Appendix to Spillovers on the Mean and Tails: A Semiparametric Dynamic Panel Modeling Approach

Yu-Fan Huang[1], Taining Wang[2]

International School of Economics and Management,
Capital University of Economics and Business,
Beijing, 100070, China.

Subal Kumbhakar[3]

Department of Economics
Binghamton University, NY, USA
&
Department of Economics
Inland Norway University of Applied Sciences
Lillehammer, Norway

This appendix provides supporting materials for "Spillovers on the Mean and Tails: A Semiparametric Dynamic Panel Modeling Approach." Appendix 1 lists technical assumptions and provides proofs of theorems 1 and 2. Appendix 2 illustrates numeric properties of our proposed two-step estimator through Monte-Carlo studies. Appendix 3 outlines MCMC algorithm for skewed-normal stochastic volatility model in Section 3 of the paper.

Keywords: US Financial Uncertainty; International Credit Channel; Nonlinear Spillover Effect; Expectations.

JEL Code: C14; C33; E32; E44; F44.

Appendix 1: Technical Assumptions and Proofs

Appendix 1.1: Assumptions

This section lists the assumptions used to prove Theorems 1 and 2. We denote a generic constant by C, whose magnitude is inconsequential for the asymptotic analysis and can vary from one place to another. A generic function $\zeta(v) \in C^j$ if $\zeta(v)$ and all of its partial derivatives of order less than or equal to j are continuous and uniformly bounded on its support. For any vector A, define the norm $||A|| = \sqrt{A^\top A}$. Assumptions:

A1 (1) Random sample $\{y_{i,t}, \omega_{i,t}^\top\}_{i=1}^n$ is identical and independent (i.i.d.) across $i = 1, \ldots, n$ for each fixed t, $\{y_{i,t}, \omega_{i,t}^\top, x_t, z_t\}_{t=1}^T$ is stationary and strongly mixing, with α-mixing coefficient $\alpha_i(t)$ for each fixed

[1]Email: yufanh@cueb.edu.cn
[2]Email: taining.wang@cueb.edu.cn
[3]Email: kkar@binghamton.edu

i. (2) Let $p_n = p + n - 1$ be the dimension of $\vartheta = (\theta^\top, \alpha_{-1}^\top)$, where $p = d_w + K_\sigma + K_\lambda + 4$ is the dimension of $\theta = (\delta^\top, \gamma^\top)$. Then $\alpha_i(t) \le \alpha(t) \le Ct^{-B}$ for some constant $B > p_n + \frac{1}{2}$. (3) z_t has a compact support $\mathcal{Z} \subset \Re$. (4) Fixed effect α_i satisfies $\sum_{i=1}^n \alpha_i = 0$ and allows $E(\alpha_i | \omega_{i,t}, x_t, z_t) \neq 0$.

A2 (1) $e_{i,t}$ is i.i.d. conditioning on $(\omega_{i,t}^\top, x_t, z_t, \alpha_i)$. (2) $E(|\bar{\epsilon}_t|^{2+\delta}|x_t, z_t) = E(|\bar{\epsilon}_t|^{2+\delta}|z_t) \in C^1, \forall \delta > 0$. (3) $E(\bar{\epsilon}_t^2|z_t) \equiv \sigma_{\bar{\epsilon}}^2(z_t) \in C^2$.

A3 (1) $\Theta \subset \Re^{p_n}$ is a compact subset of $\Re^{p_n}$, and its interior points contain $\vartheta_0 = (\theta_0^\top, \alpha_{0,-1}^\top)$. (2) $\forall \vartheta \in \Theta$, if $\vartheta \neq \vartheta_0$, $l(e_{i,t}(\vartheta); x_t, z_t) \neq l(e_{i,t}(\vartheta_0); x_t, z_t)$, where $e_{i,t}(\vartheta) = \tilde{y}_{i,t} - \tilde{\omega}_{i,t}^\top \delta - d_{i,-1}^\top \alpha_{-1} + \frac{1}{n}\sum_{i=1}^n e_{i,t}$. (4) If $\{\vartheta_s\}_{s=1,2,\ldots}$ is a random sequence in Θ such that $\vartheta_s \to \vartheta$ as $s \to \infty$, then $\ln l(e(\vartheta_s); x, z) \to \ln l(e(\vartheta); x, z), \forall \vartheta \in \Theta$. (5) For ϑ_j being the j-th element of ϑ with $j = 1, \ldots, p_n$, $E\left[\sup_{\vartheta_j \in \Theta} |\ln l(e_{i,t}(\vartheta); x_t, z_t)|^\delta\right] < C$, $E\left[\sup_{\vartheta_j \in \Theta} |\frac{\partial}{\partial \vartheta_j} \ln l(e_{i,t}(\vartheta); x_t, z_t)|^\delta\right] < C$.
(6) $\sup_{\gamma \in \Theta} |\lambda(x_t, z_t; \gamma)|^2 < C$, $\sup_{\gamma \in \Theta} |\sigma(x_t, z_t; \gamma)|^2 < C$, and $\inf_{\gamma \in \Theta} |\sigma(x_t, z_t; \gamma)| > C > 0, \forall (x_t, z_t)$.

A4 (1) $\ln l(e(\vartheta); x, z) \in C^2$ with respect to $\vartheta \in \Theta$. (2) $\ln l(e(\vartheta); x, z)$ is continuously differentiable on an open ball $S(\vartheta, d(\vartheta))$ centered at ϑ with radius $d(\vartheta) > 0$. (3) $E(v_{i,t}) = E(u_{i,t}) = 0$, where $v_{i,t} = \frac{\partial}{\partial \alpha_i} \ln l(e_{i,t}(\vartheta); x_t, z_t)|_{\alpha_i = \alpha_{0i}}$ and $u_{i,t} = \frac{\partial}{\partial \theta} \ln l(e_{i,t}(\vartheta); x_t, z_t)|_{\theta = \theta_0}$. (4) Define $v_{i,t}^j \equiv \partial v_{i,t}/\partial j$ and $u_{i,t}^j \equiv \partial u_{i,t}/\partial j$ for $j \in \{\alpha_i, \theta\}$, $\mathcal{H}_{\alpha_{0i}} = E(v_{i,t}^{\alpha_i})$ and $\mathcal{H}_{\theta_0} = E\left[u_{it}^\theta - \frac{E(u_{i,t}v_{i,t})}{E(v_{i,t})} v_{i,t}^\theta\right]$ are finite and negative definite matrix. $\Sigma_{\alpha_{0i}} = E(v_{i,t}^2)$ and $\Sigma_{\theta_0} = E\left[u_{i,t} - \frac{E(u_{i,t}v_{i,t})}{E(v_{i,t})} v_{i,t}\right]\left[u_{i,t} - \frac{E(u_{i,t}v_{i,t})}{E(v_{i,t})} v_{i,t}\right]^\top$ are symmetric and positive definite matrix. (5) $\int \sup_{\vartheta \in S(\vartheta_0, d(\vartheta_0))} ||\frac{\partial}{\partial \vartheta} \ln l(e; x, z)|| de < C$, with $e \equiv e(\vartheta_0)$.

B1 (1) $k(v) : \Re \to \Re$ is univariate and symmetric kernel function. (2) For $j = 0, 1, 2, 3$, (i) $|k(v)v^j| < C$; (ii) $\int |k(v)v^j| dv < C$; (iii) $\int k(v) dv = 1$, $\int k(v)^s dv = 0$ for $s = 1, 3, \ldots$, $\int k(v)v^2 dv \equiv \mu_{k,2} < C$, and $\int k(v)v^4 dv < C$.

B2 (1) Marginal density of z is $f_z(z) \in C^2$, $\inf_{z \in \mathcal{Z}} f_z(z) > 0$. (2) For $t \neq s = 1, \ldots, T$, $E(|x_t|^{2+\delta}|z_t) < C, \forall \delta > 0$. $E(x_t^j x_s^{j'}|z_t) \in C^2$, $\inf_{z \in \mathcal{Z}} E(x_t^j x_s^{j'}|z) > 0, \forall (j, j') \in \{0, 1\}$ and $j \neq j'$, $||\omega||^2 < C$. (3) $\beta(z) \in C^2$. (4) $E(\bar{\epsilon}_1 \bar{\epsilon}_{t+1}|x_1, x_{t+1}, z_1, z_{t+1}) < C$ and continuously differentiable in (z_1, z_{t+1}). (5) $\Sigma_{\beta_0} = [1 - E(x_t|z_t)^2 E(x_t^2|z_t)^{-1}] > 0$ is finite. Define finite constants $C_{\bar{\epsilon}}(t) = E(\bar{\epsilon}_1 \bar{\epsilon}_{t+1}) < C$ and $C_{x|z}(t) = E\left[(1 - x_1 E(x^2|z_1)^{-1} E(x|z_1))(1 - x_{t+1} E(x^2|z_{t+1})^{-1} E(x|z_{t+1}))\right] < C$, $\Omega_{\beta_0} = \Sigma_{\beta_0} \sigma_{\bar{\epsilon}}^2(z_t) + 2\sum_{t=1}^\infty C_{\bar{\epsilon}}(t) C_{x|z}(t) < C$.

B3 (1) As $n \to \infty$ and $T \to \infty$, $n/T \to 0$. (2) As $T \to \infty$, $Tb^3 \to \infty$, $Tb^7 \to 0$. (3) For $s > 2$ and some $\delta > 0$, $T^{B+1.5(0.2+\delta)-B/2+1.25} b^{-1.75-0.5(1+B)} \ln(T)^{0.5B-0.25} \to 0$, $T^{1-2/5-2\delta} b \to \infty$. (4) For $\delta > 0$, $T^{\frac{B-1}{B}} b^{\frac{2+\delta}{1+\delta}} \to \infty$ as $T \to \infty$ with B defined in A1(3).

Assumption A1(1) focuses on panel data in which observations are i.i.d. across i and stationary across t with α-mixing coefficients. A2(2) requires that the mixing coefficients satisfy a certain order, which allows us to characterize the covariance structure of the panel data asymptotically (Chen et al., 2013). A1(3) assumes the support of z_t to be compact that eases arguments on the local approximation of $\beta(\cdot)$. A1(4) normalizes fixed effects α_i for identification (Su and Ullah, 2006; Sun et al., 2009), and allows α_i to be arbitrarily correlated with observables. A2(1) assumes the conditional i.i.d. property of the error term facilitating the construction of likelihood function (Wooldridge, 2010). A2(2)-(3) place the usual boundedness conditions on the higher conditional moment of $e_{i,t}$ for the analysis of asymptotic distribution analysis. A3(1)-(5) guarantee that a unique maximum of $E(\ln(e(\vartheta); x, z))$ exists at ϑ by Theorem 2.5 in Newey and McFadden (1994). A3(6) ensures that the scale and shape functions are bounded and well defined for all x_t and z_t. A4(1)-(5) assumes that higher-order derivatives of the log likelihood function are smooth and bounded to derive asymptotic normality of $\hat{\theta}$ and $\hat{\alpha}_i$. In particular, A4(3) ensures that the zero moment conditions hold for maximizer $\hat{\theta}$ and $\hat{\alpha}_i$ to exist with the corresponding (asymptotic) covariance matrix given in A4(4). A4(5) follows Lemma 3.6 in Newey and McFadden (1994) to allow the order of differentiation and integration to be interchanged, which facilitates the investigation of the normality of $\hat{\vartheta}$.

