mathbf{E}_{i,1}, \ldots, \mathbf{E}_{i,T}]'$, rewrite model (14) in matrix form as follows

$$\mathbf{y}_i \approx \mu_i \mathbf{i}_T + \boldsymbol{\lambda}_0 + \alpha_0 \mathbf{x}_{i,-1} + \mathbf{E}_i \boldsymbol{\vartheta}_0 + \boldsymbol{\varepsilon}_i, \tag{A.1}$$

and premultiplying J_T to both sides of model (A.1) gives

$$\tilde{\mathbf{y}}_i \approx J_T \boldsymbol{\lambda}_0 + \alpha_0 \tilde{\mathbf{x}}_{i,-1} + \tilde{\mathbf{E}}_i \boldsymbol{\vartheta}_0 + \tilde{\boldsymbol{\varepsilon}}_i \tag{A.2}$$

$$= F_T \tilde{\boldsymbol{\lambda}}_0 + \alpha_0 \tilde{\mathbf{x}}_{i,-1} + \tilde{\mathbf{E}}_i \boldsymbol{\vartheta}_0 + \tilde{\boldsymbol{\varepsilon}}_i \tag{A.3}$$

where $\tilde{\mathbf{y}}_i = J_T \mathbf{y}_i$, $\tilde{\mathbf{E}}_i = J_T \mathbf{E}_i$ and $\tilde{\boldsymbol{\varepsilon}}_i = J_T \boldsymbol{\varepsilon}_i$ are all demeaned data. Because J_T does not have full rank, we redefine $J_T \boldsymbol{\lambda}_0 = F_T \tilde{\boldsymbol{\lambda}}_0$, where $F_T = [I_{T-1}, -\mathbf{i}_{T-1}]'$, $\tilde{\boldsymbol{\lambda}}_0 = \left[\tilde{\lambda}_1, \ldots, \tilde{\lambda}_{T-1}\right]'$, and $\tilde{\lambda}_t = \lambda_t - T^{-1} \sum_{s=1}^{T} \lambda_s$.

In the following, we construct two blocks of orthogonal moment conditions, in the spirit of Meijer, Spierdijk and Wansbeek (2017). The moments in Section A2 are sufficient to identify our model, and we use them to produce the empirical results in this paper. In contrast, the moments in Section A1 are empirically optional, and we include them in the econometric theory for completeness.

A1. Intertemporal covariance matrix

These moment conditions are based on the intertemporal covariance matrix of $\boldsymbol{\omega}_i$ in model (10), $\Sigma_\omega = \mathbb{E}\left(\boldsymbol{\omega}_i \boldsymbol{\omega}_i'\right)$, by exploiting the cross-sectional independence across firms:

$$\begin{aligned}
\mathbb{E}\left(\tilde{\boldsymbol{\varepsilon}}_i \otimes \tilde{\mathbf{y}}_i\right) &\approx \mathbb{E}\left[\left(\tilde{\boldsymbol{\omega}}_i - \alpha_0 \tilde{\mathbf{e}}_{i,-1}\right) \otimes \left(F_T \tilde{\boldsymbol{\lambda}}_0 + \alpha_0 \tilde{\mathbf{x}}_{i,-1}^{*} + \tilde{\mathbf{E}}_i \boldsymbol{\vartheta}_0 + \tilde{\boldsymbol{\omega}}_i\right)\right] \\
&= \mathbb{E}\left(\tilde{\boldsymbol{\omega}}_i \otimes \tilde{\boldsymbol{\omega}}_i\right) = \mathbb{E}\left[\operatorname{vec}\left(J_T \boldsymbol{\omega}_i \boldsymbol{\omega}_i' J_T\right)\right] = \left(J_T \otimes J_T\right) \mathbb{E}\left[\operatorname{vec}\left(\boldsymbol{\omega}_i \boldsymbol{\omega}_i'\right)\right] \\
&= \left(J_T \otimes J_T\right) \operatorname{vec}\left(\Sigma_\omega\right) = \left(J_T \otimes J_T\right) D_T \pi_\omega \qquad \text{(A.4)}
\end{aligned}$$

where π_ω =vech(Σ_ω), and D_T is the duplication matrix of dimension $T^2 \times m_0$, with $m_0 = T(T+1)/2$, such that $\operatorname{vec}(\Sigma_\omega) = D_T \pi_\omega$.[1] Let $\left[\left(J_T \otimes J_T\right) D_T\right]_\perp$ be the orthogonal complement of $\left(J_T \otimes J_T\right) D_T$, i.e., the $T^2 \times m_0$ matrix with full column rank satisfying

$$\left[\left(J_T \otimes J_T\right) D_T\right]_\perp' \left[\left(J_T \otimes J_T\right) D_T\right] = 0_{m_0 \times m_0}.$$

Pre-multiplying both sides of (A.4) by $\left[\left(J_T \otimes J_T\right) D_T\right]_\perp'$ gives

$$\left[\left(J_T \otimes J_T\right) D_T\right]_\perp' \mathbb{E}\left(\tilde{\boldsymbol{\varepsilon}}_i \otimes \tilde{\mathbf{y}}_i\right) \approx 0_{m_0}.$$

which is equivalent to[2]

$$\left[\left(J_T \otimes J_T\right) D_T\right]_\perp' \mathbb{E}\left[\left(I_T \otimes \tilde{\mathbf{y}}_i\right) \tilde{\boldsymbol{\varepsilon}}_i\right] \approx 0_{m_0}. \qquad \text{(A.5)}$$

As a result, the $(T \times m_0)$ matrix $\tilde{\mathbf{d}}_{1,i} = \left(I_T \otimes \tilde{\mathbf{y}}_i\right)' \left[\left(J_T \otimes J_T\right) D_T\right]_\perp$ acts as valid instru-

[1] If $A = (a_{ij})$ is a 3×3 symmetric matrix, $\operatorname{vec}(A) = [a_{11}, a_{12}, a_{13}, a_{21}, a_{22}, a_{23}, a_{31}, a_{32}, a_{33}]'$ and $\operatorname{vech}(A) = [a_{11}, a_{12}, a_{13}, a_{22}, a_{23}, a_{33}]'$.

[2] For any vectors a $(m \times 1)$ and b $(n \times 1)$, $a \otimes b = (I_m \otimes b)\, a$.

ments that are orthogonal to $\tilde{\boldsymbol{\varepsilon}}_i$. Moreover, $\tilde{\mathbf{d}}_{1,i}$ are relevant instruments if

$$\mathbb{E}\left(\tilde{\mathbf{d}}_{1,i}'\tilde{\mathbf{x}}_{i,-1}\right) \neq 0_{m_0} \tag{A.6}$$

which generally holds true as $[(J_T \otimes J_T)\, D_T]_{\perp}' \, \mathbb{E}\left(\tilde{\mathbf{x}}_{i,-1} \otimes \left(\tilde{\mathbf{E}}_i \boldsymbol{\vartheta}_0\right)\right) \neq 0_{m_0}$ if $\boldsymbol{\vartheta}_0 \neq 0_{3k}$ and $\tilde{\mathbf{x}}_{i,-1}$ and $\tilde{\mathbf{E}}_i$ are correlated.

A2. Exogenous regressors

These orthogonal moment conditions are based on the strict exogeneity of time-fixed effects and $\tilde{\mathbf{E}}_i$ in model (A.3), and include

$$\mathbb{E}\left[(\tilde{\mathbf{y}}_i - \tilde{\mathbf{q}}_i\boldsymbol{\theta}_0) \otimes \tilde{\mathbf{E}}_i\right] \approx \mathbb{E}\left(\tilde{\boldsymbol{\varepsilon}}_i \otimes \tilde{\mathbf{E}}_i\right) = 0_{T^2 \times (3k_{nT})} \tag{A.7}$$

$$\mathbb{E}\left[(\tilde{\mathbf{y}}_i - \tilde{\mathbf{q}}_i\boldsymbol{\theta}_0) \otimes F_T\right] \approx \mathbb{E}\left(\tilde{\boldsymbol{\varepsilon}}_i \otimes F_T\right) = 0_{T^2 \times (T-1)} \tag{A.8}$$

where $\tilde{\mathbf{q}}_i = \left[F_T, \tilde{\mathbf{x}}_{i,-1}, \tilde{\mathbf{E}}_i\right]$ contains all the regressors in model (A.3) and $\boldsymbol{\theta}_0 = \left[\tilde{\boldsymbol{\lambda}}_0', \alpha_0, \boldsymbol{\vartheta}_0'\right]'$ includes all the parameters to be estimated. These moment conditions are valid because after the control function approach, the only endogenous variable in model (A.3) is $\tilde{\mathbf{x}}_{i,-1}$. Consequently, denoting a $T \times [T^2 (T - 1 + 3k_{nT})]$ matrix

$$\tilde{\mathbf{d}}_{2,i} = \left[\begin{array}{cc} I_T \otimes \mathrm{vec}\left(\tilde{\mathbf{E}}_i\right)' & I_T \otimes \mathrm{vec}\left(F_T\right)' \end{array}\right]$$

we obtain the following orthogonal moment conditions

$$\mathbb{E}\left(\tilde{\mathbf{d}}_{2,i}'\tilde{\boldsymbol{\varepsilon}}_i\right) \approx 0_{T^2(T-1+3k_{nT})}. \tag{A.9}$$

Moreover, $\tilde{\mathbf{d}}_{2,i}$ are relevant instruments if $\mathbb{E}\left(\tilde{\mathbf{d}}_{2,i}'\tilde{\mathbf{x}}_{i,-1}\right) \neq 0$, which can be easily tested in the data. We further select a smaller set of instruments based on the method developed in Belloni, Chen, Chernozhukov and Hansen (2012) to mitigate the concern of many-instruments bias in a finite sample.

Appendix B. Limiting results

For a sufficiently large n and a fixed T, we show three theorems in the following (with proofs in the Online Appendix), ensuring that both the penalty estimator, $\hat{\boldsymbol{\theta}}$, and the post-penalty estimator, $\tilde{\boldsymbol{\theta}}$, converge to the true parameter values $\boldsymbol{\theta}_0$. The estimators of the Tobin's q coefficient and the three unknown curves are consistent and have asymptotic normal distributions. Theorem B.1 proves that if we knew which elements in $\boldsymbol{\theta}_0$ were zero, then dropping the corresponding regressors and conducting traditional GMM estimation by minimizing (18) in Li and Sun (2022) would produce a uniformly consistent estimator for the nonzero elements in $\boldsymbol{\theta}_0$. At the same time, the estimators of α_0, $f_0(z)$, $g_0(s)$, and $r_0(v)$ possess asymptotic normal distributions. Theorem B.2 demonstrates that for any given tuning parameter ψ, there exists a penalty estimator $\hat{\boldsymbol{\theta}}(\psi)$ that minimizes the objective function (19) in Li and Sun (2022). For sufficiently large n and with a probability approaching 1, this penalty estimator equals the traditional GMM estimator in Theorem B.1 as if we knew which elements in $\boldsymbol{\theta}_0$ were zero. That is, the penalty estimator is *oracle efficient*. Theorem B.3 reveals that the post-penalty estimator $\tilde{\boldsymbol{\theta}}$ that solves (20) in Li and Sun (2022) is same as $\hat{\boldsymbol{\theta}}$ asymptotically with a probability approaching 1.

In Appendix A, we construct the following instruments to deal with the error-ridden q:

$$\tilde{\mathbf{d}}_i = [\tilde{\mathbf{d}}_{1,i}\ \tilde{\mathbf{d}}_{2,i}] = \left[\ (I_T \otimes \tilde{\mathbf{y}}_i)' \left[(J_T \otimes J_T) D_T\right]_{\perp} \quad I_T \otimes \text{vec}\left(\tilde{\mathbf{E}}_i\right)' \quad I_T \otimes \text{vec}\left(F_T\right)'\ \right], \tag{B.1}$$

which is a $T \times m_{iv}$ matrix, with the number of instruments equal to

$$m_{iv} = m_0 + T^2 \left(T - 1 + 3k_{nT}\right).$$

We then obtain the following orthogonal moment conditions (equivalent to (17) in Li and Sun (2022)):

$$\mathbb{E}\left[\tilde{\mathbf{d}}_i' \left(\tilde{\mathbf{y}}_i - \tilde{\mathbf{q}}_i \boldsymbol{\theta}_0\right)\right] \approx 0_{m_{iv}}. \tag{B.2}$$

Decompose $\boldsymbol{\theta}_0 = \left[\boldsymbol{\theta}_{1,0}', 0_{p-J}'\right]'$, where $\boldsymbol{\theta}_{1,0}'$ is the $J \times 1$ vector containing all the non-zero parameters, and J is the number of elements in $\mathcal{J} = supp(\boldsymbol{\theta}_0)$. Divide the regressors accordingly into two parts, $\tilde{\mathbf{q}}_i = [\tilde{\mathbf{q}}_{1,i}, \tilde{\mathbf{q}}_{2,i}]$, where $\tilde{\mathbf{q}}_{1,i}$ and $\tilde{\mathbf{q}}_{2,i}$ are $T \times J$ and $T \times (p - J)$ matrix, respectively. Our orthogonal moment conditions in (B.2) can be rewritten in this partition by removing all the regressors with zero coefficients, as $\tilde{\mathbf{q}}_i \boldsymbol{\theta}_0 = \tilde{\mathbf{q}}_{1,i} \boldsymbol{\theta}_{1,0}$. Let $\check{\boldsymbol{\theta}}_1$ be the estimator of $\boldsymbol{\theta}_{1,0}$, which solves the following optimization:

$$\min_{\boldsymbol{\theta}_1 \in \Theta_1} \bar{\boldsymbol{\varphi}}_n \left(\boldsymbol{\theta}_1\right)' \boldsymbol{\Omega}_n \bar{\boldsymbol{\varphi}}_n \left(\boldsymbol{\theta}_1\right), \tag{B.3}$$

where Θ_1 is a compact subset of R^J, and

$$\bar{\varphi}_n \left(\boldsymbol{\theta}_1\right) = \frac{1}{n} \sum_{i=1}^{n} \widetilde{\tilde{\mathbf{d}}}_i' \left(\tilde{\mathbf{y}}_i - \widetilde{\tilde{\mathbf{q}}}_{1,i} \boldsymbol{\theta}_1\right), \tag{B.4}$$

where $\widetilde{\tilde{\mathbf{q}}}_{1,i}$ equals $\tilde{\mathbf{q}}_{1,i}$ with $\mathbf{P}^{k_{nT}}\left(v_{i,t}\right)$ replaced by $\mathbf{P}^{k_{nT}}\left(\hat{v}_{i,t}\right)$ for all t. Let $P_{\mathcal{J}}$ be the $J \times p$

selection matrix that satisfies $\boldsymbol{\theta}_0 = P'_{\mathcal{J}}\boldsymbol{\theta}_{1,0}$.[3] Consequently, $\check{\boldsymbol{\theta}} = P'_{\mathcal{J}}\check{\boldsymbol{\theta}}_1$ is an estimator for $\boldsymbol{\theta}_0$, when $\mathcal{J}$ is known. We list additional regularity conditions that support the limiting results in the following two assumptions.

Assumption 3. (i) $\mathbb{E}\left\|T^{-1}\mathbf{W}_i\mathbf{W}_i'\right\| \leq M$; also, $T^{-1}\mathbb{E}\left((\mathbf{i}_T, \mathbf{W}_i)'(\mathbf{i}_T, \mathbf{W}_i)\right)$, Σ_v, and $T^{-1}\mathbb{E}\left(\mathbf{W}_i\Sigma_v\mathbf{W}_i'\right)$ are all positive definite matrix; (ii) There exist two constants, ς_0 and ς_1, satisfying

$$0 < \varsigma_0 \leq \lambda_{\min}\left(\mathbb{E}\left(\tilde{\mathbf{q}}_i'\tilde{\mathbf{d}}_i\right)\boldsymbol{\Omega_n}\mathbb{E}\left(\tilde{\mathbf{d}}_i'\tilde{\mathbf{q}}_i\right)\right) \leq \lambda_{\max}\left(\mathbb{E}\left(\tilde{\mathbf{q}}_i'\tilde{\mathbf{d}}_i\right)\boldsymbol{\Omega_n}\mathbb{E}\left(\tilde{\mathbf{d}}_i'\tilde{\mathbf{q}}_i\right)\right) \leq \varsigma_1 < \infty,$$

where $\lambda_{\min}(\cdot)$ and $\lambda_{\max}(\cdot)$ denote the smallest and largest eigenvalues of a matrix, respectively; (iii) $\boldsymbol{\Omega}_n$ is a non-stochastic positive definite matrix with $\lambda_{\max}(\boldsymbol{\Omega}_n) \leq M < \infty$.

Assumption 4. As $n \to \infty$, the following conditions hold: (i) $k_{nT} \to \infty$; (ii) $\sqrt{n}k_{nT}^{-\zeta} \to 0$; (iii) $k_{nT}^2/n \to 0$; and (iv) $n^{-1}k_{nT}\left\|\mathbf{P}^{k_{nT}}\right\|_1^2 \to 0$, where $\left\|\mathbf{P}^{k_{nT}}\right\|_l = \max_{0\leq j\leq l}\sup_{x\in R}\left\|\frac{\partial^j\mathbf{P}^{k_{nT}}(x)}{\partial x^j}\right\|$ for $l \geq 0$, with $\|\cdot\|$ being the Euclidean norm.

Assumption 3(i) regulates the stochastic property of $(\mathbf{w}_{it}, v_{it})$ in the time dimension to ensure that $\boldsymbol{\pi}_0$ in model (6) in Li and Sun (2022) can be estimated at the root-n convergence rate so that the estimated residuals $\hat{v}_{i,t}$ have stochastic properties mimicking those of the true error terms $v_{i,t}$. Assumption 3 (ii) is standard in the literature on series approximation, ensuring the existence of the proposed estimators with nonsingular variance and covariance matrices. In Assumption 4, (i) and (iii) are standard conditions in series approximation, (ii) is a technical condition, and (iv) is used to remove the asymptotic impact of the first-stage estimation of $\boldsymbol{\pi}_0$ on the estimation of model (13) in Li and Sun (2022). For the smoothing parameter, the existing literature commonly sets $k_{nT} = c_k(nT)^r$, where $c_k > 0$ and $r > 0$ are constants. In this setup, Assumption 4 requires that $(2\zeta)^{-1} < r < 0.25$ if the Hermite series are adopted in the series approximation because $\left\|\mathbf{P}^{k_{nT}}\right\|_1 = O\left(k_{nT}^{3/2}\right)$.

Theorem B.1 *Denoting $a_n = k_{nT}^{-\zeta} + \sqrt{J/n}$, under Assumptions 1-4, we have*

(i) $\left\|\check{\boldsymbol{\theta}}_1 - \boldsymbol{\theta}_{1,0}\right\| = O_p(a_n)$;

(ii)

$$\sup_{z\in\mathcal{S}_z}\left\|\check{f}(z) - f_0(z)\right\| = O_p\left(a_n k_{nT}^{1/2}\right),$$

$$\sup_{s\in\mathcal{S}_s}\left\|\check{g}(s) - g_0(s)\right\| = O_p\left(a_n k_{nT}^{1/2}\right),$$

$$\sup_{v\in\mathcal{S}_v}\left\|\check{r}(v) - r_0(v)\right\| = O_p\left(a_n k_{nT}^{1/2}\right),$$

where $\check{f}(z) = \check{\boldsymbol{\theta}}_z'\vec{\mathbf{P}}_z^{k_{nT}}$, $\check{g}(s) = \check{\boldsymbol{\theta}}_s'\vec{\mathbf{P}}_s^{k_{nT}}$, *and* $\check{r}(v) = \check{\boldsymbol{\theta}}_v'\vec{\mathbf{P}}_v^{k_{nT}}$, *with* $\vec{\mathbf{P}}^{k_{nT}}(\cdot)$ *equal to* $\mathbf{P}^{k_{nT}}(\cdot)$

[3] $P_{\mathcal{J}}$ is the $J \times p$ sub-matrix of the identity matrix I_p that satisfies $\boldsymbol{\theta}_{1,0} = P_{\mathcal{J}}\boldsymbol{\theta}_0$ and $P'_{\mathcal{J}}P_{\mathcal{J}} = I_J$.

after removing the terms with zero coefficients, and $\mathcal{S}_z$*,* $\mathcal{S}_s$*, and* $\mathcal{S}_v$ *denote the support of* z_{it}*,* s_{it}*, and* v_{it}*, respectively; (iii) if* $\sqrt{n/J}k_{nT}^{-\zeta} \to 0$*, we have* $\sqrt{n/\hat{\sigma}_{\alpha}^2}\left(\check{\alpha} - \alpha_0\right) \xrightarrow{d} N\left(0, 1\right)$ *and*

$$\sqrt{n\hat{\Xi}_n^{-1}} \begin{bmatrix} \check{f}\left(z\right) - f_0\left(z\right) \\ \check{g}\left(s\right) - g_0\left(s\right) \\ \check{r}\left(v\right) - r_0\left(v\right) \end{bmatrix} \xrightarrow{d} N\left(0_3, I_3\right),$$

provided that $\hat{\sigma}_{\alpha}^2 \xrightarrow{p} \sigma_{\alpha}^2 > 0$ *and that* $\hat{\Xi}_n^{-1}$ *is nonsingular, where* $\hat{\sigma}_{\alpha}^2$*,* σ_{α}^2*, and* $\hat{\Xi}_n$ *are denoted in (F.13) and (F.14) in the Online Appendix.*

When $\mathcal{J}$ is known, Theorem B.1(i) shows the uniform consistency of $\check{\boldsymbol{\theta}}$. Meanwhile, the estimated coefficient $\check{\alpha}$ is a root-n consistent estimator of α_0. The pointwise convergence rate of $\check{f}\left(z\right)$, $\check{g}\left(s\right)$, and $\check{r}\left(v\right)$ are of order $O_p\left(\sqrt{J/n}\right)$ because $\left\|\hat{\Xi}_n\right\| = O_p\left(J\right)$.

Theorem B.2 *Under Assumptions 1-4, there exists a local minimizer* $\hat{\boldsymbol{\theta}}\left(\psi_n\right)$ *for (19) that satisfies* $\Pr\left(\hat{\boldsymbol{\theta}}\left(\psi_n\right) = \check{\boldsymbol{\theta}}\right) \to 1$ *asymptotically, provided that the tuning parameter* ψ_n *meets the following conditions: (i)* $\psi_n \to 0$*; (ii)* $\psi_n/a_n \to \infty$*; and (iii)* $\psi_n \leq \min_{l\in\mathcal{J}}\left\{\theta_{l,0}\right\}/\left(q_0 c\right)$ *for some* $c > 1$ *and* $q_0 < 1/2$*.*

Theorem B.2 implies that the penalty estimator $\hat{\boldsymbol{\theta}}\left(\psi_n\right)$ performs as if we knew $\mathcal{J}$, i.e., which parameters in $\boldsymbol{\theta}_0$ do not equal zero. This property is known as the estimator being *oracle efficient*. The next theorem shows that the post-MCP-penalty estimator $\tilde{\boldsymbol{\theta}}$ converges to the true parameters, $\boldsymbol{\theta}_0$, at the same convergence rate as the penalty estimator. This implies that the post-MCP estimator is as good as the MCP-penalty estimator.

Theorem B.3 *Under the assumptions in Theorems B.1 and B.2, we obtain*

$$\left\|\tilde{\boldsymbol{\theta}} - \boldsymbol{\theta}_0\right\| = O_p\left(k_{nT}^{-2\zeta} + J/n\right),$$

and that Theorem B.1 (ii) and (iii) hold for the post-MCP estimator, $\tilde{f}\left(z\right)$*,* $\tilde{g}\left(s\right)$ *and* $\tilde{r}\left(v\right)$*.*

Appendix C. Hypothesis tests

Based on the limiting results, we introduce asymptotically valid inference procedures for two types of hypotheses broadly. First, we conduct standard Wald-type tests for the significance of the *levels* of the estimates. That is, we test for whether each of the time effects, $\{\tilde{\lambda}_t\}_{t=1}^{T-1}$, and the Tobin's q coefficient, α_0, are zero, and whether the three unknown functions $f_0(z) = 0$, $g_0(s) = 0$, and $r_0(v) = 0$, respectively. The tests for $f_0(z) = 0$ and $g_0(s) = 0$ shed light on whether cash flow and individual stock volatility contribute to investment in addition to q. The test for $r_0(v) = 0$ instead indicates whether the variable $s_{i,t}$ is correlated with $u_{i,t}$, as $r_0(v_{i,t}) = \mathbb{E}(u_{i,t}|v_{i,t})$; put differently, it tests for whether the individual stock volatility is a contemporaneous endogenous variable.

Second, we construct the following group of pointwise tests for whether the *first-order derivatives* of the three unknown curves, $f_0'(z)$, $g_0'(s)$, and, $r_0'(v)$ are constant. We do so to test for the existence of nonlinearities.

$$H_0^f: \ f_0(z) = \beta_{f,0} z, \text{ and } H_1^f: \ \text{not } H_0^f; \tag{C.1}$$

$$H_0^g: \ g_0(s) = \beta_{g,0} s, \text{ and } H_1^g: \ \text{not } H_0^g; \tag{C.2}$$

$$H_0^r: \ r_0(v) = \beta_{v,0} v, \text{ and } H_1^r: \ \text{not } H_0^r. \tag{C.3}$$

We impose an additional assumption that is stronger than Assumption 2 to asymptotically remove the series approximation bias of the first-order derivatives.

Assumption 5. There exists $\boldsymbol{\vartheta}_0 = \left[\boldsymbol{\vartheta}_{z,0}', \boldsymbol{\vartheta}_{s,0}', \boldsymbol{\vartheta}_{v,0}'\right]'$ such that

$$\max_{0\le l\le l_0} \sup_{z\in\mathcal{S}_z} \left| f_0^{(l)}(z) - \boldsymbol{\vartheta}_{z,0}' d^l \mathbf{P}^{k_{nT}}(z) \right| \le M_1 k_{nT}^{-\zeta}$$

$$\max_{0\le l\le l_0} \sup_{s\in\mathcal{S}_s} \left| g_0^{(l)}(z) - \boldsymbol{\vartheta}_{s,0}' d^l \mathbf{P}^{k_{nT}}(s) \right| \le M_2 k_{nT}^{-\zeta}$$

$$\max_{0\le l\le l_0} \sup_{v\in\mathcal{S}_v} \left| r_0^{(l)}(v) - \boldsymbol{\vartheta}_{v,0}' d^l \mathbf{P}^{k_{nT}}(v) \right| \le M_3 k_{nT}^{-\zeta}$$

for some $\zeta > 2$ and a nonnegative integer $l_0 \le 1$ as $k_{nT} \to \infty$, where $f_0^{(l)}(\cdot)$ and $d^l\mathbf{P}^{k_{nT}}(\cdot)$ denote the l-th order partial derivatives of $f_0(\cdot)$ and $\mathbf{P}^{k_{nT}}(\cdot)$, respectively.

We use the test of (C.1) as an example to illustrate the basic ideas. Under H_0^f, model (14) in the paper becomes

$$y_{i,t} \approx \mu_i + \lambda_t + \alpha_0 x_{i,t-1} + \beta_{f,0} z_{i,t-1} + \boldsymbol{\vartheta}_{s,0}' \mathbf{P}^k(s_{i,t-1}) + \boldsymbol{\vartheta}_{v,0}' \mathbf{P}^k(v_{i,t}) + \varepsilon_{i,t}, \tag{C.4}$$

in which we replace the nonparametric Hermite expansion $\boldsymbol{\vartheta}_{z,0}'\mathbf{P}^k(z_{i,t-1})$ with the linear specification $\beta_{f,0} z_{i,t-1}$. The equivalent hypothesis is

$$H_0: \ f_0'(z) = \beta_{f,0} \text{ for all } z, \text{ and } H_1: \ \text{not } H_0.$$

The post-penalty estimator for $f_0'(z)$ is defined as

$$\tilde{f}'(z) = d\vec{\mathbf{P}}^{k_{nT}}(z)' \tilde{\boldsymbol{\vartheta}}_{z,1},$$

where $d\vec{\mathbf{P}}^{k_{nT}}(z) = \frac{\partial}{\partial z}\vec{\mathbf{P}}^{k_{nT}}(z)$. We introduce some other notations: $\widehat{\mathcal{J}} = supp\left(\hat{\boldsymbol{\theta}}\right)$, and $\hat{J}$ is the dimension of $\widehat{\mathcal{J}}$; split $\widetilde{\mathbf{q}}_i = \left[\widetilde{\mathbf{q}}_{\widehat{\mathcal{J}},1,i}, \widetilde{\mathbf{q}}_{\widehat{\mathcal{J}},2,i}\right]$ and $\tilde{\boldsymbol{\theta}} = \left[\tilde{\boldsymbol{\theta}}_1', 0_{p-\hat{J}}'\right]'$, where the parameters in front of $\widetilde{\mathbf{q}}_{\widehat{\mathcal{J}},2,i}$ are all zeros and $\widetilde{\mathbf{q}}_i\,\tilde{\boldsymbol{\theta}} = \widetilde{\mathbf{q}}_{\widehat{\mathcal{J}},1,i}\tilde{\boldsymbol{\theta}}_1$; $\hat{\sigma}_n^2(z) = \left(d\vec{\mathbf{P}}^{k_{nT}}(z)\right)'\mathbf{S}_z\tilde{\boldsymbol{\Sigma}}_n\mathbf{S}_z' d\vec{\mathbf{P}}^{k_{nT}}(z)$; $\tilde{\boldsymbol{\Sigma}}_n$ equals $\hat{\boldsymbol{\Sigma}}_n$ with $\widetilde{\mathbf{q}}_{1,i}$ replaced with $\widetilde{\mathbf{q}}_{\widehat{\mathcal{J}},1,i}$;[4] $\mathbf{S}_z$ is part of $I_{\hat{J}}$ such that $\tilde{\boldsymbol{\vartheta}}_{z,1} = \mathbf{S}_z\tilde{\boldsymbol{\theta}}_1$; and $\widehat{\tilde{\boldsymbol{\varepsilon}}}_i$ are the residuals calculated under the alternative hypothesis H_1.

Under the null hypothesis H_0, when Assumptions 1-5 hold, following the proof of Theorem B.1, we obtain

$$\sqrt{n/\hat{\sigma}_n^2(z)}\left(\tilde{f}'(z) - \beta_{f,0}\right) \xrightarrow{d} N(0,1).$$

Calculating $\hat{\beta}_f$ from model (C.4), we have

$$\sqrt{n/\hat{\sigma}_n^2(z)}\left(\tilde{f}'(z) - \hat{\beta}_f\right) = \sqrt{n/\hat{\sigma}_n^2(z)}\left(\tilde{f}'(z) - \beta_{f,0}\right) + O_p\left(\hat{\sigma}_n^{-1}(z)\right),$$

because $\hat{\beta}_f - \beta_{f,0} = O_p\left(n^{-1/2}\right)$, and $\hat{\sigma}_n^2(z) \geq \lambda_{\max}\left(\mathbf{S}_z\tilde{\boldsymbol{\Sigma}}_n\mathbf{S}_z'\right)\left\|d\vec{\mathbf{P}}^{k_{nT}}(z)\right\|^2 \geq M\left\|d\vec{\mathbf{P}}^{k_{nT}}(z)\right\|^2$ under Assumption 3. This is proved by Lemma F.3 in the Online Appendix. Hence, we have

$$\sqrt{n/\hat{\sigma}_n^2(z)}\left(\tilde{f}'(z) - \hat{\beta}_f\right) \xrightarrow{d} N(0,1).$$

On the other hand, under the alternative hypothesis H_1,

$$\sqrt{n/\hat{\sigma}_n^2(z)}\left(\tilde{f}'(z) - \hat{\beta}_f\right) = \sqrt{n/\hat{\sigma}_n^2(z)}\left(f_0'(z) - \beta_f\right) + o_p(1) \xrightarrow{p} \infty,$$

provided that there exists a constant β_f such that $\hat{\beta}_f - \beta_f = O_p\left(n^{-1/2}\right)$. Hence, we can construct a t statistic for any given z point to test H_0 and H_1.