Assumption B1 gives standard moment and smoothness conditions on the kernel function, which are satisfied by popular second-order Gaussian kernel. As stated in B2(3), the coefficient function $\beta(\cdot)$ to be estimated is continuously differentiable up to the second degree, so there is no need to explore the gain of using a higher-order kernel for a higher degree of the smoothness of $\beta(\cdot)$. B2(1) requires that the marginal density of z is smooth, finite, and bounded away from zero. Those conditions allow us to perform an asymptotic analysis of the estimator $\hat{\beta}(z)$. B2(2) places conditional moment conditions involving x that are continuously differentiable, making it feasible for the asymptotic analysis of the intercept β_0 and the function $\beta(\cdot)$. B2(4) states that the conditional variance of the error term is to be bounded and locally expanded around (z_1, z_{t+1}) by the stationary property of our data in time dimension. B2(5) defines a finite variance term associated with $\hat{\beta}_0$, which reduces to $\Sigma_{\beta_0}^{-1}\sigma_{\bar{\epsilon}}^2(z)$ for independent data in the time dimension (Fan and Huang, 2005). Finally, B3(1) requires $n \to \infty$ slower than $T \to \infty$, so $n/T \to 0$ eliminates the bias of $\hat{\theta}$ driven by the estimated fixed effects $\hat{\alpha}_i$. B3(2) governs the rates at which T diverges toward infinity while b shrinks toward zero. This rate allows for MSE-optimal bandwidth b_{cv} to be used (see Remark 3). B3(3) provides specific rates of T and b for Lemma 1 to hold, which corresponds to those rates in nonparametric regression model with dependent data (Martins-Filho and Yao, 2009). Finally, B3(4) is the sufficient condition for applying Lyapounov central limit theorem on the distribution of second-step estimator (Cai and Li, 2008).

Appendix 1.2: Proofs

The following Lemmas 1-3 are used to prove Theorems 1 and 2.

Lemma 1. *Define* $\hat{S}_j(z) = \frac{1}{Tb}\sum_{t=1}^{T} k\left(\frac{z_t - z}{b}\right)\left(\frac{z_t - z}{b}\right)^j x_t^s$*, for* $j = 0, 1, 2$ *and* $s = 1, 2$*. As* $T \to \infty$*, with assumptions A1(1)-(2), B1, B2(1)-(2), B3(2)-(3), for* $z \in \mathcal{Z}$*,*

(a) $\sup_{z\in\mathcal{Z}} |\hat{S}_0(z) - E(x^s|z)f_z(z)| = O_p\left(\sqrt{\frac{\ln(T)}{Tb}}\right)$.

(b) $\sup_{z\in\mathcal{Z}} |\hat{S}_1(z)| = O_p\left(b + \sqrt{\frac{\ln(T)}{Tb}}\right)$.

(c) $\sup_{z\in\mathcal{Z}} |\hat{S}_2(z) - E(x^s|z)f_z(z)\mu_{k,2}| = O_p\left(\sqrt{\frac{\ln(T)}{Tb}}\right)$.

Proof. The arguments for the uniform convergence rate of $\hat{S}_j(z)$ follows tightly to Theorem 1 in Martins-Filho and Yao (2009), except that one needs to account for the dependence of x_t^s and z_t through the conditional mean function $E(x^s|z_z)$, which is bounded and continuously differentiable by B2(2). The rest of the proof is thus omitted for brevity. □

Lemma 2. *Define* $\mathcal{F}_a^b$ *as the* σ*-algebra of events generated by a random variable* $\{x_t : a \leq t \leq b\}$*. For* u *and* v *as two random variables that are* $\mathcal{F}_{-\infty}^{s}$*- and* $\mathcal{F}_{s+t}^{\infty}$*-measurable, respectively, denote* $||u||_p < C$ *and* $||v||_q < C$*, where* $||u||_p = [E|u|^p]^{\frac{1}{p}}$ *such that* $p, q > 1$ *and* $\frac{1}{p} + \frac{1}{q} < 1$*. Then*

$$|E(uv) - E(u)E(v)| \leq 8\alpha(t)^r ||u||_p ||v||_q,$$

where $r = 1 - \frac{1}{p} - \frac{1}{q}$*.*

Proof. The result is Davydov's inequality in Corollary A2 in Hall and Heyde (1980). □

Lemma 3. *Let* $v_1, \ldots, v_L$ *be* α*-mixing stationary random variables that are* $\mathcal{F}_{i_1}^{j_1}, \ldots, \mathcal{F}_{i_q}^{j_q}$*-measurable, respectively, where* $1 \leq i_1 < j_1 < i_2 < j_2 \cdots < j_q$*,* $i_{l+1} - j_l \geq t$*, and* $|v_l| \leq 1$ *for* $l = 1, \ldots, q$*. Then*

$$\left|E\left(\prod_{l=1}^{q} v_l\right) - \prod_{l=1}^{q} E(v_l)\right| \leq 16(q-1)\alpha(t).$$

Proof. See Lemma 6.1 in Fan and Gijbels (1996). □

Proof of Theorem 1

We prove the consistency property in Theorem 1(a) and normality in Theorem 1(b).

Proof of Theorem 1(a). Define $\mathcal{L}^*_{nT}(\vartheta) = \frac{1}{nT}\sum_{i=1}^{n}\sum_{t=1}^{T}\ln l(e^*_{i,t}(\vartheta)|x_t, z_t)$ with $e^*_{i,t}(\vartheta) = \tilde{y}_{i,t} - \tilde{\omega}_{i,t}^\top\delta - d_{i,-1}^\top\alpha_{-1} + \mu(x_t, z_t;\gamma)$. Also, define $\mathcal{L}_0(\vartheta) = E(\ln l(e_{i,t}(\vartheta)|x_t, z_t))$ be the non-stochastic, probability limit function of $\mathcal{L}_{nT}(\vartheta)$ with $e_{i,t}(\vartheta) = \tilde{y}_{i,t} - \tilde{\omega}_{i,t}^\top\delta - d_{i,-1}^\top\alpha_{-1} + \bar{e}_t$. By definition, $\hat{\vartheta} = \underset{\vartheta\in\Theta}{\operatorname{argmax}}\ \mathcal{L}^*_{nT}(\vartheta)$ and $\vartheta_0 = \underset{\vartheta\in\Theta}{\operatorname{argmax}}\ \mathcal{L}_0(\vartheta)$. With assumptions A3(1)-(3) and Amemiya (1985), if $\mathcal{L}^*_{nT}(\vartheta) \xrightarrow{p} \mathcal{L}_0(\vartheta)$ uniformly over $\vartheta \in \Theta$, then $\hat{\vartheta} \xrightarrow{p} \vartheta_0$ as $n, T \to \infty$.

To see this, let $N_\eta(\vartheta_0) = \{\vartheta : ||\vartheta - \vartheta_0|| < \eta, \eta > 0\}$ be an open ball centered at ϑ_0 with positive radius $\eta > 0$. Also, let $N^c_\eta(\vartheta_0)$ be the complement set of $N_\eta(\vartheta)$, so $N_\eta(\vartheta_0) \cup N_\eta(\vartheta_0)^c = \Theta$. This also implies that $N^c_\eta(\vartheta_0)$ is compact so that the maximizer of $\mathcal{L}^*_{nT}(\vartheta)$ exists by A3(1). Define the event $A_\epsilon = \{\omega : |\mathcal{L}^*_{nT}(\vartheta) - \mathcal{L}_0(\vartheta)| < \frac{\epsilon}{2}, \forall\vartheta \in \Theta\}$. In addition, define the difference $\zeta = \mathcal{L}_0(\vartheta_0) - \max_{\vartheta\in N_\eta(\vartheta_0)\cap\Theta} \mathcal{L}_0(\vartheta)$. First, replacing ϑ by $\hat{\vartheta}$ in A_ϵ gives

$$\mathcal{L}_0(\hat{\vartheta}) > \mathcal{L}_{nT}(\hat{\vartheta}) - \frac{\epsilon}{2} \tag{A.1}$$

by the definition of unique maximizer $\hat{\vartheta}$ for $\mathcal{L}^*_{nT}(\cdot)$. Second, replacing ϑ by ϑ_0 in A_ϵ gives

$$\mathcal{L}^*_{nT}(\vartheta_0) > \mathcal{L}_0(\vartheta_0) - \frac{\epsilon}{2} \tag{A.2}$$

by the definition of unique maximizer ϑ_0 for $\mathcal{L}_0(\cdot)$. However, (A.1) also implies that $\mathcal{L}^*_{nT}(\hat{\vartheta}) > \mathcal{L}^*_{nT}(\vartheta_0)$, indicating that

$$\mathcal{L}_0(\hat{\vartheta}) > \mathcal{L}_{nT}(\vartheta_0) - \frac{\epsilon}{2}. \tag{A.3}$$

Then adding (A.2) and (A.3) gives $\mathcal{L}_0(\hat{\vartheta}) > \mathcal{L}_0(\vartheta_0) - \zeta$, or equivalently $\mathcal{L}_0(\hat{\vartheta}) > \max_{\vartheta\in N_\eta(\vartheta_0)\cap\Theta} \mathcal{L}_0(\vartheta)$. This implies $\hat{\vartheta} \in N_\eta(\vartheta_0)$, so $A_\epsilon \implies \hat{\vartheta} \in N_\eta(\vartheta_0)$ and thus $P(A_\epsilon) \leq P(\hat{\vartheta} \in N_\eta(\vartheta_0)) = 1$ by the condition that $\mathcal{L}^*_{nT}(\vartheta) \xrightarrow{p} \mathcal{L}_0(\vartheta)$ uniformally. Thus, $P(A_\epsilon) \to 1$ and thus $\hat{\vartheta} \xrightarrow{p} \vartheta_0$ as $n, T \to \infty$ as claimed above.