Our test statistic can also be extended to simultaneously consider a range of z values. Let z_i^*, where $i = 1, \ldots, m$, be m distinct points. Then, under H_0, we have

$$\sqrt{n\hat{\Upsilon}_n^{-1}}\begin{pmatrix} \tilde{f}'(z_1^*) - \hat{\beta}_f \\ \vdots \\ \tilde{f}'(z_m^*) - \hat{\beta}_f \end{pmatrix} \xrightarrow{d} N(0, I_m), \tag{C.5}$$

where

$$\hat{\Upsilon}_n = \begin{pmatrix} d\vec{\mathbf{P}}^{k_{nT}}(z_1^*)' \\ \vdots \\ d\vec{\mathbf{P}}^{k_{nT}}(z_m^*)' \end{pmatrix}\mathbf{S}_z\tilde{\boldsymbol{\Sigma}}_n\mathbf{S}_z'\left[d\vec{\mathbf{P}}^{k_{nT}}(z_1^*)\ldots d\vec{\mathbf{P}}^{k_{nT}}(z_m^*)\right]$$

[4] See Online Appendix (F.12).

is an $m \times m$ non-singular matrix. We can construct a χ^2 statistic as follows: under H_0,

$$T_{f,n} = n \left\| \hat{\Upsilon}_n^{-1/2} \left[\tilde{f}'\left(z_1^*\right) - \hat{\beta}_f, \ldots, \tilde{f}'\left(z_m^*\right) - \hat{\beta}_f \right]' \right\| \xrightarrow{d} \chi^2\left(m\right) \tag{C.6}$$

for a finite $m < \hat{J}$. The latter condition ensures the non-singularity of $\hat{\Upsilon}_n$.

Appendix D. Instruments for volatility

This section describes how to construct the 10 instrumental variables $\mathbf{w}_{i,t}$ to address the endogeneity in individual stock volatility. The work here largely follows Alfaro, Bloom and Lin (2018) and consists of several steps. First, we collect each individual firm's daily stock returns from CRSP. For each firm-day combination in our sample, we regress the firm's stock returns in excess of the risk-free rate on the four asset-pricing factors in Carhart (1997) using a rolling window of the previous 2,520 trading days, if data are available.[5] The regression residual is referred to as $r_{i,t}^{risk\text{-}adj}$, where i denotes the firm and t represents the day.

Second, for each year t_o in 2010-2017, we run the following pooled regression for all the firms in a given industry-j using all the days in the recursive rolling window that starts from the year 2010 until the year t_o.

$$r_{i,t}^{risk\text{-}adj} = \alpha_{j,t_o} + \beta_{j,t_o}^{(1)} \times r_t^{(1)} + \beta_{j,t_o}^{(2)} \times r_t^{(2)} ... + \beta_{j,t_o}^{(10)} \times r_t^{(10)} + \epsilon_{i,t}, \tag{D.1}$$

where α_{j,t_o} and $\beta_{j,t_o}^{(c)}$ are the industry-and-year specific intercept and coefficients; $\epsilon_{i,t}$ is the error term; and r_t^c for $c \in \{1, ..., 10\}$ correspond to the 10 different sources of aggregate shocks in Alfaro, Bloom and Lin (2018): (i) When c represents oil prices, r_t^c is the daily growth rate of the price of the crude oil futures contract CL1 COMB Comdty on Bloomberg; (ii) when c represents U.S. 10-year Treasury, r_t^c is the daily first difference of the 10-year Treasury yield multiplied by negative 1;[6] (iii) when c represents U.S. policy uncertainty, r_t^c is the daily growth rate of the policy uncertainty index in Baker, Bloom and Davis (2016); and (iv)-(x) when c represents the exchange rates between the U.S. dollars and seven major currencies around the world, r_t^c is the daily growth rate of the corresponding exchange rate. The seven major currencies are the Australian dollar, British pound, Canadian dollar, Euro, Japanese yen, Swedish krona, and Swiss franc. The daily growth rate of a variable a is calculated as $[a_t - a_{t-1}]/[(a_t + a_{t-1})/2]$, where t is a day. For each estimated coefficient β_{j,t_o}^c, we obtain its t-statistic t_{j,t_o}^c and replace the t-statistic with 0 if the absolute value of the t-statistic is smaller than 1, indicating that the coefficient is insignificant.[7] We generate a weighted

[5]At least 252 trading days of non-missing data are required in this estimation.

[6]This calculation uses the first-order approximation of duration.

[7]The usual criterion of statistical significance at the 10 percent level is 1.645. We instead use 1 because a

coefficient $\beta^{c,weighted}_{j,t_o} = (|t^c_{j,t_o}|/\sum_c |t^c_{j,t_o}|) \cdot \beta^c_{j,t_o}$.

Third, the instruments for $\Delta s_{i,t_o}$ are constructed as $w^c_{i,t_o} = |\beta^{c,weighted}_{j,t_o}| \cdot \Delta\sigma^c_{t_o}$, where firm-$i$ belongs to industry-j, and $\sigma^c_{t_o}$ represents aggregate uncertainty measures calculated as follows. (i) When c represents oil prices, $\sigma^c_{t_o}$ is the annual average of daily (30-day) volatility of the crude oil futures contract CL1 COMB Comdty (Bloomberg); (ii) when c represents the U.S. 10-year Treasury rate, $\sigma^c_{t_o}$ is the annual average of daily TYVIX; (iii) when c represents U.S. policy uncertainty, $\sigma^c_{t_o}$ is the annual average of daily policy uncertainty index in Baker, Bloom and Davis (2016); and (iv)-(x) when c represents the exchange rates between U.S. dollars and seven abovementioned currencies, $\sigma^c_{t_o}$ is the annual average of daily (three-month) volatility of each exchange rate (Bloomberg CMPN). The first-stage regression results are reported in Table D.1.[8]

Note that the weighting schemes directly follow Alfaro, Bloom and Lin (2018) and aim to reduce noise in the estimation of the industry exposure to different macro uncertainty shocks. Without the weighting schemes, the coefficients in regression (D.1) would represent the industry exposure in a way that a larger coefficient indicates higher exposure. However, a large coefficient can be statistically insignificant if its standard error is proportionally large. The weighting schemes instead weight each coefficient by its t-statistics value, where a less significant coefficient (indicated by a smaller t value) receives a smaller weight, and therefore, can result in a better measure of the industry exposure. That said, for our empirical sample, dropping the weighting schemes leads to similar estimation results, shown in Appendix E.

We further follow Alfaro, Bloom and Lin (2018) to construct 10 first-moment control variables to make sure that our results are not driven by movements in oil prices, Treasury yields, U.S. government policies, and exchange rates themselves but rather by the movements in their volatility (see section section IV.C.3 in the main text). The 10 control variables are constructed by $\beta^{c,weighted}_{j,t_o} \cdot r^c_{t_o}$, where $r^c_{t_o}$ is the first-moment aggregate shock in source c in year t_o. When c represents oil prices, U.S. 10-year Treasury, or the seven exchange rates, $r^c_{t_o}$ is the annual average of the corresponding daily r^c_t in equation (D.1). When c represents U.S. policy uncertainty, $r^c_{t_o}$ is the annual growth of the U.S. government expenditure as a share of GDP. In addition, in a robustness check (see section IV.C.3), we use out-of-sample data to construct $\mathbf{w}_{i,t}$ to ensure that these instruments are exogenous. That is, we fit regression (D.1) using the individual stock return data between 1998 and 2009 to obtain $\beta^{c,weighted}_{j,t_o}$. However, we still use the aggregate data between 2009 and 2017 to calculate $\Delta\sigma^c_{t_o}$ because

t-value larger than 1 (or smaller than -1) means that this coefficient is significant in economic terms in the sense that dropping the variable from the regression model can considerably influence the estimation results.

[8] We further add the cash-flow-to-capital ratio and time dummies to the first-stage regression, because they are exogenous variables. The first-stage regression is estimated using pooled OLS; the constant term is included in the estimation but omitted from the table.

aggregate uncertainty shocks are assumed to be exogenous to individual firms' investment plans.

Table D.1: First-stage pooled regression

	(1)
	Volatility changes
W oil	2.644***
	(0.206)
W 10-year Treasury	28.92
	(37.17)
W policy uncertainty	-26.48
	(29.51)
W AUD	2.299
	(2.906)
W CAD	2.000
	(2.069)
W CHF	6.438***
	(1.281)
W EUR	3.177
	(2.233)
W GBP	4.954*
	(2.640)
W JPY	4.339*
	(2.553)
W SEK	7.791***
	(2.155)
CF/K (z)	0.00427
	(0.00496)
Year dummy 2010	-0.205***
	(0.00663)
Year dummy 2011	0.0743***
	(0.00631)
Year dummy 2012	-0.0484***
	(0.00634)
Year dummy 2013	-0.00855
	(0.00672)

Continued on the next page

Table D.1 – continued from the previous page

	(1)
	Volatility changes
Year dummy 2014	0.0516***
	(0.00622)
Year dummy 2015	0.0437***
	(0.00742)
Year dummy 2016	0.0629***
	(0.00672)
Number of firms	1,025
Number of obs	8,200
R-squared	0.353
F-test	248.2

*** p-val < 0.01; ** p-val < 0.05; * p-val < 0.1.

Appendix E. Robustness results

Figure E.1 shows the estimates of the first-order derivative, $d\hat{g}(s)/ds$, of the investment-uncertainty relation for the following robustness checks: panel (a) for $k = 6$ (footnote 20 in Li and Sun (2022)); (b) for less strict sample selection, and (c) for stricter sample selection (footnote 25 in Li and Sun (2022)); (d)-(l) for robustness outlined in section IV.C.3 in Li and Sun (2022), (1)-(8).[9] The results are similar to Figure 1 in Li and Sun (2022).

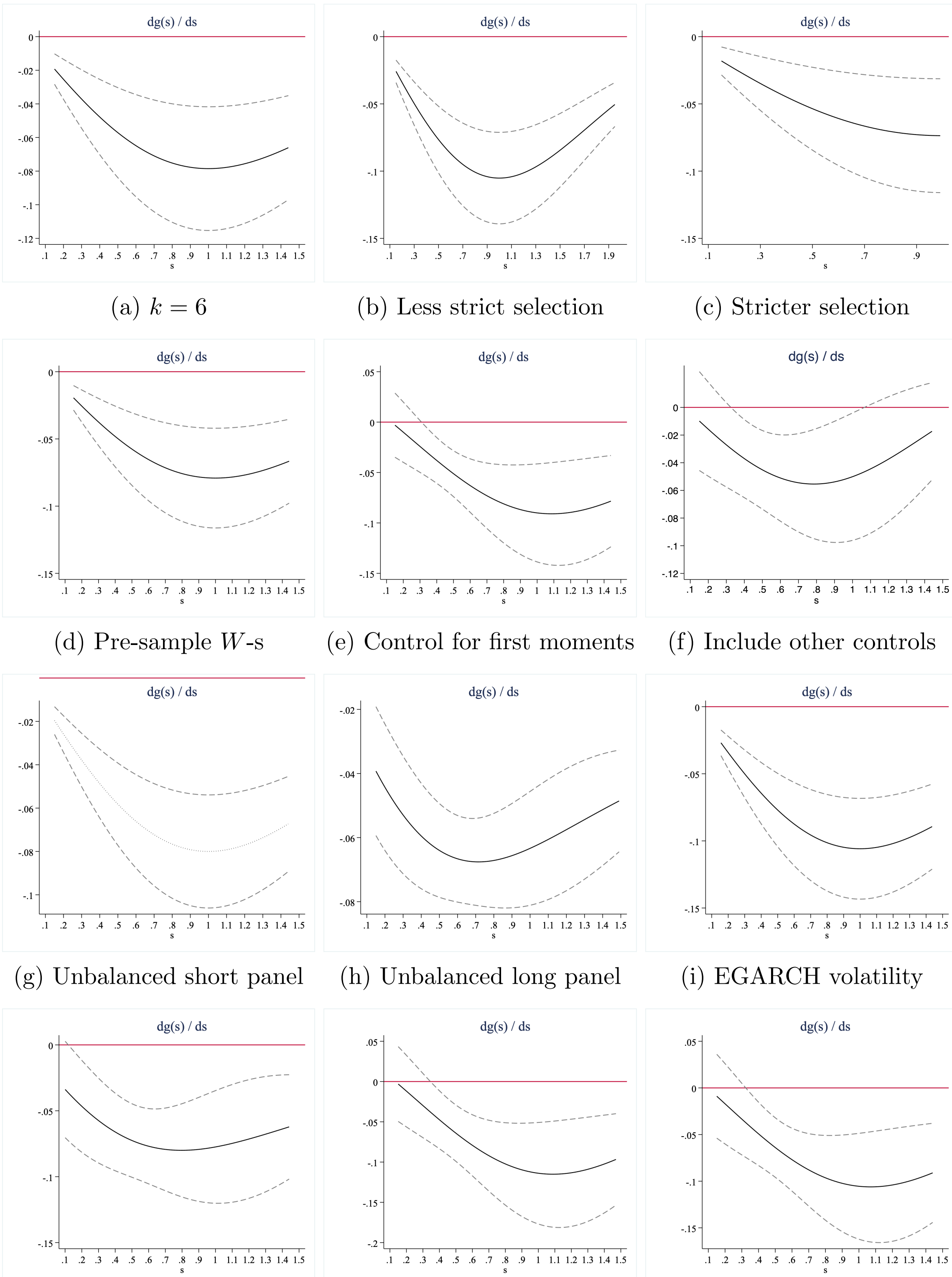
Figure E.2 shows results of more robustness checks: panel (a) for using linear specifications for f_0 and r_0;[10] (b) for dropping all the instruments related to time dummies when treating mismeasured q; (c) for dropping all the instruments related to cash flow when treating mismeasured q; (d) for dropping all the instruments related to volatility when treating mismeasured q; (e) for dropping all the instruments related to the control function when treating mismeasured q;[11] (f) for using granular instrumental variables (GIV) as in Gabaix

[9] Note that for panel (f) where control variables from Kim and Kung (2017) and Panousi and Papanikolaou (2012) are included, we follow their work to treat those variables as exogenous variables in the estimation.

[10] See discussions in section II in Li and Sun (2022).

[11] For panels (b)-(e), see discussions in footnote 18 in Li and Sun (2022).

Figure E.1: Investment-Uncertainty Relation: Robustness



(a) $k = 6$

(b) Less strict selection

(c) Stricter selection

(d) Pre-sample W-s

(e) Control for first moments

(f) Include other controls

(g) Unbalanced short panel

(h) Unbalanced long panel

(i) EGARCH volatility

(j) Idiosyncratic volatility

(k) Alternative cash flow

(l) Include R&D

and Koijen (2020);[12] (g) for using W-variables without the weighting scheme.[13] The results are similar to Figure 1 in Li and Sun (2022).

Figure E.3 shows the comparison of investment-uncertainty relations between low- and high-irreversibility subsamples, where irreversibility is measured by a firm-level scale inflexibility index calculated from quarterly data in our sample period, following Gu, Hackbarth and Johnson (2018). This comparison is similar to Figure 5 in Li and Sun (2022).

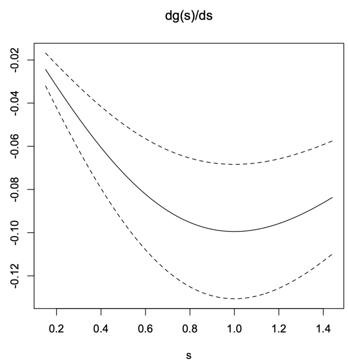
Table E.1 corresponds to the discussion in footnote 29 in Li and Sun (2022) in which we compare three different estimates of the linear model (1). Column (1) includes the results from our baseline estimator but without series approximation. Column (2) shows the results from a modified Erickson, Jiang and Whited (2014) (EJW) estimator, in which we combine their measurement error remedy with the control function approach in our method to simultaneously account for mismeasurement in q and the regressor endogeneity of individual volatility. Column (3) represents the estimates obtained from the original EJW method and their empirical model (note that their model does not include individual volatility as a regressor). The EJW estimates utilize up-to-4th order cumulants following the suggestion in Erickson, Parham and Whited (2017). We observe that the estimates are close to each other in columns (1) and (2), implying that the measurement error remedy in our estimator and that in EJW perform equally well after the regressor endogeneity of uncertainty is treated. Moreover, the coefficient of cash flow is significant in all the three columns. Thus, cash flow is a crucial factor influencing investment. This result is different from those in Erickson and Whited (2000) and Erickson, Jiang and Whited (2014). The comparison here shows that the difference is likely due to different empirical samples rather than estimation methods.

Table E.2 reports the percentage of firms in an extended sample for which key variables may contain unit roots. The extended sample includes 3,730 individual firms that have at least 3 years of non-missing data between 1986-2017. For each firm, and for each key variable, we conduct the Augmented Dickey-Fuller (ADF) test and the Phillips-Perron (PP) test for unit roots. The null hypothesis of both tests is that the variable contains a unit root, and the alternative is that the variable is stationary. For each variable, we report the percentage of firms for which we cannot reject the null of unit roots under the usual 5 percent statistical level. For example, the number for "y: I/K" under the ADF test is 68.31, indicating that for 68.31 percent of the 3,730 firms, the variable of investment-to-capital ratio is likely to contain

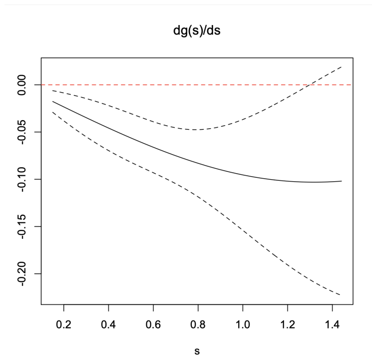
[12] We follow the idea of equations (5) and (6) in Gabaix and Koijen (2020) to construct a GIV as the difference between size-weighted average stock volatility and equal-weighted average stock volatility for each business sector (classified by two-digit SIC codes). We choose to let our instruments vary with both sectors and time, rather than just time as in Gabaix and Koijen (2020), to increase estimation efficiency. Moreover, given that both our main regression model and the first-stage regression model contain time dummies as regressors, using GIVs which vary only with time as instruments is not feasible due to multicollinearity.

[13] See discussions in Appendix D.

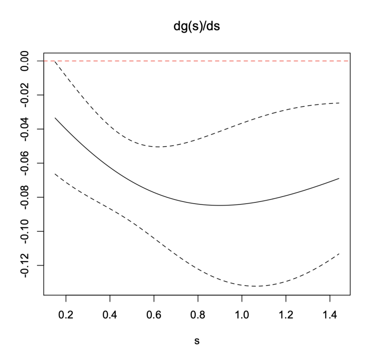
Figure E.2: Investment-Uncertainty Relation: More Robustness



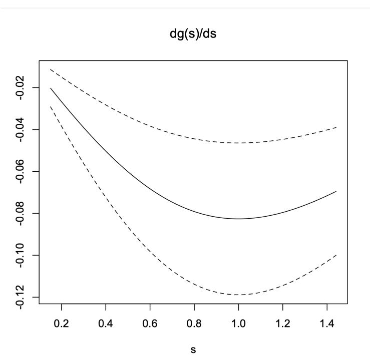
(a) Linear f_0 and r_0



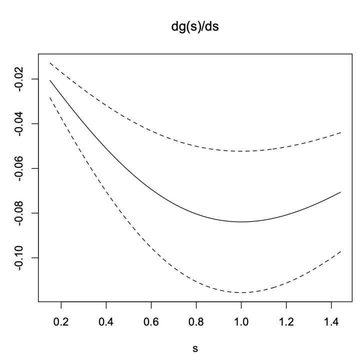
(b) Drop T instruments



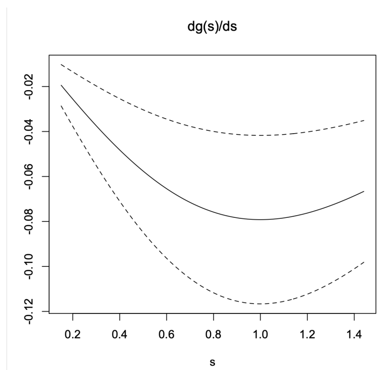
(c) Drop z instruments



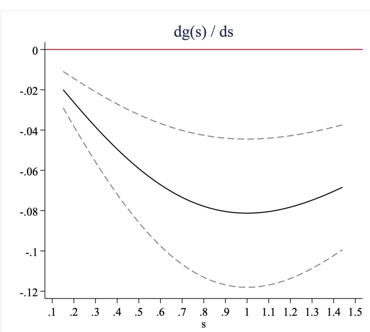
(d) Drop s instruments



(e) Drop v instruments

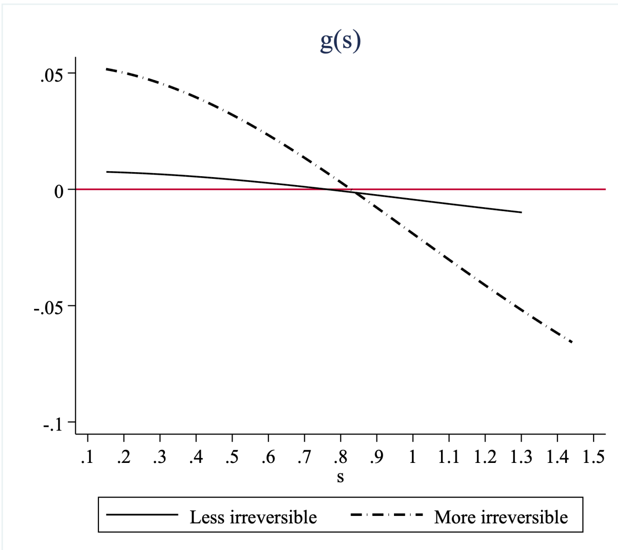


(f) GIV

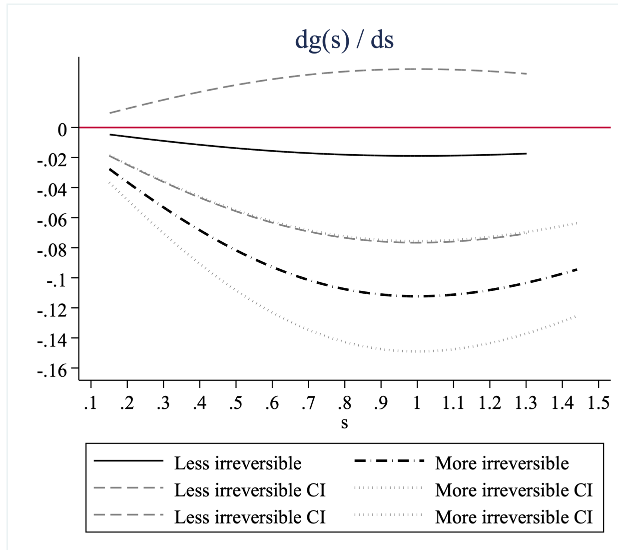


(g) No weights in W-s

Figure E.3: Low Versus High Irreversibility (Measured by Inflexibility)



(a) $\hat{g}(s)$



(b) $d\hat{g}(s)/ds$

Table E.1: Linear estimates

	(1) Our estimator	(2) Modified EJW	(3) EJW
Tobin's q	0.01***	0.01*	0.01*
	(0.002)	(0.004)	(0.003)
Cash flow	0.03***	0.05***	0.05***
	(0.008)	(0.014)	(0.013)
Individual volatility	-0.08***	-0.08***	
	(0.012)	(0.016)	
Control function	-0.03***	-0.03***	
	(0.009)	(0.011)	

Standard errors in parentheses: *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$

Table E.2: Unit-root tests

	y: I/K	x: Tobin's q	z: cash flow	s: volatility
ADF	68.31	77.51	74.08	76.6
PP	66.06	74.64	72.04	72.49

a unit root according to the ADF test. The test results reported here are obtained without a time trend and under the optimal number of lags for each firm, equal to $int\{4(T/100)^{2/9}\}$, where int takes the integer value, and T refers to each firm's time-series length (see official documents of Stata 16). The results are also robust to a uniform setting of lags between 1-3 and both with and without a time trend.[14]

Appendix F: Mathematical Proofs

In this section, $\mathbb{E}_n(\cdot)$ denotes the average $n^{-1}\sum_{i=1}^{n}$ over index i and $\mathbb{E}(\cdot)$ denotes the population expectation. Additionally, (i) $\tilde{p}_l(v_{i,t}) = p_l(v_{i,t}) - T^{-1}\sum_{s=1}^{T} p_l(v_{i,s})$ and $\tilde{p}'_l(v_{i,t}) = p'_l(v_{i,t}) - T^{-1}\sum_{s=1}^{T} p'_l(v_{i,s})$; (ii) $\left\|P^{k_{nT}}\right\|_l = \max_{0\leq j\leq l}\sup_{x\in R}\left\|d^j P^{k_{nT}}(x)/dx^j\right\|$ for $l = 0, 1, 2$; (iii) $\|A\| = [\mathrm{tr}(AA')]^{1/2}$ and $\|A\|_{sp} = \lambda_{\max}^{1/2}(AA')$ denote the Euclidean and spectral norm of matrix A, respectively. In addition, to fit our mathematical equations with page margin, we introduce the symbols ♠, ♣, ♦, ★, ▼, and ▽; they may have different meanings in different places.

Lemma F.1 *Under Assumptions 1-4 and denoting*

$$\delta_{nT} = (nT)^{-1/2}\left\|P^{k_{nT}}\right\|_1,$$

we have

$$\left\|\mathbb{E}_n\left(\widetilde{\tilde{\mathbf{q}}}_i'\widetilde{\tilde{\mathbf{d}}}_i - \tilde{\mathbf{q}}_i'\tilde{\mathbf{d}}_i\right)\right\| = O_p\left(\delta_{nT}k_{nT}^{1/2}\right) \tag{F.1}$$

$$\left\|\mathbb{E}_n\left[\left(\widetilde{\tilde{\mathbf{d}}}_i - \tilde{\mathbf{d}}_i\right)'\tilde{\boldsymbol{\Delta}}_i\right]\right\| = O_p\left(\delta_{nT}k_{nT}^{-\zeta}\right) \tag{F.2}$$

$$\left\|\mathbb{E}_n\left[\left(\widetilde{\tilde{\mathbf{d}}}_i - \tilde{\mathbf{d}}_i\right)'\tilde{\boldsymbol{\varepsilon}}_i\right]\right\| = O_p\left(\delta_{nT}/\sqrt{n}\right) \tag{F.3}$$

where $\tilde{\boldsymbol{\Delta}}_i = J_T\boldsymbol{\Delta}_i$, $\boldsymbol{\Delta}_i = [\Delta_{i,1},\ldots,\Delta_{i,T}]'$ *and* $\Delta_{i,t} = f_0(z_{i,t-1}) + g_0(s_{i,t-1}) + r_0(v_{i,t}) - E'_{i,t}\boldsymbol{\vartheta}_0$.

[14] A uniform setting of lags means that we set the same number of lags to all firms, regardless of their time-series length. We consider the maximum number of lags to be 3 because the maximum time-series length is 32 years, and $int\{4(T/100)^{2/9}\} = 3$.

Proof of Lemma F.1: The matrix form of model (6) in the main text is given by

$$\Delta \mathbf{s} = \mathbf{i}_{nT}\eta_0 + \mathbf{W}\pi_0 + \mathbf{v}$$

where $\mathbf{s} = [\mathbf{s}_1', \ldots, \mathbf{s}_n']'$, $\mathbf{v} = [\mathbf{v}_1', \ldots, \mathbf{v}_n']'$, and $\mathbf{W} = [\mathbf{W}_1, \ldots, \mathbf{W}_n]'$ is an $(nT) \times d_w$ matrix. Applying the least squares estimation gives $\|\hat{\pi} - \pi_0\| = O_p\left((nT)^{-1/2}\right)$ and $\hat{\eta} - \eta_0 = O_p\left((nT)^{-1/2}\right)$ under Assumptions 1(i) and 3(i).

First, we verify (F.1). It is readily seen that

$$\begin{aligned}
&\mathbb{E}_n\left(\widetilde{\hat{\mathbf{q}}}_i'\widetilde{\hat{\mathbf{d}}}_i - \tilde{\mathbf{q}}_i'\tilde{\mathbf{d}}_i\right) \qquad \text{(F.4)}\\
&= \mathbb{E}_n\left[\left(\widetilde{\hat{\mathbf{q}}}_i - \tilde{\mathbf{q}}_i\right)'\left(\widetilde{\hat{\mathbf{d}}}_i - \tilde{\mathbf{d}}_i\right)\right] + \mathbb{E}_n\left[\left(\widetilde{\hat{\mathbf{q}}}_i - \tilde{\mathbf{q}}_i\right)'\tilde{\mathbf{d}}_i\right] + \mathbb{E}_n\left[\tilde{\mathbf{q}}_i'\left(\widetilde{\hat{\mathbf{d}}}_i - \tilde{\mathbf{d}}_i\right)\right]\\
&= \mathbb{E}_n\left\{\left(P_{\hat{v},i}^{k_{nT}} - P_{v,i}^{k_{nT}}\right)' J_T\left[I_T \otimes \operatorname{vec}\left(J_T\left(P_{\hat{v},i}^{k_{nT}} - P_{v,i}^{k_{nT}}\right)\right)'\right]\right\}\\
&+ \mathbb{E}_n\left[\left(P_{\hat{v},i}^{k_{nT}} - P_{v,i}^{k_{nT}}\right)' J_T\tilde{\mathbf{d}}_i\right] + \mathbb{E}_n\left\{\tilde{\mathbf{q}}_i'\left[I_T \otimes \operatorname{vec}\left(J_T\left(P_{\hat{v},i}^{k_{nT}} - P_{v,i}^{k_{nT}}\right)\right)'\right]\right\},
\end{aligned}$$

where for the first term, we have

$$\begin{aligned}
&\left\|\mathbb{E}_n\left\{\left(P_{\hat{v},i}^{k_{nT}} - P_{v,i}^{k_{nT}}\right)' J_T\left[I_T \otimes \operatorname{vec}\left(J_T\left(P_{\hat{v},i}^{k_{nT}} - P_{v,i}^{k_{nT}}\right)\right)'\right]\right\}\right\|\\
&\leq \mathbb{E}_n\left(\left\|\left(P_{\hat{v},i}^{k_{nT}} - P_{v,i}^{k_{nT}}\right)' J_T\left[I_T \otimes \operatorname{vec}\left(J_T\left(P_{\hat{v},i}^{k_{nT}} - P_{v,i}^{k_{nT}}\right)\right)'\right]\right\|\right)\\
&= \mathbb{E}_n\left\{\operatorname{tr}\left[\left(P_{\hat{v},i}^{k_{nT}} - P_{v,i}^{k_{nT}}\right)' J_T\left(P_{\hat{v},i}^{k_{nT}} - P_{v,i}^{k_{nT}}\right)\right]\right\}\\
&= \mathbb{E}_n\left(\sum_{l=1}^{k_{nT}}\sum_{t=1}^{T}\left[\tilde{p}_l\left(\hat{v}_{i,t}\right) - \tilde{p}_l\left(v_{i,t}\right)\right]^2\right) = O_p\left(\delta_{nT}^2\right) \qquad \text{(F.5)}
\end{aligned}$$

under Assumptions 1(i) and 2(ii) because $\hat{v}_{i,t} - v_{i,t} = \eta_0 - \hat{\eta} + \mathbf{w}_{i,t}'\left(\boldsymbol{\pi}_0 - \hat{\boldsymbol{\pi}}\right)$, and applying