We now show $\mathcal{L}^*_{nT}(\vartheta) \xrightarrow{p} \mathcal{L}_0(\vartheta)$ uniformally over $\vartheta \in \Theta$. It is sufficient to show that

$$\begin{aligned}\sup_{\vartheta\in\Theta} |\mathcal{L}^*_{nT}(\vartheta) - \mathcal{L}_0(\vartheta)| &\leq \sup_{\vartheta\in\Theta} |\mathcal{L}^*_{nT}(\vartheta) - \mathcal{L}_{nT}(\vartheta)| + \sup_{\vartheta\in\Theta} |\mathcal{L}_{nT}(\vartheta) - \mathcal{L}_0(\vartheta)| \\ &\equiv A_1(\vartheta) + A_2(\vartheta) \\ &= o_p(1).\end{aligned}$$

(1) Focus on $A_1(\vartheta)$, we follow Graves (1927) to apply Gateaux differentials for its order. If G is a normed space and there is a functional $T(g) : G \to \Re$, the Gateaux differentials of $T(\cdot)$ at g with increment $s \in G$ of order 1 is given by $\delta^1 T(g, s) = \frac{\partial}{\partial a} T(g + as)|a = 0$. By Theorem 5 in Graves (1927), a Taylor expansion of $T(g + s)$ around g gives $T(g + s) = T(g) + \int_0^1 \delta^1 T(g + sr, s)dr = T(g) + \int_0^1 \left[\frac{\partial}{\partial a} T(g + s(r + a))|a = 0\right] dr$. In our case, $g = \tilde{y}_{i,t} - \tilde{\omega}_{i,t}^\top\delta - \alpha_i + \bar{e}_t$ and $s = \mu(x_t, z_t;\gamma) - \bar{e}_t$. Following notations in Section 2.2 of the paper,

we observe that

$$\begin{aligned}\mathcal{L}^*_{nT}(\vartheta) &= \frac{1}{nT}\sum_{i=1}^{n}\sum_{t=1}^{T}\ln l(e^*_{i,t}(\vartheta);x_t,z_t) = \frac{1}{nT}\sum_{i=1}^{n}\sum_{t=1}^{T}\ln l(\tilde{y}_{i,t}-\tilde{\omega}_{i,t}^{\top}\delta-\alpha_i+\mu(x_t,z_t;\gamma);x_t,z_t)\\ &= \frac{1}{nT}\sum_{i=1}^{n}\sum_{t=1}^{T}\ln l(\tilde{y}_{i,t}-\tilde{\omega}_{i,t}^{\top}\delta-\alpha_i+\bar{e}_t-(\bar{e}_t-\mu(x_t,z_t;\gamma));x_t,z_t)\\ &\equiv \frac{1}{nT}\sum_{i=1}^{n}\sum_{t=1}^{T}T_{i,t}(g+s)\\ &= \frac{1}{nT}\sum_{i=1}^{n}\sum_{t=1}^{T}\ln l(e_{i,t}(\vartheta);x_t,z_t)+\frac{1}{nT}\sum_{i=1}^{n}\sum_{t=1}^{T}\int_0^1\left[\frac{\partial}{\partial a}T_{i,t}(g+s(r+a))|a=0\right]dr.\end{aligned}$$

Define $\lambda_t=\lambda(x_t,z_t;\gamma_\lambda)$ and $\sigma_t=\sigma(x_t,z_t;\gamma_\sigma)$, the integral part in the second term of the last equality above gives

$$\begin{aligned}&\int_0^1\left[\frac{\partial}{\partial a}T_{i,t}(g+s(r+a))|a=0\right]dr\\ &\quad=\int_0^1\left[\frac{\partial}{\partial a}\ln l(\tilde{y}_{i,t}-\tilde{\omega}_{i,t}^{\top}\delta-\alpha_i+\bar{e}_t-(r+a)(\bar{e}_t-\mu(x_t,z_t;\gamma));x_t,z_t)|a=0\right]dr\\ &\quad=\int_0^1\left[\ln\phi\left(\frac{1}{\sigma_t}(e_{i,t}-(\bar{e}_t-\mu(x_t,z_t;\gamma)))(r+a)\right)|_a=0\right]dr\\ &\qquad+\int_0^1\left[\ln\Phi\left(\frac{\lambda_t}{\sigma_t}(e_{i,t}-(\bar{e}_t-\mu(x_t,z_t;\gamma)))(r+a)\right)|_a=0\right]dr\\ &\quad\equiv R_1(\vartheta)+R_2(\vartheta).\end{aligned}$$

It is easy to see that $R_1(\vartheta)=\int_0^1\frac{1}{\sigma_t^2}\left[e_{i,t}-(\bar{e}_t-\mu(x_t,z_t;\gamma))r\right]dr\left[\bar{e}_t-\mu(x_t,z_t;\gamma)\right]=$ $\frac{1}{\sigma_t}\left[e_{i,t}-\frac{1}{2}r^2(\bar{e}_t-\mu(x_t,z_t;\gamma))|_0^1\right]=\frac{1}{\sigma_t}\left[e_{i,t}-\frac{1}{2}(\bar{e}_t-\mu(x_t,z_t;\gamma))\right]\left[\bar{e}_t-\mu(x_t,z_t;\gamma)\right]$, so

$$\begin{aligned}\sup_{\gamma\in\Theta}|R_1(\vartheta)| &\leq \frac{C}{\inf_{\gamma\in\Theta}\sigma(x_t,z_t;\gamma)}\sup_{\vartheta\in\Theta}|e_{i,t}-\frac{1}{2}(\bar{e}_t-\mu(x_t,z_t;\gamma))|\sup_{\gamma\in\Theta}|\bar{e}_t-\mu(x_t,z_t;\gamma)|\\ &= O_p(n^{-1/2})\end{aligned}$$

because $|e_{i,t}|=O(1)$ by A2(2), and $E\sup_{\gamma\in\Theta}|\bar{e}_t-\mu(x_t,z_t;\gamma)|=O_p(n^{-1/2})$ by the weak law of large number and assumptions A3(5)-(6). $R_2(\vartheta)=\int_0^1\frac{\phi(Q_r)}{\Phi(Q_r)}dr\left[-\frac{\lambda_t}{\sigma_t}(\bar{e}_t-\mu(x_t,z_t;\gamma))\right]$, where $Q_r=e_{i,t}-(\bar{e}_t-\mu(x_t,z_t;\gamma))r$. Note that $\frac{\phi(Q_r)}{\Phi(Q_r)}$ is the inverse Mill's ratio such that $Q_r\to\infty$ leads to $\frac{\phi(Q_r)}{\Phi(Q_r)}\to 0$, while $Q_r\to-\infty$ leads to $\frac{\phi(Q_r)}{\Phi(Q_r)}\to -Q_r$ by L'Hôpital's rule. Thus, $R_2(\vartheta)<\int_0^1|Q_r|dr\left[-\frac{\lambda_t}{\sigma_t}(\bar{e}_t-\mu(x_t,z_t;\gamma))\right]=$ $\left[e_{i,t}-\frac{1}{2}(\bar{e}_t-\mu(x_t,z_t;\gamma))\right]\left[-\frac{\lambda_t}{\sigma_t}(\bar{e}_t-\mu(x_t,z_t;\gamma))\right]$, which gives $\sup_{\vartheta\in\Theta}|R_2(\vartheta)|=O_p(n^{-1/2})$ following similar arguments in $R_1(\vartheta)$. The results together gives $A_1(\vartheta)=o_p(1)$ as $n\to\infty$ in B3(1), as claimed above.

(2) Focus on $A_2(\vartheta)$, let's denote $N=nT$. Given that Θ is compact by A3(1), by Heine-Borel Theorem, the open-covering Θ can have a finite (K_N) number of sub-covering $\{S(\vartheta_k,d(\vartheta_k))\}_{k=1}^{K_N}$, each of which is an

open ball centered at ϑ_k with radius $d(\vartheta_k) > 0$. Then for $\vartheta_k \in S(\vartheta_k, d(\vartheta_k))$,

$$\begin{aligned}
|\mathcal{L}_{nT}(\vartheta) - \mathcal{L}_0(\vartheta)| &= |\mathcal{L}_{nT}(\vartheta) - \mathcal{L}_{nT}(\vartheta_k) + \mathcal{L}_0(\vartheta_k) - \mathcal{L}_0(\vartheta) + \mathcal{L}_{nT}(\vartheta_k) - \mathcal{L}_0(\vartheta_k)| \\
&\leq |\frac{1}{nT}\sum_{i=1}^{n}\sum_{t=1}^{T} \ln l(e_{i,t}(\vartheta); x_t, z_t) - \ln l(e_{i,t}(\vartheta_k); x_t, z_t)| \\
&\quad + |\frac{1}{nT}\sum_{i=1}^{n}\sum_{t=1}^{T} E(\ln l(e_{i,t}(\vartheta_k); x_t, z_t) - \ln l(e_{i,t}(\vartheta); x_t, z_t))| \\
&\quad + |\frac{1}{nT}\sum_{i=1}^{n}\sum_{t=1}^{T} \ln l(e_{i,t}(\vartheta_k); x_t, z_t) - E(\ln l(e_{i,t}(\vartheta_k); x_t, z_t))| \\
&\equiv A_{21} + A_{22} + A_{23}.
\end{aligned}$$

$A_{21} \leq C \sup_{\vartheta_k \in S(\vartheta_k, d(\vartheta_k))} |l(e_{i,t}(\vartheta); x_t, z_t) - \ln l(e_{i,t}(\vartheta_k); x_t, z_t)| = o_p(1)$ almost everywhere by A3(4). Similar arguments applied to A_{22} with dominated convergence theorem, implying that $A_{22} = o_p(1)$. $A_{23} \leq \frac{1}{nT}\sum_{i=1}^{n}\sum_{t=1}^{T} |\ln l(e_{i,t}(\vartheta_k); x_t, z_t) - E(\ln l(e_{i,t}(\vartheta_k); x_t, z_t))|$. For some constant $\Delta\epsilon > 0$, we show that

$$P\left(\left(\frac{\ln(N)}{N}\right)^{-\frac{1}{2}} \max_{1\leq k\leq K_N} |\ln l(e_{i,t}(\vartheta_k); x_t, z_t) - E(\ln l(e_{i,t}(\vartheta_k); x_t, z_t))| > \Delta\epsilon\right) \to 0,\ \forall N > \Delta\epsilon,$$

which implies $A_{23} = o_p(1)$. Let $G_{i,t} = l(e_{i,t}(\vartheta_k); x_t, z_t) - E(\ln l(e_{i,t}(\vartheta_k); x_t, z_t))$ and $\sup_{1\leq i\leq n, 1\leq t\leq T} |G_{i,t}|_\infty < b$ for finite constant b, where $|G|_\infty \equiv \inf a : P(G > a) = 0$ places uniform boundedness on $G_{i,t}$. Let $q = 1, 2, \ldots, \left[\frac{N}{2}\right]$, where $[v]$ takes integer part of a real number v. By Bernstein's inequality in Bosq (1996), for $\epsilon_N = \frac{\ln(N)}{N}^{\frac{1}{2}}\Delta\epsilon$ we have

$$P\left(\frac{1}{N}|\sum_{i=1}^{n}\sum_{t=1}^{T} G_{i,t}| > \epsilon_N\right) \leq 4\left(-\frac{\epsilon_N^2}{8V^2(q)}q\right) + 22(1 + \frac{4b}{\epsilon_N})^{-\frac{1}{2}} q\alpha\left(\left[\frac{N}{2q}\right]\right), \quad \text{(A.4)}$$

where $V^2(q) = \frac{2}{m^2}\sigma^2(q) + \frac{b\epsilon_N}{2}$, $m = \frac{N}{2q}$, and

$$\sigma^2(q) = \max_{0\leq j\leq 2q-1} E\left[([jm] + 1 - jm)Z_{[jm]+1} + Z_{[jm]+2} + \cdots + Z_{[jm]+m} + ((j+1)m - [(j+1)m])Z_{[(j+1)m+1]}\right]^2,$$

where $\{Z_j\}_{j=1}^N = \{G_{i,t}\}_{i=1,t=1}^{n,T}$. We know that

$$\sigma^2(q) \leq \max_{0\leq j\leq 2q-1}\left(\sum_{[jm]<i\leq[(j+1)m+1]} E(Z_i^2) + \sum_{[jm]+1<l\leq[(j+1)m]}\sum_{[jm]+1<i\leq[(j+1)m+1]} |E(Z_i Z_l)|\right).$$