Taylor expansion yields

$$\tilde{p}_l(\hat{v}_{i,t}) - \tilde{p}_l(v_{i,t}) = p_l'(\mathring{v}_{i,t})(\hat{v}_{i,t} - v_{i,t}) - T^{-1}\sum_{s=1}^{T} p_l'(\mathring{v}_{i,s})(\hat{v}_{i,s} - v_{i,s})$$
$$= p_l'(\mathring{v}_{i,t})\left[\eta_0 - \hat{\eta} + \mathbf{w}_{i,t}'(\boldsymbol{\pi}_0 - \hat{\boldsymbol{\pi}})\right]$$
$$- T^{-1}\sum_{s=1}^{T} p_l'(\mathring{v}_{i,s})\left(\eta_0 - \hat{\eta} + \mathbf{w}_{i,s}'(\boldsymbol{\pi}_0 - \hat{\boldsymbol{\pi}})\right)$$
$$= \tilde{p}_l'(\mathring{v}_{i,t})(\eta_0 - \hat{\eta}) - (\boldsymbol{\pi}_0 - \hat{\boldsymbol{\pi}})'\left[\mathbf{w}_{i,t}p_l'(\mathring{v}_{i,t}) - T^{-1}\sum_{s=1}^{T}\mathbf{w}_{i,s}p_l'(\mathring{v}_{i,s})\right]$$

where $\mathring{v}_{i,t}$ lies between $\hat{v}_{i,t}$ and $v_{i,t}$. Additionally, we have

$$\left\|\mathbb{E}_n\left[\left(P_{\hat{v},i}^{k_{nT}} - P_{v,i}^{k_{nT}}\right)' J_T\tilde{\mathbf{d}}_i\right]\right\| \leq \mathbb{E}_n\left\|\left(P_{\hat{v},i}^{k_{nT}} - P_{v,i}^{k_{nT}}\right)' J_T\tilde{\mathbf{d}}_i\right\|$$
$$\leq 2\mathbb{E}_n\left|\mathrm{tr}\left[\left(P_{\hat{v},i}^{k_{nT}} - P_{v,i}^{k_{nT}}\right)' J_T\left(P_{\hat{v},i}^{k_{nT}} - P_{v,i}^{k_{nT}}\right)\right]\left[\mathrm{tr}\left(\tilde{\mathbf{E}}_i'\tilde{\mathbf{E}}_i\right) + \mathrm{tr}\left(F_T'F_T\right)\right]\right|^{1/2}$$
$$+ 2\mathbb{E}_n\left|\mathrm{tr}\left(\left(P_{\hat{v},i}^{k_{nT}} - P_{v,i}^{k_{nT}}\right)' J_T\left(I_T\otimes\tilde{\mathbf{y}}_i\right)'\spadesuit\left(I_T\otimes\tilde{\mathbf{y}}_i\right)J_T\left(P_{\hat{v},i}^{k_{nT}} - P_{v,i}^{k_{nT}}\right)\right)\right|^{1/2}$$
$$= O_p\left(\delta_{nT}k_{nT}^{1/2}\right) + O_p\left(\delta_{nT}\right)$$

under Assumptions 1(i), 2(ii) and 3(i), where we obtain the last line using Hölder's inequality, (F.5), Lemma F.2(i), tr(AB) ≤tr(A)tr(B) for non-negative definite matrices A and B and $\spadesuit = \left[\left(J_T\otimes J_T\right)D_T\right]_\perp\left[\left(J_T\otimes J_T\right)D_T\right]_\perp'$, tr$[\spadesuit] = m_0$. Similarly, we obtain

$$\left\|\mathbb{E}_n\left\{\widetilde{\mathbf{q}}_i'\left[I_T\otimes\mathrm{vec}\left(J_T\left(P_{\hat{v},i}^{k_{nT}} - P_{v,i}^{k_{nT}}\right)\right)'\right]\right\}\right\|$$
$$\leq \mathbb{E}_n\left\|\widetilde{\mathbf{q}}_i'\left[I_T\otimes\mathrm{vec}\left(J_T\left(P_{\hat{v},i}^{k_{nT}} - P_{v,i}^{k_{nT}}\right)\right)'\right]\right\|$$
$$= \mathbb{E}_n\left|\mathrm{tr}\left[\left(P_{\hat{v},i}^{k_{nT}} - P_{v,i}^{k_{nT}}\right)' J_T\left(P_{\hat{v},i}^{k_{nT}} - P_{v,i}^{k_{nT}}\right)\right]\mathrm{tr}\left(\widetilde{\mathbf{q}}_i'\widetilde{\mathbf{q}}_i\right)\right|^{1/2}$$
$$= O_p\left(\delta_{nT}k_{nT}^{1/2}\right).$$

Hence, under Assumptions 1-4, taking together the results above gives (i).

Second, applying a similar proof method used above and by (F.5), we have

$$\begin{aligned}
&\left\|\mathbb{E}_n\left[\left(\widetilde{\hat{\mathbf{d}}}_i-\tilde{\mathbf{d}}_i\right)'\tilde{\boldsymbol{\Delta}}_i\right]\right\|=\left\|\mathbb{E}_n\left\{\left[I_T\otimes\operatorname{vec}\left(J_T\left(P_{\hat{v},i}^{k_{nT}}-P_{v,i}^{k_{nT}}\right)\right)\right]\tilde{\boldsymbol{\Delta}}_i\right\}\right\|\\
&\leq\sqrt{T}\mathbb{E}_n\left(\tilde{\boldsymbol{\Delta}}_i'\tilde{\boldsymbol{\Delta}}_i\operatorname{tr}\left[\left(P_{\hat{v},i}^{k_{nT}}-P_{v,i}^{k_{nT}}\right)'J_T\left(P_{\hat{v},i}^{k_{nT}}-P_{v,i}^{k_{nT}}\right)\right]\right)^{1/2}=O_p\left(\delta_{nT}k_{nT}^{-\zeta}\right),
\end{aligned}$$

where $\max_{1\leq i\leq n}\left|\tilde{\boldsymbol{\Delta}}_i'\tilde{\boldsymbol{\Delta}}_i\right|\leq Mk_{nT}^{-2\zeta}$ under Assumption 2(i), and

$$\begin{aligned}
&\mathbb{E}\left\|\mathbb{E}_n\left[\left(\widetilde{\hat{\mathbf{d}}}_i-\tilde{\mathbf{d}}_i\right)'\tilde{\boldsymbol{\varepsilon}}_i\right]\right\|^2\\
&=\mathbb{E}\left\|\mathbb{E}_n\left\{\left[I_T\otimes\operatorname{vec}\left(J_T\left(P_{\hat{v},i}^{k_{nT}}-P_{v,i}^{k_{nT}}\right)\right)\right]\tilde{\boldsymbol{\varepsilon}}_i\right\}\right\|^2\\
&=n^{-1}\mathbb{E}\left\{\left[I_T\otimes\operatorname{vec}\left(J_T\left(P_{\hat{v},i}^{k_{nT}}-P_{v,i}^{k_{nT}}\right)\right)\right]\tilde{\boldsymbol{\varepsilon}}_i\tilde{\boldsymbol{\varepsilon}}_i'\left[I_T\otimes\operatorname{vec}\left(J_T\left(P_{\hat{v},i}^{k_{nT}}-P_{v,i}^{k_{nT}}\right)\right)\right]\right\}\\
&\leq n^{-1}\mathbb{E}\left\{\tilde{\boldsymbol{\varepsilon}}_i'\tilde{\boldsymbol{\varepsilon}}_i\operatorname{tr}\left[\left(P_{\hat{v},i}^{k_{nT}}-P_{v,i}^{k_{nT}}\right)'J_T\left(P_{\hat{v},i}^{k_{nT}}-P_{v,i}^{k_{nT}}\right)\right]\right\}=O\left(\delta_{nT}^2/n\right).
\end{aligned}$$

This completes the proof of this lemma.

Lemma F.2 *Under Assumptions 1-4, we have*

$$\begin{aligned}
\left\|\mathbb{E}_n\left(\tilde{\mathbf{d}}_i'\widetilde{\mathbf{q}}_i\right)-\mathbb{E}\left(\tilde{\mathbf{d}}_i'\widetilde{\mathbf{q}}_i\right)\right\|&=O_p\left(k_{nT}/\sqrt{n}\right)\\
\left\|\mathbb{E}_n\left(\tilde{\mathbf{d}}_i'\tilde{\boldsymbol{\Delta}}_i\right)-\mathbb{E}\left(\tilde{\mathbf{d}}_i'\tilde{\boldsymbol{\Delta}}_i\right)\right\|&=O_p\left(k_{nT}^{1/2-\zeta}/\sqrt{n}\right).
\end{aligned}$$

Proof of Lemma F.2: First, we consider

$$
\begin{aligned}
&\mathbb{E}\left\|\mathbb{E}_n\left(\tilde{\mathbf{d}}_i'\widetilde{\mathbf{q}}_i\right)-\mathbb{E}\left(\tilde{\mathbf{d}}_i'\widetilde{\mathbf{q}}_i\right)\right\|^2\\
&=n^{-1}\mathbb{E}\left\{\operatorname{tr}\left[\left(\tilde{\mathbf{d}}_i'\widetilde{\mathbf{q}}_i-\mathbb{E}\left(\tilde{\mathbf{d}}_i'\widetilde{\mathbf{q}}_i\right)\right)'\left(\tilde{\mathbf{d}}_i'\widetilde{\mathbf{q}}_i-\mathbb{E}\left(\tilde{\mathbf{d}}_i'\widetilde{\mathbf{q}}_i\right)\right)\right]\right\}\\
&\leq n^{-1}\mathbb{E}\left[\operatorname{tr}\left(\tilde{\mathbf{d}}_i'\widetilde{\mathbf{q}}_i\widetilde{\mathbf{q}}_i'\tilde{\mathbf{d}}_i\right)\right]\leq n^{-1}\mathbb{E}\left[\operatorname{tr}\left(\widetilde{\mathbf{q}}_i\widetilde{\mathbf{q}}_i'\right)\operatorname{tr}\left(\tilde{\mathbf{d}}_i\tilde{\mathbf{d}}_i'\right)\right]\\
&=n^{-1}\mathbb{E}\left[\operatorname{tr}\left(F_TF_T'+\tilde{\mathbf{x}}_{i,-1}\tilde{\mathbf{x}}_{i,-1}'+\tilde{\mathbf{E}}_i\tilde{\mathbf{E}}_i'\right)\right.\\
&\left.\times\operatorname{tr}\left(\left(I_T\otimes\tilde{\mathbf{y}}_i\right)'\spadesuit\left(I_T\otimes\tilde{\mathbf{y}}_i\right)+\operatorname{tr}\left(\tilde{\mathbf{E}}_i\tilde{\mathbf{E}}_i'+F_TF_T'\right)I_T\right)\right]\\
&=n^{-1}\mathbb{E}\left[\left(2\left(T-1\right)+\tilde{\mathbf{x}}_{i,-1}'\tilde{\mathbf{x}}_{i,-1}+\operatorname{tr}\left(\tilde{\mathbf{E}}_i\tilde{\mathbf{E}}_i'\right)\right)\right.\\
&\left.\times\operatorname{tr}\left(\left(I_T\otimes\tilde{\mathbf{y}}_i\right)'\spadesuit\left(I_T\otimes\tilde{\mathbf{y}}_i\right)+\left(\operatorname{tr}\left(\tilde{\mathbf{E}}_i\tilde{\mathbf{E}}_i'\right)+2\left(T-1\right)\right)T\right)\right]\\
&\leq n^{-1}T\mathbb{E}\left[\left(2\left(T-1\right)+\tilde{\mathbf{x}}_{i,-1}'\tilde{\mathbf{x}}_{i,-1}+\operatorname{tr}\left(\tilde{\mathbf{E}}_i\tilde{\mathbf{E}}_i'\right)\right)\left(\tilde{\mathbf{y}}_i'\tilde{\mathbf{y}}_i+\operatorname{tr}\left(\tilde{\mathbf{E}}_i\tilde{\mathbf{E}}_i'\right)+2\left(T-1\right)\right)\right]\\
&=O\left(n^{-1}k_{nT}^2\right),
\end{aligned}
$$

where $\spadesuit=\left[\left(J_T\otimes J_T\right)D_T\right]_\perp\left[\left(J_T\otimes J_T\right)D_T\right]_\perp'$. Second, we obtain

$$
\begin{aligned}
&\mathbb{E}\left\|\mathbb{E}_n\left(\tilde{\mathbf{d}}_i'\tilde{\mathbf{\Delta}}_i\right)-\mathbb{E}\left(\tilde{\mathbf{d}}_i'\tilde{\mathbf{\Delta}}_i\right)\right\|^2\\
&=n^{-1}\mathbb{E}\left\{\operatorname{tr}\left[\left(\tilde{\mathbf{d}}_i'\tilde{\mathbf{\Delta}}_i-\mathbb{E}\left(\tilde{\mathbf{d}}_i'\tilde{\mathbf{\Delta}}_i\right)\right)'\left(\tilde{\mathbf{d}}_i'\tilde{\mathbf{\Delta}}_i-\mathbb{E}\left(\tilde{\mathbf{d}}_i'\tilde{\mathbf{\Delta}}_i\right)\right)\right]\right\}\\
&\leq n^{-1}\mathbb{E}\left[\operatorname{tr}\left(\tilde{\mathbf{d}}_i'\tilde{\mathbf{\Delta}}_i\tilde{\mathbf{\Delta}}_i'\tilde{\mathbf{d}}_i\right)\right]\leq n^{-1}\mathbb{E}\left[\tilde{\mathbf{\Delta}}_i'\tilde{\mathbf{\Delta}}_i\operatorname{tr}\left(\tilde{\mathbf{d}}_i\tilde{\mathbf{d}}_i'\right)\right]=O\left(n^{-1}k_{nT}^{1-2\zeta}\right)
\end{aligned}
$$

under Assumptions 1-2. This completes the proof of this lemma.

Lemma F.3 *Under Assumptions 1-4, we have*

$$
\begin{aligned}
\left\|\Lambda_n-\Lambda\right\|_{sp}&=O_p\left(k_{nT}/\sqrt{n}+\delta_{nT}J^{1/2}\right)\\
\left\|A_n-A\right\|&=O_p\left(k_{nT}^{-\zeta}\left(k_{nT}/\sqrt{n}+\delta_{nT}J^{1/2}\right)\right)\\
\left\|\Lambda_n^{-1}-\Lambda^{-1}\right\|_{sp}&=O_p\left(k_{nT}/\sqrt{n}+\delta_{nT}J^{1/2}\right)
\end{aligned}
$$

where we denote

$$
\Lambda_n=\mathbb{E}_n\left(\widetilde{\hat{\mathbf{q}}}_{1,i}'\widetilde{\hat{\mathbf{d}}}_i\right)\Omega_n\mathbb{E}_n\left(\widetilde{\hat{\mathbf{d}}}_i'\widetilde{\hat{\mathbf{q}}}_{1,i}\right),\ \Lambda=\mathbb{E}\left(\tilde{\mathbf{q}}_{1,i}'\tilde{\mathbf{d}}_i\right)\Omega_n\mathbb{E}\left(\tilde{\mathbf{d}}_i'\tilde{\mathbf{q}}_{1,i}\right) \tag{F.6}
$$

$$
A_n=\mathbb{E}_n\left(\widetilde{\hat{\mathbf{q}}}_{1,i}'\widetilde{\hat{\mathbf{d}}}_i\right)\Omega_n\mathbb{E}_n\left(\widetilde{\hat{\mathbf{d}}}_i'\tilde{\mathbf{\Delta}}_i\right),\ A=\mathbb{E}\left(\tilde{\mathbf{q}}_{1,i}'\tilde{\mathbf{d}}_i\right)\Omega_n\mathbb{E}\left(\tilde{\mathbf{d}}_i'\tilde{\mathbf{\Delta}}_i\right). \tag{F.7}
$$

Proof of Lemma F.3: For any comparable matrices A, B, C, and D, we have $A\Omega_n C' - B\Omega_n D' = (A-B)\,\Omega_n\,(C-D)' + (A-B)\,\Omega_n D' + B\Omega_n\,(C-D)'$, so that

$$\begin{aligned}\|A\Omega_n C' - B\Omega_n D'\|_{sp} \leq \|A-B\|_{sp}\|\Omega_n\|_{sp}\|C-D\|_{sp} + \|A-B\|_{sp}\|\Omega_n\|_{sp}\|D\|_{sp} \\ + \|B\|_{sp}\|\Omega_n\|_{sp}\|C-D\|_{sp}.\end{aligned} \tag{F.8}$$

Hence, by Lemmas F.1 and F.2 and under Assumption 3, we have

$$\begin{aligned}&\|\Lambda_n - \Lambda\|_{sp} \\ &\leq \left\|\mathbb{E}_n\left(\widetilde{\hat{\mathbf{q}}}_{1,i}'\widetilde{\hat{\mathbf{d}}}_i\right) - \mathbb{E}\left(\tilde{\mathbf{q}}_{1,i}'\tilde{\mathbf{d}}_i\right)\right\|_{sp}^2 \|\Omega_n\|_{sp} \\ &+ 2\left\|\mathbb{E}_n\left(\widetilde{\hat{\mathbf{q}}}_{1,i}'\widetilde{\hat{\mathbf{d}}}_i\right) - \mathbb{E}\left(\tilde{\mathbf{q}}_{1,i}'\tilde{\mathbf{d}}_i\right)\right\|_{sp} \|\Omega_n^{1/2}\|_{sp}\left\|\Omega_n^{1/2}\mathbb{E}\left(\tilde{\mathbf{d}}_i'\tilde{\mathbf{q}}_{1,i}\right)\right\|_{sp} \\ &= O_p\left(k_{nT}/\sqrt{n} + \delta_{nT}J^{1/2}\right),\end{aligned}$$

and

$$\begin{aligned}&\|A_n - A\|_{sp} \\ &\leq \left\|\mathbb{E}_n\left(\widetilde{\hat{\mathbf{q}}}_{1,i}'\widetilde{\hat{\mathbf{d}}}_i\right) - \mathbb{E}\left(\tilde{\mathbf{q}}_{1,i}'\tilde{\mathbf{d}}_i\right)\right\|_{sp}\|\Omega_n\|_{sp}\left\|\mathbb{E}_n\left(\widetilde{\hat{\mathbf{d}}}_i'\tilde{\mathbf{\Delta}}_i\right) - \mathbb{E}\left(\tilde{\mathbf{d}}_i'\tilde{\mathbf{\Delta}}_i\right)\right\|_{sp} \\ &+ \left\|\mathbb{E}_n\left(\widetilde{\hat{\mathbf{q}}}_{1,i}'\widetilde{\hat{\mathbf{d}}}_i\right) - \mathbb{E}\left(\tilde{\mathbf{q}}_{1,i}'\tilde{\mathbf{d}}_i\right)\right\|_{sp}\|\Omega_n\|_{sp}\left\|\mathbb{E}\left(\tilde{\mathbf{d}}_i'\tilde{\mathbf{\Delta}}_i\right)\right\|_{sp} \\ &+ \left\|\mathbb{E}\left(\tilde{\mathbf{q}}_{1,i}'\tilde{\mathbf{d}}_i\right)\Omega_n^{1/2}\right\|_{sp}\|\Omega_n^{1/2}\|_{sp}\left\|\mathbb{E}_n\left(\widetilde{\hat{\mathbf{d}}}_i'\tilde{\mathbf{\Delta}}_i\right) - \mathbb{E}\left(\tilde{\mathbf{d}}_i'\tilde{\mathbf{\Delta}}_i\right)\right\|_{sp} \\ &= O_p\left(k_{nT}^{-\zeta}\left(k_{nT}/\sqrt{n} + \delta_{nT}J^{1/2}\right)\right).\end{aligned}$$

Next, we have

$$\begin{aligned}\left\|\Lambda_n^{-1} - \Lambda^{-1}\right\|_{sp} &= \left\|\Lambda_n^{-1}\left(\Lambda_n - \Lambda\right)\Lambda^{-1}\right\|_{sp} \\ &\leq \left\|\Lambda_n^{-1}\right\|_{sp}\|\Lambda_n - \Lambda\|_{sp}\left\|\Lambda^{-1}\right\|_{sp} = O_p\left(k_{nT}/\sqrt{n} + \delta_{nT}J^{1/2}\right)\end{aligned}$$

under Assumption 3(ii), where $\|\Lambda_n^{-1}\|_{sp} = \|\Lambda^{-1}\|_{sp} + O\left(\|\Lambda_n - \Lambda\|_{sp}\right)$. This completes the proof of this lemma.

Proof of Theorem B.1: We first verify (i). Applying simple algebra yields

$$\begin{aligned}
\check{\boldsymbol{\theta}}_1 &= \Lambda_n^{-1}\mathbb{E}_n\left(\widetilde{\hat{\mathbf{q}}}_{1,i}'\widetilde{\hat{\mathbf{d}}}_i\right)\Omega_n\mathbb{E}_n\left(\widetilde{\hat{\mathbf{d}}}_i'\tilde{\mathbf{y}}_i\right) \\
&= \boldsymbol{\theta}_{1,0} + \Lambda_n^{-1}\mathbb{E}_n\left(\widetilde{\hat{\mathbf{q}}}_{1,i}'\widetilde{\hat{\mathbf{d}}}_i\right)\Omega_n\left[\mathbb{E}_n\left(\widetilde{\hat{\mathbf{d}}}_i'\tilde{\boldsymbol{\Delta}}_i\right) + \mathbb{E}_n\left(\widetilde{\hat{\mathbf{d}}}_i'\tilde{\boldsymbol{\varepsilon}}_i\right)\right] \\
&= \boldsymbol{\theta}_{1,0} + \Lambda_n^{-1}A_n + \Lambda_n^{-1}\mathbb{E}_n\left(\widetilde{\hat{\mathbf{q}}}_{1,i}'\widetilde{\hat{\mathbf{d}}}_i\right)\Omega_n\mathbb{E}_n\left(\widetilde{\hat{\mathbf{d}}}_i'\tilde{\boldsymbol{\varepsilon}}_i\right) \\
&= \boldsymbol{\theta}_{1,0} + \Lambda^{-1}A + \left(\Lambda_n^{-1}A_n - \Lambda^{-1}A\right) + \Lambda_n^{-1}\mathbb{E}_n\left(\widetilde{\hat{\mathbf{q}}}_{1,i}'\widetilde{\hat{\mathbf{d}}}_i\right)\Omega_n\mathbb{E}_n\left(\widetilde{\hat{\mathbf{d}}}_i'\tilde{\boldsymbol{\varepsilon}}_i\right),
\end{aligned}$$

where $\|b\| = \|b\|_{sp}$ for any vector b, and by Lemmas F.1-F.3 we obtain

$$\begin{aligned}
&\left\|\Lambda_n^{-1}A_n - \Lambda^{-1}A\right\| \\
&\leq \left\|\left(\Lambda_n^{-1} - \Lambda^{-1}\right)(A_n - A)\right\| + \left\|\left(\Lambda_n^{-1} - \Lambda^{-1}\right)A\right\| + \left\|\Lambda^{-1}(A_n - A)\right\| \\
&= O_p\left(\delta_{nT}k_{nT}^{1/2-\zeta}\right),
\end{aligned}$$

and

$$\begin{aligned}
&\left\|\Lambda_n^{-1}\mathbb{E}_n\left(\widetilde{\hat{\mathbf{q}}}_{1,i}'\widetilde{\hat{\mathbf{d}}}_i\right)\Omega_n\mathbb{E}_n\left(\widetilde{\hat{\mathbf{d}}}_i'\tilde{\boldsymbol{\varepsilon}}_i\right)\right\| \\
&\leq \left\|\left(\Lambda_n^{-1} - \Lambda^{-1}\right)\mathbb{E}_n\left(\widetilde{\hat{\mathbf{q}}}_{1,i}'\widetilde{\hat{\mathbf{d}}}_i\right)\Omega_n\mathbb{E}_n\left(\widetilde{\hat{\mathbf{d}}}_i'\tilde{\boldsymbol{\varepsilon}}_i\right)\right\| + \left\|\Lambda^{-1}\mathbb{E}_n\left(\widetilde{\hat{\mathbf{q}}}_{1,i}'\widetilde{\hat{\mathbf{d}}}_i\right)\Omega_n\mathbb{E}_n\left(\widetilde{\hat{\mathbf{d}}}_i'\tilde{\boldsymbol{\varepsilon}}_i\right)\right\| \\
&\leq \left(\left\|\Lambda_n^{-1} - \Lambda^{-1}\right\|_{sp} + \left\|\Lambda^{-1}\right\|_{sp}\right)\left\|\mathbb{E}_n\left(\widetilde{\hat{\mathbf{q}}}_{1,i}'\widetilde{\hat{\mathbf{d}}}_i\right)\Omega_n\mathbb{E}_n\left(\widetilde{\hat{\mathbf{d}}}_i'\tilde{\boldsymbol{\varepsilon}}_i\right)\right\|_{sp} \\
&= O_p(1)\left(\left\|\left(\mathbb{E}_n\left(\widetilde{\hat{\mathbf{q}}}_{1,i}'\widetilde{\hat{\mathbf{d}}}_i\right) - \mathbb{E}\left(\tilde{\mathbf{q}}_{1,i}'\tilde{\mathbf{d}}_i\right)\right)\Omega_n\mathbb{E}_n\left(\widetilde{\hat{\mathbf{d}}}_i'\tilde{\boldsymbol{\varepsilon}}_i\right)\right\|_{sp} + \|\blacklozenge\|_{sp}\right) \\
&= O_p\left(\left(\delta_{nT} + \sqrt{J}\right)/\sqrt{n}\right) \qquad \text{(F.9)}
\end{aligned}$$

where $\blacklozenge = \mathbb{E}\left(\tilde{\mathbf{q}}_{1,i}'\tilde{\mathbf{d}}_i\right)\Omega_n\mathbb{E}_n\left(\widetilde{\hat{\mathbf{d}}}_i'\tilde{\boldsymbol{\varepsilon}}_i\right)$, because $\mathbb{E}\left(\left\|\mathbb{E}\left(\tilde{\mathbf{q}}_{1,i}'\tilde{\mathbf{d}}_i\right)\Omega_n\mathbb{E}_n\left(\tilde{\mathbf{d}}_i'\tilde{\boldsymbol{\varepsilon}}_i\right)\right\|^2\right) = n^{-1}\mathbb{E}\left[\tilde{\boldsymbol{\varepsilon}}_i'\tilde{\boldsymbol{\varepsilon}}_i\mathrm{tr}(\clubsuit)\right]$ $= O(J/n)$;[15] $\clubsuit = \mathbb{E}\left(\tilde{\mathbf{q}}_{1,i}'\tilde{\mathbf{d}}_i\right)\Omega_n\tilde{\mathbf{d}}_i'\tilde{\mathbf{d}}_i\Omega_n\mathbb{E}\left(\tilde{\mathbf{q}}_{1,i}'\tilde{\mathbf{d}}_i\right)$. Hence, under Assumption 4, we have $\left\|\check{\boldsymbol{\theta}}_1 - \boldsymbol{\theta}_{1,0}\right\| = O_p\left(k_{nT}^{-\zeta} + \sqrt{J/n}\right)$ because the bias term, $\Lambda^{-1}A$, is of order $O\left(k_{nT}^{-\zeta}\right)$ under

[15] This is from Assumption 3.

Assumptions 2(i) and 3. Next, since

$$\begin{bmatrix} \check{f}(z) - f_0(z) \\ \check{g}(s) - g_0(s) \\ \check{r}(v) - r_0(v) \end{bmatrix} = \begin{bmatrix} \check{f}(z) - \vartheta'_{z,1,0}\vec{P}^{k_{nT}}(z) \\ \check{g}(s) - \vartheta'_{s,1,0}\vec{P}^{k_{nT}}(s) \\ \check{r}(v) - \vartheta'_{v,1,0}\vec{P}^{k_{nT}}(v) \end{bmatrix} + \begin{bmatrix} \vartheta'_{z,1,0}\vec{P}^{k_{nT}}(z) - f_0(z) \\ \vartheta'_{s,1,0}\vec{P}^{k_{nT}}(s) - g_0(s) \\ \vartheta'_{v,1,0}\vec{P}^{k_{nT}}(v) - r_0(v) \end{bmatrix}$$
$$= \begin{bmatrix} \vec{P}^{k_{nT}}(z)' & 0 & 0 \\ 0 & \vec{P}^{k_{nT}}(s)' & 0 \\ 0 & 0 & \vec{P}^{k_{nT}}(v)' \end{bmatrix} \begin{bmatrix} \check{\boldsymbol{\vartheta}}_{z,1} - \vartheta_{z,1,0} \\ \check{\boldsymbol{\vartheta}}_{s,1} - \vartheta_{s,1,0} \\ \check{\boldsymbol{\vartheta}}_{v,1} - \vartheta_{v,1,0} \end{bmatrix} + O_p\left(k_{nT}^{-\zeta}\right)$$

under Assumption 2(i); then, under Assumption 2(ii), we obtain result (ii) of this theorem, where the subscript "1" has the same meaning as that in $\check{\boldsymbol{\theta}}_1$ and $\boldsymbol{\theta}_{1,0}$.

Under Assumption 4 and by the proof above, we have

$$\begin{bmatrix} \check{\alpha} - \alpha_0 \\ \check{\boldsymbol{\vartheta}}_{z,1} - \vartheta_{z,1,0} \\ \check{\boldsymbol{\vartheta}}_{s,1} - \vartheta_{s,1,0} \\ \check{\boldsymbol{\vartheta}}_{v,1} - \vartheta_{v,1,0} \end{bmatrix} \approx \left[\mathbb{E}\left(\tilde{\mathbf{q}}'_{1,i}\tilde{\mathbf{d}}_i\right)\Omega_n\mathbb{E}\left(\tilde{\mathbf{d}}'_i\tilde{\mathbf{q}}_{1,i}\right)\right]^{-1}\mathbb{E}\left(\tilde{\mathbf{q}}'_{1,i}\tilde{\mathbf{d}}_i\right)\Omega_n\mathbb{E}_n\left(\tilde{\mathbf{d}}'_i\tilde{\boldsymbol{\varepsilon}}_i\right), \tag{F.10}$$

where for any $p_1 \times m_{iv}$ selection matrix S with finite p_1, applying the multivariate central limit theorem yields

$$n^{-1/2}S\sum_{i=1}^{n}\tilde{\mathbf{d}}'_i\tilde{\boldsymbol{\varepsilon}}_i \xrightarrow{d} N\left(0_{p_1}, S\mathbb{E}\left(\tilde{\mathbf{d}}'_i\tilde{\boldsymbol{\varepsilon}}_i\tilde{\boldsymbol{\varepsilon}}'_i\tilde{\mathbf{d}}_i\right)S\right). \tag{F.11}$$

Denoting a $J \times J$ matrix

$$\hat{\Sigma}_n = \Lambda_n^{-1}\mathbb{E}_n\left(\widetilde{\hat{\mathbf{q}}}'_{1,i}\widetilde{\hat{\mathbf{d}}}_i\right)\Omega_n\mathbb{E}_n\left(\widetilde{\hat{\mathbf{d}}}'_i\widehat{\tilde{\boldsymbol{\varepsilon}}}_i\widehat{\tilde{\boldsymbol{\varepsilon}}}'_i\widetilde{\hat{\mathbf{d}}}_i\right)\Omega_n\mathbb{E}_n\left(\widetilde{\hat{\mathbf{d}}}'_i\widetilde{\hat{\mathbf{q}}}'_{1,i}\right)\Lambda_n^{-1} \tag{F.12}$$

and applying the delta method yields

$$\sqrt{n/\hat{\sigma}_\alpha^2}\left(\check{\alpha} - \alpha_0\right) \xrightarrow{d} N(0,1)$$

where we have

$$\hat{\sigma}_\alpha^2 = e'_{1,J}\hat{\Sigma}_n e_{1,J} \xrightarrow{p} \sigma_\alpha^2, \tag{F.13}$$

and σ_α^2 equals $\hat{\sigma}_\alpha^2$ with $\mathbb{E}_n$, $\widetilde{\hat{\mathbf{q}}}_{1,i}$, $\widetilde{\hat{\mathbf{d}}}_i$ and $\widehat{\tilde{\boldsymbol{\varepsilon}}}_i$ replaced by $\mathbb{E}$, $\tilde{\mathbf{q}}_{1,i}$, $\tilde{\mathbf{d}}_i$ and $\tilde{\boldsymbol{\varepsilon}}_i$, respectively, and $e_{1,J}$ is the first column of I_J. In addition, we have

$$\sqrt{n\hat{\Xi}_n^{-1}}\begin{bmatrix}\check{f}(z)-f_0(z)\\ \check{g}(s)-g_0(s)\\ \check{r}(v)-r_0(v)\end{bmatrix}\xrightarrow{d}N(0_3,I_3)$$

where, letting S_0 equal the last $J-1$ rows of the identity matrix I_J, we define

$$\hat{\Xi}_n=\begin{bmatrix}\vec{P}^{k_{nT}}(z)' & 0 & 0\\ 0 & \vec{P}^{k_{nT}}(s)' & 0\\ 0 & 0 & \vec{P}^{k_{nT}}(v)'\end{bmatrix}S_0\hat{\Sigma}_nS_0'\begin{bmatrix}\vec{P}^{k_{nT}}(z) & 0 & 0\\ 0 & \vec{P}^{k_{nT}}(s) & 0\\ 0 & 0 & \vec{P}^{k_{nT}}(v)\end{bmatrix}. \tag{F.14}$$

This completes the proof of this theorem.