In the first term, by A1(2) we have $\sum_{[jm]<i\leq[(j+1)m+1]} E(Z_i^2) = O(m)$ by stationary property. By A3(5), $E|Z_i|^\delta < C$ for $\delta > 0$. Then by Lemma 2, $E(Z_i Z_l) \leq C\alpha(i-l)^{1-\frac{2}{\delta}}$. Let's fix index l such that, for any $[jm] + 1 < l \leq [(j+1)m]$, $\sum_{[jm]+1<i\leq[(j+1)m+1]} |E(Z_i Z_l)| \leq \sum_{i=1}^{m^*-1} E|Z_l Z_{l+i}| + \sum_{i=1}^{m^*-1} E|Z_l Z_{l-i}|$, for $m^* = [(j+1)m+1] - [jm]$. By A1(2), $\sum_{i=1}^{m^*-1} E|Z_l Z_{l+i}| \leq C\sum_{i=1}^{m^*-1}\alpha(i)^{1-\frac{2}{\delta}} \leq C\sum_{i=1}^{m^*-1} i^{-B(1-\frac{2}{\delta})} < C$, provided that $B > (1-2/\delta)^{-1}$ by A1(2). $\sum_{i=1}^{m^*-1} E|Z_l Z_{l-i}| < C$ by similar arguments, so $\sigma^2(q) < C$. Thus, we find that $mV^2(q) = \frac{2}{m}V^2(q) + \frac{bm}{2}\epsilon_N < C$ if we restrict $m = \frac{1}{b\epsilon_N}$. This implies $m = \left(b\left(\frac{\ln(N)}{N}\right)^{\frac{1}{2}}\Delta\epsilon\right)^{-1}$, and $q = N/2m = \frac{b\Delta\epsilon}{2}(N\ln(N))^{\frac{1}{2}}$.

Now the first term in (A.4) gives $4\exp\left(-\frac{\epsilon_N^2}{8V^2(q)}\frac{N}{2}\frac{2q}{N}\right) = 4\exp\left(-\frac{\epsilon_N^2 N}{16V^2(q)m}\right) < 4\exp\left(-\frac{\Delta\epsilon^2}{16C}\right) = 4N^{-\frac{\Delta\epsilon^2}{16C}}$, where the inequality applies because $\epsilon_N^2 N = \ln(N)\Delta\epsilon^2$ and $mV^2(q) < C$. For N sufficiently large and recall that $m = (b\epsilon_N)^{-1}$, the second term in (A.4) gives $22(1+\frac{4b}{\epsilon_N})^{-\frac{1}{2}} q\alpha\left(\left[\frac{N}{2q}\right]\right) \le C(\frac{b}{\epsilon_N})^{\frac{1}{2}}\frac{N}{2m}[m]^{-B} \le Cb^{1.5+B}N\epsilon_N^{\frac{1}{2}+B}$. Combining above results, the right hand side of (A.4) is bounded by $4N^{-\frac{\Delta\epsilon^2}{16C}} + Cb^{1.5+B}N\epsilon_N^{\frac{1}{2}+B}$. This implies that

$$
\begin{aligned}
P\left(\max_{1\le k\le K_N}|G_{i,t}| > \epsilon_N\right) &\le K_N(4N^{-\frac{\Delta\epsilon^2}{16C}} + Cb^{1.5+B}N\epsilon_N^{\frac{1}{2}+B}) \\
&\le CN^{\frac{p_n}{2}-\frac{\Delta\epsilon^2}{16C}} + CN^{\frac{p_n}{2}+1}b^{1.5+B}\epsilon_N^{\frac{1}{2}+B} \\
&= o_p(1),
\end{aligned}
$$

where the second inequality holds by restricting $K_N < CN^{\frac{p_n}{2}}$, and the last inequality holds by making $\Delta\epsilon^2$ sufficiently large in the first term and letting $B > p_n + \frac{1}{2}$ by A1(2) in the second term. In all, results (1)-(2) completes the proof. □

Proof of Theorem 1(b). Establishing asymptotic normality of $\hat{\vartheta}$ calls for special treatment because fixed effect estimates $\hat{\alpha}_i$ converges slower at rate of $1/\sqrt{T}$ compared to parameters for other interested parameters $\hat{\theta} = (\hat{\delta}^\top, \hat{\gamma}^\top)$. Recall that $\hat{\alpha}_i$ is the maximizer from

$$
\hat{\alpha}_i \equiv \hat{\alpha}_i(\theta) = \operatorname*{argmax}_{\alpha_i\in\Theta} \frac{1}{T}\sum_{t=1}^{T} \ln l(e_{i,t}(\vartheta); x_t, z_t), \tag{A.5}
$$

where the dependence of $\hat{\alpha}_i(\theta)$ with θ is highlighted. Clearly, only T observations are used to provide information for each fixed effect. It has been documented in linear/nonlinear panel regression model (i.e., the error component has zero conditional mean), estimator $\hat{\theta}$ would appear a bias term of order $O(1/T)$ induced by $\hat{\alpha}_i$, which is not asymptotically negligible if T grows at the rate with n such that $n/T \to \rho \neq 0$ (Arellano et al., 2007; Hahn and Moon, 2006; Hahn and Newey, 2004). Below, we show that this observation is continue to hold in our skewed panel model, in which $\hat{\alpha}_i$ induces non-negligible bias in $\hat{\theta}$ of order $1/T$ due to non-centered skew error and dynamic regressor $y_{i,t}$. We eliminate this bias by letting $\rho = 0$, which is a reasonable assumption in our empirical dataset with large T and small n.

First, recall that we partition $\vartheta = (\theta^\top, \alpha_{-1}^\top)$. Expanding $\frac{\partial}{\partial\theta}\mathcal{L}_{nT}(\vartheta)$, evaluated at $(\hat{\alpha}_i, \theta_0)$, around (α_{0i}, θ_0) gives

$$
\frac{\partial}{\partial\theta}\mathcal{L}_{nT}(\vartheta)|_{\hat{\alpha}_i,\theta_0} = \frac{\partial}{\partial\theta}\mathcal{L}_{nT}(\vartheta)|_{\alpha_{0i},\theta_0} + \frac{\partial^2}{\partial\alpha_i\partial\theta}\mathcal{L}_{nT}(\vartheta)|_{\alpha_i^*,\theta_0}(\hat{\alpha}_i - \alpha_{0i}), \tag{A.6}
$$

where α_i^* is trapped between $\hat{\alpha}_i$ and α_{0i}. Similarly, expanding $\frac{\partial}{\partial \alpha_i}\mathcal{L}_{nT}(\vartheta)$, evaluated at $(\hat{\alpha}_i, \theta_0)$, around (α_{0i}, θ_0) gives

$$\frac{\partial}{\partial \alpha_i}\mathcal{L}_{nT}(\vartheta)|_{\hat{\alpha}_i,\theta_0} = \frac{\partial}{\partial \alpha_i}\mathcal{L}_{nT}(\vartheta)|_{\alpha_{0i},\theta_0} + \frac{\partial^2}{\partial^2 \alpha_i^2}\mathcal{L}_{nT}(\vartheta)|_{\alpha_i^*,\theta_0}(\hat{\alpha}_i - \alpha_{0i}). \tag{A.7}$$

As $T \to \infty$, the indicator function $1(\hat{\alpha}_i \in S(\alpha_{0i}, d(\alpha_{0i})) \xrightarrow{p} 1$ by results in Theorem 1(a). Thus, the left-hand-side of (A.7) gives $\left[\frac{\partial}{\partial \alpha_i}\mathcal{L}_{nT}(\vartheta)|_{\hat{\alpha}_i,\theta_0}\right] 1(\hat{\alpha}_i \in S(\alpha_{0i}, d(\alpha_{0i})) = 0$. By A4(2), $\mathcal{L}_{nT}(\vartheta)$ is continuous in $(\alpha_i, \theta^\top)$, so $\sup_{\alpha_i \in S(\alpha_{0i}, d(\alpha_{0i}))} |\frac{\partial^2}{\partial^2 \alpha_i^2}\mathcal{L}_{nT}(\vartheta) - \frac{\partial^2}{\partial^2 \alpha_i^2}\mathcal{L}_0(\vartheta)| = o_p(1)$. Then by Theorem 21.6 in Davidson (1994), $\frac{\partial^2}{\partial^2 \alpha_i^2}\mathcal{L}_{nT}(\vartheta)|_{\alpha_i^*,\theta_0} = \frac{\partial^2}{\partial^2 \alpha_i^2}\mathcal{L}_0(\vartheta)|_{\alpha_{0i},\theta_0} + o_p(1)$, so

$$(\hat{\alpha}_i - \alpha_i) = -\left[\frac{\partial^2}{\partial^2 \alpha_i^2}\mathcal{L}_0(\vartheta)|_{\alpha_{0i},\theta_0}\right]^{-1} \frac{\partial}{\partial \alpha_i}\mathcal{L}_{nT}(\vartheta)|_{\alpha_{0i},\theta_0}(1 + o_p(1)). \tag{A.8}$$