Proof of Theorem B.2: By Theorem B.1, we have $\left\|\check{\boldsymbol{\theta}}_1-\boldsymbol{\theta}_{1,0}\right\|=O_p(a_n)$, where $a_n\to 0$ as $n\to\infty$. It follows that $\left\|\check{\boldsymbol{\theta}}-\boldsymbol{\theta}_0\right\|=\left\|P_{\mathcal{J}}'\left(\check{\boldsymbol{\theta}}_1-\boldsymbol{\theta}_{1,0}\right)\right\|=O_p(a_n)$, where $\check{\boldsymbol{\theta}}=P_{\mathcal{J}}'\check{\boldsymbol{\theta}}_1$. Since $Q_n\left(\check{\boldsymbol{\theta}};\psi\right)\le Q_n(\boldsymbol{\theta}_0;\psi)$, for any $\varpi\in R^p$, we have

$$\begin{aligned}
&Q_n(\boldsymbol{\theta}_0+\varpi a_n;\psi)-Q_n\left(\check{\boldsymbol{\theta}};\psi\right)\\
&\ge Q_n(\boldsymbol{\theta}_0+\varpi a_n;\psi)-Q_n(\boldsymbol{\theta}_0;\psi)\\
&=\bar{G}_n(\boldsymbol{\theta}_0+\varpi a_n)'\,\Omega_n\bar{G}_n(\boldsymbol{\theta}_0+\varpi a_n)-\bar{G}_n(\boldsymbol{\theta}_0)'\,\Omega_n\bar{G}_n(\boldsymbol{\theta}_0)\\
&+\sum_{l=1}^{p}\left[p_c(|\theta_{l,0}+\varpi_l a_n|,\psi)-p_c(|\theta_{l,0}|,\psi)\right]\\
&=-2a_n\bar{G}_n\left(\bar{\boldsymbol{\theta}}\right)'\Omega_n\mathbb{E}_n\left(\widetilde{\hat{\mathbf{d}}}_i'\widetilde{\hat{\mathbf{q}}}_i\right)\varpi\\
&+\sum_{l=1}^{p}\left[a_n\varpi_l p_c'(|\theta_{l,0}|,\psi)+\frac{a_n^2\varpi_l^2}{2}p_c''(|\theta_{l,0}+\upsilon_l\varpi_l a_n|,\psi)\right]
\end{aligned}\tag{F.15}$$

where $\bar{\boldsymbol{\theta}}=a\boldsymbol{\theta}_0+(1-a)(\boldsymbol{\theta}_0+\varpi a_n)$ for some $a\in(0,1)$, $\upsilon_l\in(0,1)$ for $l=1,\ldots,p$, and $\partial\bar{G}_n(\boldsymbol{\theta})/\partial\boldsymbol{\theta}'=-\mathbb{E}_n\left(\widetilde{\hat{\mathbf{d}}}_i'\widetilde{\hat{\mathbf{q}}}_i\right)$. First, we have

$$\begin{aligned}
\bar{G}_n\left(\bar{\boldsymbol{\theta}}\right)&=\mathbb{E}_n\left[\widetilde{\hat{\mathbf{d}}}_i'\left(\tilde{\mathbf{y}}_i-\widetilde{\hat{\mathbf{q}}}_i\bar{\boldsymbol{\theta}}\right)\right]\\
&=\mathbb{E}_n\left[\widetilde{\hat{\mathbf{d}}}_i'\left(\tilde{\mathbf{q}}_i\boldsymbol{\theta}_0-\widetilde{\hat{\mathbf{q}}}_i\bar{\boldsymbol{\theta}}\right)\right]+\mathbb{E}_n\left(\widetilde{\hat{\mathbf{d}}}_i'\tilde{\boldsymbol{\Delta}}_i\right)+\mathbb{E}_n\left(\widetilde{\hat{\mathbf{d}}}_i'\tilde{\boldsymbol{\varepsilon}}_i\right)\\
&=\mathbb{E}_n\left[\widetilde{\hat{\mathbf{d}}}_i'J_T\left(P_v^{k_{nT}}-P_{\hat{v}}^{k_{nT}}\right)\right]\boldsymbol{\vartheta}_{v,0}+a_n(1-a)\,\mathbb{E}_n\left(\widetilde{\hat{\mathbf{d}}}_i'\widetilde{\hat{\mathbf{q}}}_i\right)\varpi\\
&+\mathbb{E}_n\left(\widetilde{\hat{\mathbf{d}}}_i'\tilde{\boldsymbol{\Delta}}_i\right)+\mathbb{E}_n\left(\widetilde{\hat{\mathbf{d}}}_i'\tilde{\boldsymbol{\varepsilon}}_i\right).
\end{aligned}$$

It follows that $\bar{G}_n(\bar{\boldsymbol{\theta}})' \Omega_n \mathbb{E}_n\left(\widetilde{\hat{\mathbf{d}}}_i' \widetilde{\hat{\mathbf{q}}}_i\right) \varpi = A_1 + a_n(1-a) A_2 + A_3 + A_4$, where the definition of A_1 to A_4 will be clarified in the context below.

(i) We have

$$\begin{aligned}
\|A_1\| &= \left| \boldsymbol{\vartheta}_{v,0}' \mathbb{E}_n \left[\widetilde{\hat{\mathbf{d}}}_i' J_T \left(P_v^{k_{nT}} - P_{\hat{v}}^{k_{nT}} \right) \right]' \Omega_n \mathbb{E}_n \left(\widetilde{\hat{\mathbf{d}}}_i' \widetilde{\hat{\mathbf{q}}}_i \right) \varpi \right| \\
&\leq \left\| \mathbb{E}_n \left[\widetilde{\hat{\mathbf{d}}}_i' J_T \left(P_v^{k_{nT}} - P_{\hat{v}}^{k_{nT}} \right) \right] \boldsymbol{\vartheta}_{v,0} \times \right\| \|\Omega_n\|_{sp} \left\| \mathbb{E}_n \left(\widetilde{\hat{\mathbf{d}}}_i' \widetilde{\hat{\mathbf{q}}}_i \right) \varpi \right\| \\
&= \|\varpi\| O_p\left(\delta_{nT}^2\right)
\end{aligned}$$

because we have

$$\begin{aligned}
&\left\| \mathbb{E}_n \left[\widetilde{\hat{\mathbf{d}}}_i' J_T \left(P_v^{k_{nT}} - P_{\hat{v}}^{k_{nT}} \right) \right] \boldsymbol{\vartheta}_{v,0} \right\| \\
&= \left\| \mathbb{E}_n \left[I_T \otimes \operatorname{vec}\left(J_T \left(P_{\hat{v},i}^{k_{nT}} - P_{v,i}^{k_{nT}} \right) \right)' J_T \left(P_v^{k_{nT}} - P_{\hat{v}}^{k_{nT}} \right) \right] \boldsymbol{\vartheta}_{v,0} \right\| \\
&\leq \mathbb{E}_n \left\| \left[I_T \otimes \operatorname{vec}\left(J_T \left(P_{\hat{v},i}^{k_{nT}} - P_{v,i}^{k_{nT}} \right) \right)' J_T \left(P_v^{k_{nT}} - P_{\hat{v}}^{k_{nT}} \right) \right] \boldsymbol{\vartheta}_{v,0} \right\| \\
&= \mathbb{E}_n \left(\operatorname{tr}[\bigstar] \boldsymbol{\vartheta}_{v,0}' \left(P_v^{k_{nT}} - P_{\hat{v}}^{k_{nT}} \right)' J_T \left(P_v^{k_{nT}} - P_{\hat{v}}^{k_{nT}} \right) \boldsymbol{\vartheta}_{v,0} \right)^{1/2} \\
&\leq \|\boldsymbol{\vartheta}_{v,0}\| \mathbb{E}_n (\operatorname{tr}[\bigstar]) = O_p\left(\delta_{nT}^2\right)
\end{aligned}$$

by (F.5), where $\bigstar = \left(P_{\hat{v},i}^{k_{nT}} - P_{v,i}^{k_{nT}} \right)' J_T \left(P_{\hat{v},i}^{k_{nT}} - P_{v,i}^{k_{nT}} \right)$, and

$$\begin{aligned}
&\left\| \mathbb{E}_n \left(\widetilde{\hat{\mathbf{d}}}_i' \widetilde{\hat{\mathbf{q}}}_i \right) \varpi \right\| \\
&\leq \left\| \mathbb{E}_n \left(\widetilde{\hat{\mathbf{d}}}_i' \widetilde{\hat{\mathbf{q}}}_i - \mathbb{E}\left(\tilde{\mathbf{d}}_i' \tilde{\mathbf{q}}_i \right) \right) \varpi \right\| + \left\| \mathbb{E}\left(\tilde{\mathbf{d}}_i' \tilde{\mathbf{q}}_i \right) \varpi \right\| \\
&\leq \|\varpi\| \left(\left\| \mathbb{E}_n \left(\widetilde{\hat{\mathbf{d}}}_i' \widetilde{\hat{\mathbf{q}}}_i - \mathbb{E}\left(\tilde{\mathbf{d}}_i' \tilde{\mathbf{q}}_i \right) \right) \right\|_{sp} + \left\| \mathbb{E}\left(\tilde{\mathbf{d}}_i' \tilde{\mathbf{q}}_i \right) \right\|_{sp} \right) \left(1 + O_p\left(k_{nT}/\sqrt{n}\right)\right)
\end{aligned}$$

by Lemma F.2.

(ii) We have $A_2 = \varpi' \mathbb{E}_n \left(\widetilde{\hat{\mathbf{d}}}_i' \widetilde{\hat{\mathbf{q}}}_i \right)' \Omega_n \mathbb{E}_n \left(\widetilde{\hat{\mathbf{d}}}_i' \widetilde{\hat{\mathbf{q}}}_i \right) \varpi = \varpi' \Lambda \varpi + \varpi' (\Lambda_n - \Lambda) \varpi$. As $\varpi' (\Lambda_n - \Lambda) \varpi \leq \|\varpi\| \|\Lambda_n - \Lambda\|_{sp} = \|\varpi\| O_p\left(k_{nT}/\sqrt{n} + \delta_{nT} J^{1/2}\right)$ by Lemma F.3, we obtain $A_2 = \varpi' \Lambda \varpi + \|\varpi\| O_p\left(k_{nT}/\sqrt{n} + \delta_{nT} J^{1/2}\right)$.

(iii) Under Assumption 2(i) and by Lemma F.3, we have

$$|A_3| = \left| \varpi' \mathbb{E}_n \left(\widetilde{\tilde{\mathbf{d}}}_i' \tilde{\boldsymbol{\Delta}}_i \right) \Omega_n \mathbb{E}_n \left(\widetilde{\tilde{\mathbf{d}}}_i' \widetilde{\tilde{\mathbf{q}}}_i \right) \varpi \right| = \|\varpi\| \, O_p \left(k_{nT}^{-\zeta} \left(k_{nT}/\sqrt{n} + \delta_{nT} J^{1/2} + 1 \right) \right).$$

Following the proof of (F.9) we obtain

$$\left| \mathbb{E}_n \left(\widetilde{\tilde{\mathbf{d}}}_i' \tilde{\boldsymbol{\varepsilon}}_i \right)' \Omega_n \mathbb{E}_n \left(\widetilde{\tilde{\mathbf{d}}}_i' \widetilde{\tilde{\mathbf{q}}}_i \right) \varpi \right| = \|\varpi\| \, O_p \left(\left(\delta_{nT} + \sqrt{J} \right) / \sqrt{n} \right).$$

Taking (i)-(iii) together gives

$$\begin{aligned} & \bar{G}_n \left(\bar{\boldsymbol{\theta}} \right)' \Omega_n \mathbb{E}_n \left(\widetilde{\tilde{\mathbf{d}}}_i' \widetilde{\tilde{\mathbf{q}}}_i \right) \varpi - \varpi' \Lambda \varpi \\ & = O_p \left(\delta_{nT}^2 + k_{nT}/\sqrt{n} + \delta_{nT} J^{1/2} + k_{nT}^{-\zeta} + \sqrt{J/n} \right) \|\varpi\| \end{aligned} \tag{F.16}$$

so that $\bar{G}_n \left(\bar{\boldsymbol{\theta}} \right)' \Omega_n \mathbb{E}_n \left(\widetilde{\tilde{\mathbf{d}}}_i' \widetilde{\tilde{\mathbf{q}}}_i \right) \varpi = \varpi' \Lambda \varpi + o_p(1)$ is positive in probability approaching one for any finite $\|\varpi\|$.

Second, the MCP penalty and its first-order derivative are defined as

$$p_c(t, \psi) = \begin{cases} \psi |t| - t^2 c\psi/2 & \text{if } |t| \leq c\psi \\ c\psi^2/2 & \text{otherwise} \end{cases} \tag{F.17}$$

$$p_c'(t, \psi) = \begin{cases} (\psi - |t|/c) \, sign(t) & \text{if } |t| \leq c\psi \\ 0 & \text{otherwise} \end{cases}$$

for $c > 1$ and $p_c''(t, \psi) = -c^{-1} I(|t| \leq c\psi)$, where $sign(t) = -1, 0, 1$ for negative, 0 and positive t, respectively. Hence, setting (c, ψ) satisfying $c\psi < q_0 \min_{l \in \mathcal{J}_n} \{\theta_{l,0}\}$ for some $q_0 \leq 1$, we have $\sum_{l=1}^{p} a_n \varpi_l p_c'(|\theta_{l,0}|, \psi) = 0$ and

$$\begin{aligned} & \sum_{l=1}^{p} \varpi_l^2 p_c''(|\theta_{l,0} + \upsilon_l \varpi_l a_n|, \psi) \\ & = \sum_{l \in \mathcal{J}_n} \varpi_l^2 p_c''(|\theta_{l,0} + \upsilon_l \varpi_l a_n|, \psi) + \sum_{l \notin \mathcal{J}_n} \varpi_l^2 p_c''(|\upsilon_l \varpi_l a_n|, \psi) = -c^{-1} \sum_{l \notin \mathcal{J}_n} \varpi_l^2 \end{aligned}$$

where $p_c''(|\upsilon_l \varpi_l a_n|, \psi) = -c^{-1}$ for $l \notin \mathcal{J}_n$ if $\max_{l \notin \mathcal{J}_n} |\upsilon_l \varpi_l a_n| < a_n \|\boldsymbol{\varpi}\| < c\psi$, and for $l \in \mathcal{J}_n$, $p_c''(|\theta_{l,0} + \upsilon_l \varpi_l a_n|, \psi) = 0$, because $\max_{l \in \mathcal{J}_n} |\theta_{l,0} + \upsilon_l \varpi_l a_n| > \left(q_0^{-1} - 1 \right) c\psi > c\psi$ for $q_0 < 1/2$.