Similar arguments applied to (A.6) to have

$$\frac{\partial}{\partial \theta}\mathcal{L}_{nT}(\vartheta)|_{\hat{\alpha}_i,\theta_0} = \frac{\partial}{\partial \theta}\mathcal{L}_{nT}(\vartheta)|_{\alpha_{0i},\theta_0} + \frac{\partial^2}{\partial \alpha_i \partial \theta}\mathcal{L}_0(\vartheta)|_{\alpha_{0i},\theta_0}(\hat{\alpha}_i - \alpha_{0i}), \tag{A.9}$$

with probability one as $n, T \to \infty$. For notation simplicity, let's denote scalar $v_{i,t}(\alpha_i, \theta) = \frac{\partial}{\partial \alpha_i}\ln l(e_{i,t}(\vartheta))|_{\alpha_i,\theta}$ and $p \times 1$ vector $u_{i,t}(\alpha_i, \theta) = \frac{\partial}{\partial \theta}\ln l(e_{i,t}(\vartheta))|_{\alpha_i,\theta}$. Also, let $v_{i,t}^j(\alpha_i, \theta) \equiv \frac{\partial}{\partial j}v_{i,t}(\alpha_i, \theta)$ be the derivative with respect to $j \in \{\alpha_i, \theta\}$ and similarly for $u_{i,t}^j$. We further drop arguments in $v_{i,t}(\cdot)$ and $u_{i,t}(\cdot)$ when they are evaluated at $(\alpha_{0i}, \theta_0^\top)$, e.g., $v_{i,t} \equiv v_{i,t}(\alpha_{0i}, \theta_0)$. Using conventional identity that, for $j \neq s = 1, \ldots, p$, $E\left[\frac{\partial^2}{\partial \theta_j \partial \theta_s}\ln l(e(\vartheta); x, z)\right] = -E\left[\frac{\partial}{\partial \theta_j}\ln(e(\vartheta); x, z)\frac{\partial}{\partial \theta_s}\ln(e(\vartheta); x, z)\right]$ and $E\left[\frac{\partial^2}{\partial^2 \alpha_i^2}\ln l(e(\vartheta); x, z)\right] = -E\left[\frac{\partial}{\partial \alpha_i}\ln l(e(\vartheta); x, z)\right]^2$, which holds by A4(5) allowing interchangeable integration and derivative (Newey and McFadden, 1994). Thus, as $T \to \infty$, (A.8) reduces to

$$(\hat{\alpha}_i - \alpha_{0i}) = \frac{1}{T}\sum_{t=1}^{T} E(v_{i,t}^2)^{-1} v_{i,t}(1 + o_p(1)), \tag{A.10}$$

which coincides with the stochastic expansion results in Rilstone et al. (1996). Substituting A.10 in (A.9), and notice that $\frac{\partial^2}{\partial \alpha_i \partial \theta}\mathcal{L}_0(\vartheta)|_{\alpha_{0i},\theta_0} = -E(v_{i,t}u_{i,t})$, some algebra shows that (A.9) reduces to a concentrated moment condition for $\hat{\theta}$ as

$$\frac{1}{nT}\sum_{i=1}^{n}\sum_{t=1}^{T} u_{i,t}(\hat{\alpha}_i, \theta_0) = \frac{1}{nT}\sum_{i=1}^{n}\sum_{t=1}^{T}\left[u_{i,t}(\alpha_{0i}, \theta_0) - \frac{E(u_{i,t}v_{i,t})}{E(v_{i,t}^2)}v_{i,t}(\alpha_{0i}, \theta_0)\right]. \tag{A.11}$$

Define $U_{i,t}(\alpha_i, \theta) = u_{i,t}(\alpha_i, \theta) - \frac{E(u_{i,t}v_{i,t})}{E(v_{i,t}^2)}v_{i,t}(\alpha_i, \theta)$, it can be shown that $\hat{\theta}$ is the unique maximizer of the concentrated moment above in an sense that $\sum_{i=1}^{n}\sum_{t=1}^{T} U_{i,t}(\hat{\alpha}, \hat{\theta}) = 0$, with $\hat{\alpha}$ defined in A.5 (Arellano et al., 2007; Hahn and Newey, 2004). Thus, following similar practice above, we expand $U_{i,t}(\hat{\alpha}, \hat{\theta}) = U_{i,t}(\hat{\alpha}, \theta_0) + U_{i,t}^{\theta}(\hat{\alpha}_i, \theta^*)(\hat{\theta} - \theta_0)$ for some θ_j^* trapped between $\hat{\theta}_j$ and θ_{0j}, for $j = 1, \ldots, p$.

Since $\frac{1}{nT}\sum_{i=1}^n\sum_{t=1}^T U_{i,t}^{\theta}(\hat{\alpha}_i,\theta^*)\xrightarrow{p}\mathcal{H}_{\theta_0}+o_p(1)$ defined in A4(4) by arguments in (A.8), we see that

$$\sqrt{nT}(\hat{\theta}-\theta_0) = -\mathcal{H}_{\theta_0}\frac{1}{\sqrt{nT}}\sum_{i=1}^n\sum_{t=1}^T U_{i,t}(\hat{\alpha}_i,\theta_0)(1+o_p(1)) \tag{A.12}$$

$$= -\mathcal{H}_{\theta_0}\frac{1}{\sqrt{nT}}\left[\sum_{i=1}^n\sum_{t=1}^T U_{i,t}+\sum_{i=1}^n\sum_{t=1}^T\left(U_{i,t}^{\alpha_i}(\hat{\alpha}_i-\alpha_{0i})+\frac{1}{2}U_{i,t}^{\alpha_i,\alpha_i}(\hat{\alpha}_i-\alpha_{0i})^2\right)\right](1+o_p(1)),$$

where $U_{i,t}\equiv U_{i,t}(\alpha_{0i},\theta_0)$, $U_{i,t}^{\alpha_i}\equiv\frac{\partial}{\partial\alpha_i}U_{i,t}$, and $U_{i,t}^{\alpha_i,\alpha_i}\equiv\frac{\partial^2}{\partial\alpha_i^2}U_{i,t}$. Substituting $(\hat{\alpha}_i-\alpha_{0i})$ with (A.10), and using conditional i.i.d assumption in A2(1), some tedious derivation shows that $\sum_{i=1}^n\sum_{t=1}^T(U_{i,t}^{\alpha_i}(\hat{\alpha}_i-\alpha_{0i})+\frac{1}{2}U_{i,t}^{\alpha_i,\alpha_i}(\hat{\alpha}_i-\alpha_{0i})^2)=\sum_{i=1}^n b_{i,T}$, where

$$b_{i,T}=\frac{1}{T}\sum_{t=1}^T\sum_{t'=1}^T E(v_{i,t'}^2)^{-1}U_{i,t}^{\alpha_i}v_{i,t'}+\frac{1}{T^2}\sum_{t=1}^T\sum_{t'=1}^T\frac{1}{2}U_{i,t}^{\alpha_i,\alpha_i}E(v_{i,t'}^2)^{-2}v_{i,t'}^2.$$

It is easily see that $E(b_{i,T})\equiv\bar{b}_i=E(v_{i,t}^2)^{-1}[E(U_{i,t}^{\alpha_i}v_{i,t})+\frac{1}{2}E(U_{i,t}^{\alpha_i,\alpha_i})]<C$. A further lengthy calculation (omitted) shows that $\bar{b}_i\neq 0$ in our model because $E(e_{i,t}|x_t,z_t)\neq 0$ and $y_{i,t}$ is included as a dynamic regressor in $\omega_{i,t}$. In a special case where $E(e_{i,t}|x_t,z_t)=0$ (i.e., $\lambda(x_t,z_t;\gamma_{0\lambda})=0$) and $y_{i,t}$ is not included, then $\bar{b}_i=0$ by assuming all regressors are strictly exogenous. Let $B(\theta_0)=-\mathcal{H}_{\theta_0}\frac{1}{n}\sum_{i=1}^n\bar{b}_i$, (A.12) becomes

$$\sqrt{nT}(\hat{\theta}-\theta_0)=\left(-\mathcal{H}_{\theta_0}\frac{1}{\sqrt{nT}}\sum_{i=1}^n\sum_{t=1}^T U_{i,t}+\sqrt{\frac{n}{T}}B(\theta_0)\right)(1+o_p(1)), \tag{A.13}$$

where $B(\theta_0)$ is the bias of $\hat{\theta}$ induced by the fixed effect estimates $\hat{\alpha}_i$, which vanishes by letting $n/T\to 0$ as in B3(1) so that $\sqrt{nT}(\hat{\theta}-\theta_0)=-\mathcal{H}_{\theta_0}\frac{1}{\sqrt{nT}}\sum_{i=1}^n\sum_{t=1}^T U_{i,t}(1+o_p(1))$. Notice that $E(U_{i,t})=0$ and $V(U_{i,t})\equiv\Sigma_{\theta_0}$ in A4(4). By assumptions A1(1), A2(1), and A4(4), we apply Cramer-Rao device and Lindeberg-Levy central limit theorem to have $\frac{1}{\sqrt{nT}}\sum_{i=1}^n\sum_{t=1}^T U_{i,t}\xrightarrow{d}\mathcal{N}(0,\Sigma_{\theta_0})$. Then by Slutsky's Theorem, we arrive the claim in Theorem 1(b). □

Proof of Theorem 2.

We show in turn the result of Theorem 2(a) and Theorem 2(b).

Proof of Theorem 2(a). Recall that $\mathcal{Y}_{i,t}(\vartheta)=y_{i,t}-\omega_{i,t}^\top\delta-\alpha_i-\mu(x_t,z_t;\gamma)$, so $\bar{\mathcal{Y}}_t(\vartheta)=\frac{1}{n}\sum_{i=1}^n\mathcal{Y}_{i,t}=\bar{y}_t-\bar{\omega}_t^\top\delta-\bar{\alpha}-\mu(x_t,z_t;\gamma)$. Notice that by the identification condition $\sum_{i=1}^n\alpha_{0i}=0$ and $\sum_{i=1}^n\hat{\alpha}_i=0$, both $\bar{\mathcal{Y}}_t(\hat{\vartheta})$ and $\bar{\mathcal{Y}}_t(\vartheta_0)$ wipe out the average of fixed effects. Following notations in Section 2.2 of the paper, the regression model considered in the second step (in vector form) gives

$$\bar{\mathcal{Y}}(\theta_0)=\iota_T\beta_0+\beta_T(x,z)+\bar{\epsilon}, \tag{A.14}$$

where the t-th element of $\bar{\epsilon}$ gives $\bar{\epsilon}_t = \bar{e}_t - \mu(x_t, z_t; \gamma)$, satisfying $E(\bar{\epsilon}_t|x_t, z_t) = 0$ by our construction. Our constant estimator β_0 takes the form

$$\hat{\beta}_0 = \left[\iota_T^\top (I_T - S_T)^\top (I_T - S_T)\iota_T\right]^{-1} \iota_T^\top (I_T - S_T)^\top (I_T - S_T)\bar{\mathcal{Y}}(\hat{\theta}). \tag{A.15}$$

For non-stochastic constants $\Sigma_{\beta_0} > 0$ and $\Omega_{\beta_0} < C$ defined in B2(5), we show below that

(a) $\frac{1}{T}\iota_T^\top (I_T - S_T)^\top (I_T - S_T)\iota_T = \Sigma_{\beta_0}(1 + o_p(1))$.

(b) $\sqrt{T}(\hat{\beta}_0 - \beta_0) = \sqrt{T}\Sigma_{\beta_0}^{-1}\frac{1}{T}(I_T - S_T)^\top \bar{\epsilon}(1 + o_p(1))$.

(c) $\frac{1}{\sqrt{T}}(I_T - S_T)^\top \bar{\epsilon} \xrightarrow{d} \mathcal{N}(0, \Omega_{\beta_0})$.

Then by Slutsky's Theorem, (a)-(c) implies $\sqrt{T}(\hat{\beta}_0 - \beta_0) \xrightarrow{d} N(0, \Sigma_{\beta_0}^{-2}\Omega_{\beta_0})$ as claimed in Theorem 2(a).

Proof of (a). Define a $T \times 1$ vector $G = (I_T - S_T)\iota_T = \iota_T - S_T\iota_T$ from

$$G = \begin{pmatrix} 1 - [x_1, 0]^\top \left(\mathcal{X}(z_1)^\top K(z_1)\mathcal{X}(z_1)\right)^{-1} \mathcal{X}(z_1)^\top K(z_1)\iota_T \\ \vdots \\ 1 - [x_T, 0]^\top \left(\mathcal{X}(z_T)^\top K(z_T)\mathcal{X}(z_T)\right)^{-1} \mathcal{X}(z_T)^\top K(z_T)\iota_T \end{pmatrix} \equiv \begin{pmatrix} G_1^\top \\ \vdots \\ G_T^\top \end{pmatrix}.$$

Define $D_b = \begin{pmatrix} 1 & 0 \\ 0 & b \end{pmatrix}$ be a 2×2 diagonal matrix. In $G_t^\top$, $[x_t, 0]^\top \left(\mathcal{X}(z_t)^\top K(z_t)\mathcal{X}(z_t)\right)^{-1} \mathcal{X}(z_t)^\top K(z_t)\iota_T = [x_t, 0]^\top D_b^{-1}\hat{S}(z_t)^{-1}D_b^{-1}\frac{1}{Tb}\mathcal{X}(z_t)^\top K(z_t)\iota_T$, where

$$\hat{S}(z_t) = \frac{1}{Tb}D_b^{-1}\mathcal{X}(z_t)^\top K(z_t)\mathcal{X}(z_t)D_b^{-1} \equiv \begin{pmatrix} \hat{S}_0(z) & \hat{S}_1(z) \\ \hat{S}_1(z) & \hat{S}_2(z) \end{pmatrix}, \tag{A.16}$$

with $\hat{S}_j(z)$ defined in Lemma 1. Thus, we readily obtain $\hat{S}(z_t) = S(z_t) + O_p(\sqrt{\ln(T)/Tb})$, where

$$S(z_t) = \begin{pmatrix} E(x^2|z_t)f_z(z_t) & 0 \\ 0 & E(x^2|z_t)f_z(z_t)\mu_{k,2} \end{pmatrix}$$

is non-stochastic and invertible matrix. Thus,

$$\begin{aligned} G_t^\top &= 1 - x_t E(x^2|z_t)^{-1} f_z(z_t)^{-1} \frac{1}{Tb}\sum_{\tau=1}^{T} k\left(\frac{z_\tau - z_t}{b}\right) x_\tau (1 + O_p(\sqrt{\ln(T)/Tb})) \qquad \text{(A.17)} \\ &= 1 - x_t E(x^2|z_t)^{-1} f_z(z_t)^{-1} \left[E(x|z_t)f_z(z_t) + \frac{1}{Tb}\sum_{\tau=1}^{T} k\left(\frac{z_\tau - z_t}{b}\right) x_\tau - E(x|z_t)f_z(z_t)\right] (1 + O_p(\sqrt{\ln(T)/Tb})) \\ &= 1 - x_t E(x^2|z_t)^{-1} f_z(z_t)^{-1} \left[E(x|z_t)f_z(z_t) + O_p(\sqrt{\ln(T)/Tb})\right] (1 + O_p(\sqrt{\ln(T)/Tb})) \\ &= 1 - x_t E(x^2|z_t)^{-1} E(x|z_t)(1 + O_p(\sqrt{\ln(T)/Tb})), \end{aligned}$$

where $\sup_{z_t \in \mathcal{Z}} |\frac{1}{Tb}\sum_{\tau=1}^{T} k(\frac{z_\tau - z_t}{b})x_\tau - E(x|z_t)f_z(z_t)| = O_p(\sqrt{\ln(T)/Tb})$ is applied on the third equality by Lemma 1. Thus, $\frac{1}{T}G^\top G = \frac{1}{T}\sum_{t=1}^{T} G_t G_t^\top = (1 - E^2(x|z_t)E(x^2|z_t)^{-1}(1 + O_p(\sqrt{\ln(T)/Tb})) = \Sigma_{\beta_0}(1 + o_p(1))$

by B3(2), as claimed in (a).

Proof of (b). Let $\hat{S}_G = \frac{1}{T}G^\top G$, we known from (b) that $\hat{S}_G = \Sigma_{\beta_0}(1 + O_p(\sqrt{\ln(T)/Tb}))$. Based on (A.15), we can readily decompose $\hat{\beta}_0$ to obtain

$$
\begin{aligned}
\sqrt{T}(\hat{\beta}_0 - \beta_0) = &\sqrt{T}\hat{S}_G^{-1}\frac{1}{T}\iota_T^\top(I_T - S_T)^\top(I_T - S_T)\beta_T(x, z)\\
&+ \sqrt{T}\hat{S}_G^{-1}\frac{1}{T}\iota_T^\top(I_T - S_T)^\top(I_T - S_T)(\bar{\mathcal{Y}}(\hat{\vartheta}) - \bar{\mathcal{Y}}(\vartheta_0))\\
&+ \sqrt{T}\hat{S}_G^{-1}\frac{1}{T}\iota_T^\top(I_T - S_T)^\top(I_T - S_T)\bar{\epsilon}\\
\equiv &B_1 + B_2 + V.
\end{aligned}
$$

i) We focus on B_1 by first observing that

$$
\begin{aligned}
B_1 &= \sqrt{T}\Sigma_{\beta_0}^{-1}\frac{1}{T}\iota_T^\top(I_T - S_T)^\top(I_T - S_T)\beta_T(x, z)(1 + O_p(\sqrt{\ln(T)/Tb}))\\
&= \sqrt{T}\Sigma_{\beta_0}^{-1}\frac{1}{T}\sum_{t=1}^{T} G_t\left[x_t\beta(z_t) - [x_t, 0]^\top D_b^{-1}\hat{S}(z_t)^{-1}D_b^{-1}\mathcal{X}(z_t)^\top K(z_t)x_t\beta(z_t)\right](1 + O_p(\sqrt{\ln(T)/Tb})).
\end{aligned}
$$

We know that $|G_t| < C$ as shown in (A.17) for all $t = 1, \ldots, T$. Also, for any $z \in \mathcal{Z}$, we have

$$
\begin{aligned}
&[x_t, 0]^\top D_b^{-1}\hat{S}(z_t)^{-1}D_b^{-1}\mathcal{X}(z)^\top K(z)x_t\beta(z)\\
&\quad= x_tE(x^2|z)^{-1}f_z(z)^{-1}\left[E(x_t^2|z)f_z(z) + \frac{1}{Tb}\sum_{t=1}^{T}k\left(\frac{z_t - z}{b}\right)x_t^2 - E(x_t^2|z)f_z(z)\right]\beta(z_t)\\
&\qquad\qquad\times(1 + O_p(\sqrt{\ln(T)/Tb}))\\
&\quad= x_t\beta(z)(1 + O_p(\sqrt{\ln(T)/Tb})).
\end{aligned}
$$

Thus, $B_1 = \sqrt{T}\Sigma_{\beta_0}^{-1}\frac{1}{T}\sum_{t=1}^{T}\left[1 - x_tE(x^2|z_t)^{-1}E(x|z_t)\right]x_t\beta(z_t)(1 + O_p(\sqrt{\ln(T)/Tb}))O_p(\sqrt{\ln(T)/Tb})$. Since $\left[1 - x_tE(x^2|z_t)^{-1}E(x|z_t)\right]x_t\beta(z_t) \xrightarrow{p} E(x_t\beta(z_t) - x_t^2E(x^2|z_t)^{-1}E(x|z_t)\beta(z_t) = E(\beta(z_t)(E(x|z_t) - E(x|z_t))) = 0$ by law of iterative expectation, the term $\left[x_t - x_t^2E(x^2|z_t)^{-1}E(x|z_t)\right]$ is orthogonal to any function of z_t. Thus, $\frac{1}{T}\sum_{t=1}^{T}\left[1 - x_tE(x^2|z_t)^{-1}E(x|z_t)\right]x_t\beta(z_t) = O_p(1/\sqrt{T})$ by B2(2). This gives $B_1 = \sqrt{T}O_p(1/\sqrt{T})O_p(1 + O_p(\sqrt{\ln(T)/Tb}))O_p(\sqrt{\ln(T)/Tb}) = O_p(1)o_p(1) = o_p(1)$ by B3(2).

(ii) By similar arguments in (i), it is easy to show that

$$
\begin{aligned}
|B_2| &\leq C\sqrt{T}\Sigma_{\beta_0}^{-1}|\frac{1}{T}\sum_{t=1}^{T}[1-x_tE(x^2|z_t)^{-1}E(x|z_t)]|O_p(||\omega||\cdot||\hat{\delta}-\delta_0||+|\mu(x_t,z_t;\hat{\gamma})-\mu(x_t,z_t;\gamma_0)|) \\
&\qquad\qquad\qquad\qquad\qquad\qquad \times(1+O_p(\sqrt{\ln(T)/Tb}))O_p(\sqrt{\ln(T)/Tb}) \\
&= \sqrt{T}O_p(\frac{1}{\sqrt{nT}})(1+O_p(\sqrt{\ln(T)/Tb}))O_p(\sqrt{\ln(T)/Tb}) \\
&= O_p(1/\sqrt{n})o_p(1)=o_p(1)
\end{aligned}
$$

where the first equality applies Cauchy-Schwartz inequality, the second equality uses Theorem 1(b) that $\hat{\theta}-\theta_0=O_p(1/\sqrt{nT})$, and the last equality holds given B3(2). So $B_2=o_p(1)$.