Combining this result with (F.15) and (F.16) gives

$$Q_n\left(\boldsymbol{\theta}_0+\varpi a_n;\psi\right)-Q_n\left(\check{\boldsymbol{\theta}};\psi\right)=2a_n\varpi'\Lambda\varpi-\frac{a_n^2}{2c}\sum_{l\notin\mathcal{J}_n}\varpi_l^2$$
$$+\ O_p\left(\delta_{nT}^2+k_{nT}/\sqrt{n}+\delta_{nT}J^{1/2}+k_{nT}^{-\zeta}+\sqrt{J/n}\right)a_n\left\|\varpi\right\|.$$

Hence, $Q_n\left(\boldsymbol{\theta}_0+\varpi a_n;\psi\right)\geq Q_n\left(\check{\boldsymbol{\theta}};\psi\right)$ holds in probability approaching one if the tuning parameter, ψ, falls into the interval $(a_n\left\|\boldsymbol{\varpi}\right\|/c,\min_{l\in\mathcal{J}_n}\{\theta_{l,0}\}/(cq_0)]$ for some $c>1$, $q_0<1/2$ and sufficiently large $\|\boldsymbol{\varpi}\|$. Thus, there exists a local minimizer, $\hat{\boldsymbol{\theta}}(\psi)$, of $Q_n(\boldsymbol{\theta};\boldsymbol{\psi})$ in the ball $\{\boldsymbol{\theta}_0+a_n\left\|\varpi\right\|:\left\|\varpi\right\|\leq M\}$, and combining this result with Theorem B.1 yields

$$\Pr\left(\check{\boldsymbol{\theta}}=\hat{\boldsymbol{\theta}}\left(\psi\right)\right)\to 1\text{ as }n\to\infty.$$

This completes the proof of this Theorem.

Proof of Theorem B.3: Let $\widehat{\mathcal{J}}$ =supp$\left(\hat{\boldsymbol{\theta}}\right)$ and $\hat{J}=dim\left(\widehat{\mathcal{J}}\right)$ and split $\widetilde{\tilde{\mathbf{q}}}_i=\left[\widetilde{\tilde{\mathbf{q}}}_{\widehat{\mathcal{J}},1,i},\widetilde{\tilde{\mathbf{q}}}_{\widehat{\mathcal{J}},2,i}\right]$ and $\tilde{\boldsymbol{\theta}}=\left[\tilde{\boldsymbol{\theta}}_1',0_{p-\hat{J}}'\right]'$, where the parameters in front of $\widetilde{\tilde{\mathbf{q}}}_{\widehat{\mathcal{J}},2,i}$ are all zeros and $\widetilde{\tilde{\mathbf{q}}}_i\ \tilde{\boldsymbol{\theta}}=\widetilde{\tilde{\mathbf{q}}}_{\widehat{\mathcal{J}},1,i}\tilde{\boldsymbol{\theta}}_1$. Additionally, we denote a parameter vector $\boldsymbol{\theta}_{\widehat{\mathcal{J}},0}=\left[\boldsymbol{\theta}_{\widehat{\mathcal{J}},1,0},0_{p-\hat{J}}\right]$ for a chosen $\hat{J}$ where $\boldsymbol{\theta}_{\widehat{\mathcal{J}},0}$ is the parameter vector to which $\hat{\boldsymbol{\theta}}$ converge. By definition of $\tilde{\boldsymbol{\theta}}$, $\bar{G}_n\left(\tilde{\boldsymbol{\theta}}\right)'\Omega_n\bar{G}_n\left(\tilde{\boldsymbol{\theta}}\right)\leq\bar{G}_n\left(\hat{\boldsymbol{\theta}}\right)'\Omega_n\bar{G}_n\left(\hat{\boldsymbol{\theta}}\right)$ and $\bar{G}_n\left(\tilde{\boldsymbol{\theta}}\right)'\Omega_n\bar{G}_n\left(\tilde{\boldsymbol{\theta}}\right)\leq\bar{G}_n\left(\boldsymbol{\theta}_{\widehat{\mathcal{J}},0}\right)'\Omega_n\bar{G}_n\left(\boldsymbol{\theta}_{\widehat{\mathcal{J}},0}\right)$, which implies that

$$\bar{G}_n\left(\tilde{\boldsymbol{\theta}}\right)'\Omega_n\bar{G}_n\left(\tilde{\boldsymbol{\theta}}\right)-\bar{G}_n\left(\boldsymbol{\theta}_0\right)'\Omega_n\bar{G}_n\left(\boldsymbol{\theta}_0\right)\leq\min\left(P_n,S_n\right)$$

where $P_n=\bar{G}_n\left(\hat{\boldsymbol{\theta}}\right)'\Omega_n\bar{G}_n\left(\hat{\boldsymbol{\theta}}\right)-\bar{G}_n\left(\boldsymbol{\theta}_0\right)'\Omega_n\bar{G}_n\left(\boldsymbol{\theta}_0\right)$ and

$$S_n=\bar{G}_n\left(\boldsymbol{\theta}_{\widehat{\mathcal{J}},0}\right)'\Omega_n\bar{G}_n\left(\boldsymbol{\theta}_{\widehat{\mathcal{J}},0}\right)-\bar{G}_n\left(\boldsymbol{\theta}_0\right)'\Omega_n\bar{G}_n\left(\boldsymbol{\theta}_0\right).$$

For any vector $\omega\in R^p$, applying simple algebra gives

$$\bar{G}_n\left(\boldsymbol{\theta}_0+\omega\right)'\Omega_n\bar{G}_n\left(\boldsymbol{\theta}_0+\omega\right)-\bar{G}_n\left(\boldsymbol{\theta}_0\right)'\Omega_n\bar{G}_n\left(\boldsymbol{\theta}_0\right)$$
$$=\omega'\Lambda_n\omega-2\omega'\mathbb{E}_n\left(\widetilde{\tilde{\mathbf{d}}}_i'\widetilde{\tilde{\mathbf{q}}}_i\right)'\Omega_n\bar{G}_n\left(\boldsymbol{\theta}_0\right)\tag{F.18}$$

where we denote $\Lambda_n=\mathbb{E}_n\left(\widetilde{\tilde{\mathbf{d}}}_i'\widetilde{\tilde{\mathbf{q}}}_i\right)'\Omega_n\mathbb{E}_n\left(\widetilde{\tilde{\mathbf{d}}}_i'\widetilde{\tilde{\mathbf{q}}}_i\right)$ since $\bar{G}_n\left(\boldsymbol{\theta}_0+\omega\right)=\bar{G}_n\left(\boldsymbol{\theta}_0\right)-\mathbb{E}_n\left(\widetilde{\tilde{\mathbf{d}}}_i'\widetilde{\tilde{\mathbf{q}}}_i\right)\omega$.

Additionally, we have

$$\bar{G}_n(\boldsymbol{\theta}_0)' \Omega_n \bar{G}_n(\boldsymbol{\theta}_0) = O_p\left(k_{nT}^{-2\zeta} + \delta_{nT}^2 k_{nT} + J/n\right) \tag{F.19}$$

by Lemma F.1, which implies that

$$\begin{aligned}
&\left|\omega' \mathbb{E}_n\left(\widetilde{\hat{\mathbf{d}}}_i' \widetilde{\hat{\mathbf{q}}}_i\right)' \Omega_n \bar{G}_n(\boldsymbol{\theta}_0)\right| \leq \left\|\omega' \mathbb{E}_n\left(\widetilde{\hat{\mathbf{d}}}_i' \widetilde{\hat{\mathbf{q}}}_i\right)' \Omega_n^{1/2}\right\| \left\|\Omega_n^{1/2} \bar{G}_n(\boldsymbol{\theta}_0)\right\| \\
&= (\omega' \Lambda_n \omega)^{1/2} O_p\left(k_{nT}^{-\zeta} + \delta_{nT} k_{nT}^{1/2} + \sqrt{J/n}\right).
\end{aligned}$$

Then, we have

$$\begin{aligned}
&\left|\bar{G}_n\left(\tilde{\boldsymbol{\theta}}\right)' \Omega_n \bar{G}_n\left(\tilde{\boldsymbol{\theta}}\right) - \bar{G}_n(\boldsymbol{\theta}_0)' \Omega_n \bar{G}_n(\boldsymbol{\theta}_0) - \left(\tilde{\boldsymbol{\theta}} - \boldsymbol{\theta}_0\right)' \Lambda_n \left(\tilde{\boldsymbol{\theta}} - \boldsymbol{\theta}_0\right)\right| \\
&\qquad \leq \left[\left(\tilde{\boldsymbol{\theta}} - \boldsymbol{\theta}_0\right)' \Lambda_n \left(\tilde{\boldsymbol{\theta}} - \boldsymbol{\theta}_0\right)\right]^{1/2} O_p\left(k_{nT}^{-\zeta} + \delta_{nT} k_{nT}^{1/2} + \sqrt{J/n}\right).
\end{aligned}$$

Hence, we obtain

$$\left[\left(\tilde{\boldsymbol{\theta}} - \boldsymbol{\theta}_0\right)' \Lambda_n \left(\tilde{\boldsymbol{\theta}} - \boldsymbol{\theta}_0\right)\right]^{1/2} = O_p\left(k_{nT}^{-\zeta} + \delta_{nT} k_{nT}^{1/2} + \sqrt{J/n}\right) + \sqrt{\min\left(P_{n,+}, S_{n,+}\right)} \tag{F.20}$$

where $P_{n,+} = \max(P_n, 0)$ and $S_{n,+} = \max(S_n, 0)$. By (F.18), we have

$$\begin{aligned}
P_n &= \left(\hat{\boldsymbol{\theta}} - \boldsymbol{\theta}_0\right)' \Lambda_n \left(\hat{\boldsymbol{\theta}} - \boldsymbol{\theta}_0\right) - 2\left(\hat{\boldsymbol{\theta}} - \boldsymbol{\theta}_0\right)' \mathbb{E}_n\left(\widetilde{\hat{\mathbf{d}}}_i' \widetilde{\hat{\mathbf{q}}}_i\right)' \Omega_n \bar{G}_n(\boldsymbol{\theta}_0) \\
S_n &= \left(\boldsymbol{\theta}_{\widehat{\mathcal{J}},0} - \boldsymbol{\theta}_0\right)' \Lambda_n \left(\boldsymbol{\theta}_{\widehat{\mathcal{J}},0} - \boldsymbol{\theta}_0\right) - 2\left(\boldsymbol{\theta}_{\widehat{\mathcal{J}},0} - \boldsymbol{\theta}_0\right)' \mathbb{E}_n\left(\widetilde{\hat{\mathbf{d}}}_i' \widetilde{\hat{\mathbf{q}}}_i\right)' \Omega_n \bar{G}_n(\boldsymbol{\theta}_0).
\end{aligned}$$

By Theorem B.1, we have

$$\begin{aligned}
P_{n,+} &\leq \blacktriangledown + \blacktriangledown^{1/2} O_p\left(k_{nT}^{-\zeta} + \delta_{nT_0} k_{nT}^{1/2} + \sqrt{J/n}\right) \\
&= O_p\left(k_{nT}^{-2\zeta} + \frac{J}{n}\right) + O_p\left[\left(k_{nT}^{-\zeta} + \sqrt{\frac{J}{n}}\right)\left(k_{nT}^{-\zeta} + \delta_{nT} k_{nT}^{1/2} + \sqrt{J/n}\right)\right],
\end{aligned}$$

where $\blacktriangledown = \left(\hat{\boldsymbol{\theta}} - \boldsymbol{\theta}_0\right)' \Lambda_n \left(\hat{\boldsymbol{\theta}} - \boldsymbol{\theta}_0\right)$. In addition, denoting the complement set of $\mathcal{J}$ and $\widehat{\mathcal{J}}$ by $\mathcal{J}^c = \{l \in (1, \ldots, p) : \theta_{l,0} = 0\}$ and $\widehat{\mathcal{J}}^c = \left\{l \in (1, \ldots, p) : \theta_{\widehat{\mathcal{J}},l,0} = 0\right\}$, respectively, we have $0 \leq m = \dim\left(\mathcal{J}^c \cap \widehat{\mathcal{J}}^c\right) \leq p - J$, where $m = 0$ if $\mathcal{J} \cap \widehat{\mathcal{J}} = \varnothing$ and $m = p - J$ if $\mathcal{J} = \widehat{\mathcal{J}}$.

It is readily seen that $\boldsymbol{\theta}_{\widehat{\mathcal{J}},0} = \boldsymbol{\theta}_0$ so that $S_n = 0$ if $\mathcal{J} \subseteq \widehat{\mathcal{J}}$. If $\mathcal{J} \subseteq \widehat{\mathcal{J}}$ fails to hold, we have

$$\begin{aligned} S_{n,+} &\leq \nabla + \nabla^{1/2} O_p\left(k_{nT}^{-\zeta} + \delta_{nT} k_{nT}^{1/2}\right) \\ &= O_p\left(p - m\right) + O_p\left(\sqrt{p - m}\left(k_{nT}^{-\zeta} + \delta_{nT} k_{nT}^{1/2} + \sqrt{J/n}\right)\right), \end{aligned}$$

where $\nabla = \left(\boldsymbol{\theta}_{\widehat{\mathcal{J}},0} - \boldsymbol{\theta}_0\right)' \Lambda_n \left(\boldsymbol{\theta}_{\widehat{\mathcal{J}},0} - \boldsymbol{\theta}_0\right)$. Therefore, (F.20) is bounded by

$$\left(\tilde{\boldsymbol{\theta}} - \boldsymbol{\theta}_0\right)' \Lambda_n \left(\tilde{\boldsymbol{\theta}} - \boldsymbol{\theta}_0\right) = O_p\left(k_{nT}^{-2\zeta} + \frac{J}{n}\right).$$

By Lemma F.3, we have $\|\Lambda_n - \Lambda\|_{sp} = O_p\left(k_{nT}/\sqrt{n} + \delta_{nT} J^{1/2}\right)$, where $\Lambda = \mathbb{E}\left(\tilde{\mathbf{d}}_i' \tilde{\mathbf{q}}_i\right)' \Omega_n \mathbb{E}\left(\tilde{\mathbf{d}}_i' \tilde{\mathbf{q}}_i\right)$ is a finite non-singular matrix under Assumption 3. Hence, $\left\|\tilde{\boldsymbol{\theta}} - \boldsymbol{\theta}_0\right\| = O_p\left(k_{nT}^{-\zeta} + \sqrt{J/n}\right)$. This completes the proof of this theorem.

References

Alfaro, I., Bloom, N., Lin, X., 2018. The finance uncertainty multiplier. NBER Working Paper No. 24571 .

Baker, S.R., Bloom, N., Davis, S.J., 2016. Measuring economic policy uncertainty. The Quarterly Journal of Economics 131, 1593–1636.

Belloni, A., Chen, D., Chernozhukov, V., Hansen, C., 2012. Sparse models and methods for optimal instruments with an application to eminent domain. Econometrica 80, 2369–2429.

Carhart, M.M., 1997. On persistence in mutual fund performance. The Journal of Finance 52, 57–82.

Erickson, T., Jiang, C.H., Whited, T.M., 2014. Minimum distance estimation of the errors-in-variables model using linear cumulant equations. Journal of Econometrics 183, 211–221.

Erickson, T., Parham, R., Whited, T.M., 2017. Fitting the errors-in-variables model using high-order cumulants and moments. The Stata Journal 17, 116–129.

Erickson, T., Whited, T.M., 2000. Measurement error and the relationship between investment and q. Journal of Political Economy 108, 1027–1057.

Gabaix, X., Koijen, R.S., 2020. Granular instrumental variables. Technical Report. National Bureau of Economic Research.

Gu, L., Hackbarth, D., Johnson, T., 2018. Inflexibility and stock returns. The Review of Financial Studies 31, 278–321.

Kim, H., Kung, H., 2017. The asset redeployability channel: How uncertainty affects corporate investment. The Review of Financial Studies 30, 245–280.

Li, D., Sun, Y., 2022. The impact of uncertainty on investment: Empirical challenges and a new estimator. Forthcoming to The Journal of Financial and Quantitative Analysis .

Meijer, E., Spierdijk, L., Wansbeek, T., 2017. Consistent estimation of linear panel data models with measurement error. Journal of Econometrics 200, 169–180.

Panousi, V., Papanikolaou, D., 2012. Investment, idiosyncratic risk, and ownership. The Journal of Finance 67, 1113–1148.