(iii) $V=\sqrt{T}\hat{S}_G^{-1}\frac{1}{T}G^{\top}\bar{\epsilon}-\sqrt{T}\hat{S}_G^{-1}\frac{1}{T}G^{\top}S_T\bar{\epsilon}\equiv V_1-V_2$. First, $V_1=\sqrt{T}\Sigma_{\beta_0}^{-1}\frac{1}{T}(I_T-S_T)\bar{\epsilon}(1+O_p(\sqrt{\ln(T)/Tb}))$. Second,

$$
\begin{aligned}
V_2 &= \sqrt{T}\Sigma_{\beta_0}^{-1}\frac{1}{T}G^{\top}S_T\bar{\epsilon}(1+O_p(\sqrt{\ln(T)/Tb})) \\
&= \sqrt{T}\Sigma_{\beta_0}^{-1}\frac{1}{T}\sum_{t=1}^{T}G_tx_tE(x^2|z_t)^{-1}f_z(z_t)^{-1}\left[\frac{1}{Tb}\sum_{\tau=1}^{T}k\left(\frac{z_\tau-z_t}{b}\right)x_\tau\bar{\epsilon_\tau}\right](1+O_p(\sqrt{\ln(T)/Tb})) \\
&= \sqrt{T}\Sigma_{\beta_0}^{-1}\frac{1}{T}\sum_{t=1}^{T}[1-x_tE(x^2|z_t)^{-1}E(x|z_t)]x_tE(x^2|z_t)^{-1}f_z(z_t)^{-1}O_p(\sqrt{\ln(T)/Tb})(1+O_p(\sqrt{\ln(T)/Tb})) \\
&= \sqrt{T}O_p(1/\sqrt{T})O_p(\sqrt{\ln(T)/Tb})(1+O_p(\sqrt{\ln(T)/Tb})) \\
&= o_p(1)
\end{aligned}
$$

where the third equality uses the fact that $E(\frac{1}{Tb}\sum_{\tau=1}^{T}k\left(\frac{z_\tau-z_t}{b}\right)x_\tau\bar{\epsilon_\tau})=0$ and $\sup_{z_t\in\mathcal{Z}}|\frac{1}{Tb}\sum_{\tau=1}^{T}k\left(\frac{z_\tau-z_t}{b}\right)x_\tau\bar{\epsilon_\tau}|=O_p(\sqrt{\ln(T)/Tb})$ by Lemma 1, as well as $E\{[1-x_tE(x^2|z_t)^{-1}E(x|z_t)]x_tE(x^2|z_t)^{-1}f_z(z_t)^{-1}\}=0$ as discussed in B_1. Thus, $V=V_1+o_p(1)$. Combining results (i)-(iii), we have the claimed result in (b).

Proof of (c). From (b), we see that $\sqrt{T}\hat{S}_G^{-1}\frac{1}{T}G^{\top}(I_T-S_T)\bar{\epsilon}=\sqrt{T}\Sigma_{\beta_0}^{-1}\frac{1}{T}\sum_{t=1}^{T}[1-x_tE(x^2|z_t)^{-1}E(x|z_t)]\bar{\epsilon}_t(1+o_p(1))\equiv\Sigma_{\beta_0}^{-1}\left[\frac{1}{\sqrt{T}}\sum_{t=1}^{T}\tilde{G}_t\right](1+o_p(1))$, where $\tilde{G}_t=[1-x_tE(x^2|z_t)^{-1}E(x|z_t)]\bar{\epsilon}_t$. We see $E(\frac{1}{\sqrt{T}}\sum_{t=1}^{T}\tilde{G}_t)=0$, and by Theorem 2.20 in Fan and Yao (2003), the variance

$$
V(\frac{1}{\sqrt{T}}\sum_{t=1}^{T}\tilde{G}_t)=\frac{1}{T}V(\sum_{t=1}^{T}\tilde{G}_t)=\Sigma_{\beta_0}\sigma_{\bar{\epsilon}}^2(z_t)+2\sum_{t=1}^{\infty}C_{\bar{\epsilon}}(t)C_{x|z}(t)\equiv\Omega_{\beta_0}<C
$$

as defined in B2(5). Then by the small and large block technique that handles dependent data with α-mixing coefficient (see details below in *Proof of Theorem (2)*), by Theorem 2.1 in Fan and Yao (2003),

$$
\frac{1}{\sqrt{T}}\sum_{t=1}^{T}\tilde{G}_t\xrightarrow{d}\mathcal{N}(0,\Sigma_{\beta_0}),
$$

provided that $E|\tilde{G}_t|^\delta < C$ and $\sum_{t>1}\alpha(t)^{1-\frac{2}{\delta}} < C$ for some $\delta > 2$. First, $E|(1-x_tE(x^2|z_t)^{-1}E(x|z_t))\bar{\epsilon}_t|^{2+\delta} = E(|\bar{\epsilon}_t|^{2+\delta}|z_t)E(|1-x_tE(x^2|z_t)^{-1}E(x|z_t)|^{2+\delta}) < CE(|\bar{\epsilon}_t|^{2+\delta}|z_t)(1+E(|x|^{2+\delta}|z_t)|E(x^2|z_t)^{-1}E(x|z_t)|^{2+\delta}) < C$ by A2(2) and B2(2). Second, $\sum_{t>1}\alpha(t)^{1-\frac{2}{\delta}} < \sum_{t=1}^T t^{-B(1-\frac{2}{\delta})} \to 0$ because $\delta > 2$ and $B > 0$ by A1(2). Thus, results in (c) follows.

Combining results (a)-(c), we obtain claimed result in Theorem 2(a). □

Proof of Theorem 2(b).

Consider again the regression in (A.14), suppose β_0 and ϑ were known. This gives us

$$\bar{\mathcal{Y}}_t(\vartheta_0) - \beta_0 = x_t\beta(z_t) + \bar{\epsilon}_t, \tag{A.18}$$

where $\bar{\mathcal{Y}}_t(\vartheta_0) = \bar{y}_t - \bar{\omega}_t^\top\delta_0 - \mu(x_t, z_t; \gamma_0)$. We approximate $x_t\beta(z_t)$ through 2nd-order Taylor expansion around $z \in \mathcal{Z}$ to obtain $x_t\beta(z_t) = \mathcal{X}_t(z)^\top B(z) + R_t(z)$, where $\mathcal{X}_t(z) = [x_t, x_t(z_t - z)]^\top$, $B(z) = [\beta(z), \beta^{(1)}(z)]^\top$ is unknown constant to be estimated at $z_t = z$, and $R_t(z) = \frac{1}{2}\beta^{(2)}(z_t^*)x_t(z_t - z)^2$ is the reminder for $z_t^* \in [z_t, z]$. We define $T \times 1$ vectors $\bar{\mathcal{Y}}(\vartheta) = [\bar{\mathcal{Y}}_1(\vartheta), \ldots, \bar{\mathcal{Y}}_T(\vartheta)]^\top$, $\bar{\epsilon} = [\bar{\epsilon}_1, \ldots, \bar{\epsilon}_T]^\top$, and $R(z) = [R_1(z), \ldots, R_T(z)]^\top$. We further define $T \times 2$ matrix $\mathcal{X}(z) = [\mathcal{X}_1(z), \ldots, \mathcal{X}_T(z)]^\top$ whose t-th row is given by $\mathcal{X}_t(z)^\top$. We obtain estimator $\hat{B}(z) \equiv \hat{B}(z; \hat{\vartheta}, \hat{\beta}_0)$ as

$$\hat{B}(z) = \left[\mathcal{X}(z)^\top K(z)\mathcal{X}(z)\right]^{-1}\mathcal{X}(z)^\top K(z)(\bar{\mathcal{Y}}(\hat{\vartheta}) - \iota_T\hat{\beta}_0), \tag{A.19}$$

where $K(z) = diag\{k\left(\frac{z_t - z}{b}\right)\}_{t=1}^T$ is a $T \times T$ diagonal matrix of kernel functions. With D_b defined in *Proof of (a)* above, recall that $\hat{S}(z) = \frac{1}{Tb}D_b^{-1}\mathcal{X}(z)^\top K(z)\mathcal{X}(z)D_b^{-1} = S(z) + o_p(1)$ as shown in (A.16). Then,

$$\begin{aligned}
D_b(\hat{B}(z) - B(z)) = S(z)^{-1}\Bigg[&\frac{1}{Tb}D_b^{-1}\mathcal{X}(z)^\top K(z)R(z) \\
&\frac{1}{Tb}D_b^{-1}\mathcal{X}(z)^\top K(z)[\bar{Y}(\vartheta) - \bar{Y}(\vartheta_0) - (\hat{\beta}_0 - \beta_0)] \\
&\frac{1}{Tb}D_b^{-1}\mathcal{X}(z)^\top K(z)\hat{\epsilon}\Bigg](1 + o_p(1)) \\
\equiv S(z)^{-1}&\left[\mathcal{B}_{T1}(z) + \mathcal{B}_{T2}(z) + V_T(z)\right].
\end{aligned}$$

We show below that

(a) $[1, 0]^\top\mathcal{B}_{T1}(z) = b^2\mathcal{B}(z) + o_p(b^2)$

(b) $\sqrt{Tb}[1, 0]^\top\mathcal{B}_{T2}(z) = o_p(1)$

(c) $\sqrt{Tb}V_T(z) \xrightarrow{d} \mathcal{N}(0, \sigma_{\bar{\epsilon}}^2(z)S(z)\int k^2(v)dv)$.

With Slutsky's Theorem, results (a)-(c) gives claim in Theorem 2(b).

Proof of (a). We see that $\mathcal{B}_{T1}(z) = [\mathcal{B}_{T1,0}(z), \mathcal{B}_{T1,1}(z)]^\top$, where $\mathcal{B}_{T1,j} = \frac{1}{Tb}\sum_{t=1}^T k\left(\frac{z_t-z}{b}\right) x_t \left(\frac{z_t-z}{b}\right)^j \frac{x_t}{2}(z_t - z)^2\beta^{(2)}(z_t^*)$, for $j = 0, 1$. $\mathcal{B}_{T1,0}(z) = \frac{1}{2Tb}\sum_{t=1}^T k\left(\frac{z_t-z}{b}\right) x_t^2 \left(\frac{z_t-z}{b}\right)^2 b^2\beta^{(2)}(z_t^*) = \frac{b^2}{2}\beta^{(2)}(z_t^*)\hat{S}_2(z)$. Here, $E(\mathcal{B}_{T1,0}(z)) = \frac{b^2}{2}\beta(z)^{(2)}\mu_{k,2}E(x^2|z)f_z(z) + o_p(b^2)$ by B1, B2(1)-(2), and dominated convergence theorem. Also,

$$\sup_{z\in\mathcal{Z}} |\mathcal{B}_{T1,0}(z)| \leq \frac{b^2}{2}\beta^{(2)}(z)\sup_{z\in\mathcal{Z}}|\hat{S}_2(z)| = b^2[\frac{1}{2}E(x^2|z)f_z(z)\mu_{k,2}](1+o_p(1)) \equiv b^2\mathcal{B}(z) + o_p(b^2),$$

where the first equality applies Lemma 1 and $\mathcal{B}(z) < C$ is non-stochastic leading bias for $\hat{\beta}(z)$. Thus, $\mathcal{B}_{T1,0} = [1,0]^\top\mathcal{B}_{T1} = o_p(1)$ uniformally over $z \in \mathcal{Z}$ given B3(2).

Proof of (b). As in (a), we obtain $\mathcal{B}_{T2}(z) = [\mathcal{B}_{T2,0}(z), \mathcal{B}_{T2,1}(z)]^\top$, with

$$\mathcal{B}_{T2,j}(z) = \frac{1}{Tb}\sum_{t=1}^T k\left(\frac{z_t - z}{b}\right) x_t \left(\frac{z_t - z}{b}\right)^j \left[-\bar{\omega}_t^\top(\hat{\delta} - \delta_0) - (\mu(x_t, z_t; \hat{\gamma}) - \mu(x_t, z_t; \gamma_0))\right].$$

It is easy to see that $|\sqrt{Tb}\mathcal{B}_{T2,0}(z)| < \sqrt{Tb}\sup_{z\in\mathcal{Z}}|\mathcal{B}_{T2,0}(z)| = \sqrt{Tb}\sup_{z\in\mathcal{Z}}|\hat{S}_0(z)|O_p(1/\sqrt{nT}) = O_p(\sqrt{b/n})$ by Lemma 1 and Theorem 1(b). Thus, $\sqrt{Tb}\mathcal{B}_{T2,0}(z) = \sqrt{Tb}[1,0]^\top\mathcal{B}_{T2}(z) = o_p(1)$.

Proof of (c). $\sqrt{Tb}V_T(z) = \sqrt{Tb}\frac{1}{Tb}\sum_{t=1}^T k\left(\frac{z_t-z}{b}\right)\check{\mathcal{X}}_t(z)\bar{\epsilon}_t$, where $\check{\mathcal{X}}_t(z) = [x_t, x_t((z_t - z)/b)]^\top$. We see that $E(\sqrt{Tb}V_T(z)) = 0$, and

$$\begin{aligned}
V(\sqrt{Tb}V_T(z)) &= \frac{1}{Tb}\Bigg[TE\left(k^2\left(\frac{z_t - z}{b}\right)\bar{\epsilon}_t^2\check{\mathcal{X}}_t(z)\check{\mathcal{X}}_t(z)^\top\right) \\
&\quad + 2\sum\sum_{1\leq t\leq\tau\leq T} E\left(k\left(\frac{z_t - z}{b}\right)\check{\mathcal{X}}_t(z)\bar{\epsilon}_t k\left(\frac{z_\tau - z}{b}\right)\check{\mathcal{X}}_\tau(z)^\top\bar{\epsilon_\tau}\right)\Bigg] \\
&= \frac{1}{b}E\left(k^2\left(\frac{z_t - z}{b}\right)\bar{\epsilon}_t^2\check{\mathcal{X}}_t(z)\check{\mathcal{X}}_t(z)^\top\right) \\
&\quad + \frac{2}{Tb}\left[\sum_{t=1}^{T-1}(T-t)E\left(k\left(\frac{z_t - z}{b}\right)k\left(\frac{z_{t+1} - z}{b}\right)\bar{\epsilon}_t\bar{\epsilon}_{t+1}\check{\mathcal{X}}_t(z)\check{\mathcal{X}}_{t+1}(z)^\top\right)\right] \\
&\equiv V_{T1}(z) + V_{T2}(z).
\end{aligned}$$

First, $E(V_{T1}(z)) = \int k^2(v)E(\bar{\epsilon}_t{}^2|z_t)\check{\mathcal{X}}_t(z)\check{\mathcal{X}}_t(z)^\top f_z(z+vh)dv = \sigma_{\bar{\epsilon}}^2(z)S(z) + o_p(1)$ by A2(2)-(3), B1. Applying Lemma 1 further gives $\sup_{z\in\mathcal{Z}}|V_{T1}(z) - \sigma_{\bar{\epsilon}}^2(z)S(z)| = o_p(1)$ by B3(2). Thus, $V_{T1}(z) = \sigma_{\bar{\epsilon}}^2(z)S(z) + o_p(1)$ uniformly over $z \in \mathcal{Z}$.

Second, we can write the covariance term $V_{T2}(z)$ as

$$V_{T2}(z) = \frac{2}{Tb}\left[\sum_{t=1}^{T-1}(T-t)E\left(k\left(\frac{z_1 - z}{b}\right)k\left(\frac{z_{t+1} - z}{b}\right)\bar{\epsilon}_1\bar{\epsilon}_{t+1}\check{\mathcal{X}}_1(z)\check{\mathcal{X}}_{t+1}(z)^\top\right)\right]$$

given the stationary property in A1(1). Let us partition the time dimension $T-1$ into two parts $[1,\dots,d_T]$ and $[d_T+1,\dots,T-1]$, where as $T\to\infty$, we impose conditions on the growth rate of $d_T\to\infty$ as

$$d_T \to \infty, \quad d_T b \to 0, \quad d_T^2 b = O(1). \tag{A.20}$$

Thus,

$$\begin{aligned} V_{T2}(z) &= \sum_{t=1}^{d_T} \frac{2}{Tb}\left[\sum_{t=1}^{T-1}(T-t)E\left(k\left(\frac{z_1-z}{b}\right)k\left(\frac{z_{t+1}-z}{b}\right)\bar{\epsilon}_1\bar{\epsilon}_{t+1}\check{\mathcal{X}}_1(z)\check{\mathcal{X}}_{t+1}(z)^\top\right)\right] \\ &\quad + \sum_{t=d_T+1}^{T-1}\frac{2}{Tb}\left[(T-t)E\left(k\left(\frac{z_1-z}{b}\right)k\left(\frac{z_{t+1}-z}{b}\right)\bar{\epsilon}_1\bar{\epsilon}_{t+1}\check{\mathcal{X}}_1(z)\check{\mathcal{X}}_{t+1}(z)^\top\right)\right] \\ &\equiv V_{T2,1}(z)+V_{T2,2}(z). \end{aligned}$$

First, $V_{T2,1}(z) = b\sum_{t=1}^{d_T}\frac{2}{Tb^2}(T-t)E\left[k\left(\frac{z_1-z}{b}\right)k\left(\frac{z_{t+1}-z}{b}\right)\check{\mathcal{X}}_1(z)\check{\mathcal{X}}_{t+1}(z)^\top E(\bar{\epsilon}_1\bar{\epsilon}_{t+1}|x_1,x_{t+1},z_1,z_{t+1})\right] = O(d_T b) = o_p(1)$ by B2(4) and (A.20).

Second, $V_{T2,2}(z) = \sum_{t=d_T+1}^{T-1}\frac{2}{Tb}(T-t)E\left(k\left(\frac{z_1-z}{b}\right)k\left(\frac{z_{t+1}-z}{b}\right)\bar{\epsilon}_1\bar{\epsilon}_{t+1}\check{\mathcal{X}}_1(z)\check{\mathcal{X}}_{t+1}(z)^\top\right)$. Define a short-hand notation $k_t^s(z) \equiv K\left(\frac{z_t-z}{b}\right)\left(\frac{z_t-z}{b}\right)^s$, for $s=0,1$. Given some $\delta>0$, under A1(2) and Davydov's inequality in Lemma 2, a typical element being summed over in $V_{T2,2}(z)$ can be bounded by

$$E\left[k_1^s(z)x_1\bar{\epsilon}_1k_{t+1}^s(z)x_{t+1}\bar{\epsilon}_{t+1}\right] \le C\alpha(t+1)^{1-\frac{2}{2+\delta}}\left[E|k_1^s(z)x_1\bar{\epsilon}_1|^{2+\delta}\right]^{\frac{1}{2+\delta}}\left[E|k_{t+1}^s(z)x_{t+1}\bar{\epsilon}_{t+1}\right]^{\frac{1}{2+\delta}}.$$

Since $\frac{1}{b}E|k_t^s(z)x_t\bar{\epsilon}_t|^{2+\delta} = \frac{1}{b}E\left[|k_t^s(z)|^{2+\delta}E(|x_t|^{2+\delta}|z_t)E(|\bar{\epsilon}_t|^{2+\delta}|z_t)\right] = O(1)$ by A2(2) B1, B2(1)-(2), for all $t=1,\dots,T-1$, then given that $\alpha(t+1)<\alpha(t)$ under A1(2), we have

$$\begin{aligned} E\left[k_1^s(z)x_1\bar{\epsilon}_1k_{t+1}^s(z)x_{t+1}\bar{\epsilon}_{t+1}\right] &< Ch^{\frac{2}{2+\delta}}\alpha(t)^{\frac{2}{2+\delta}}\left[\frac{1}{b}E|k_1^s(z)x_1\bar{\epsilon}_1|^{2+\delta}\right]^{\frac{1}{2+\delta}}\left[\frac{1}{b}E|k_{t+1}^s(z)x_{t+1}\bar{\epsilon}_{t+1}\right]^{\frac{1}{2+\delta}} \\ &= O(\alpha(t)^{\frac{2}{2+\delta}}b^{\frac{2}{2+\delta}}). \end{aligned}$$

Therefore, we obtain

$$V_{T2,2}(z) < C\sum_{t>d_T}\alpha(t)^{\frac{\delta}{2+\delta}}b^{-1}b^{\frac{2}{2+\delta}} \le Ch^{-\frac{\delta}{2+\delta}}\sum_{t>d_T}\left(\frac{t}{d_T}\right)^\delta\alpha(t)^{\frac{\delta}{2+\delta}} = Ch^{-\frac{\delta}{2+\delta}}d_T^{-\delta}\sum_{t>d_T}t^\delta\alpha(t)^{\frac{\delta}{2+\delta}},$$

where the second inequality holds because $(t/d_T)^\delta>1$ given that $t>d_T$. Since $\alpha(t)<Ct^{-B}$ under A1(2), we further obtain $V_{T2,2}(z) < b^{-\frac{\delta}{2+\delta}}d_T^{-\delta}\sum_{t>d_T}t^{\delta(1-\frac{B}{2+\delta})}$. It can be shown that by choosing $B>(2\delta+1)(2+\delta)/\delta$, one can ensure $\delta(1-\frac{B}{2+\delta})<-(1+\delta)$, which implies $\sum_{t>d_T}t^{-(1+\delta)}<C$ by A1(2). Thus, $V_{T2,2}(z) = O(b^{-\frac{\delta}{2+\delta}}d_T^{-\delta}) = o(1)$ by observing that$b^{-\frac{\delta}{2+\delta}}d_T^{-\delta} = \left[(hd_T^2)^{\frac{1}{2+\delta}}d_T^{\frac{\delta}{2+\delta}}\right]^{-\delta} = o_p(1)$ under (A.20). In summary, we have $V_{T2}(z) = o(1)$.

We now focus on the asymptotic distribution of $V_T(z)$, demonstrating that $V_{T1}(z)$ determines the leading variance of $\hat{\beta}(z)$. For any non-zero vector λ of size 2×1,

$$\sqrt{Tb}\lambda^\top V_T(z) = \frac{1}{\sqrt{T}}\sum_{t=1}^{T}\frac{1}{\sqrt{b}}k\left(\frac{z_t - z}{b}\right)\lambda^\top \check{\mathcal{X}}_t(z)\bar{\epsilon}_t \equiv \frac{1}{\sqrt{T}}\sum_{t=1}^{T} Q_t.$$

Recall that $E(Q_t) = 0$ and $V(\frac{1}{\sqrt{T}}\sum_{t=1}^{T} Q_t) = \lambda^\top \sigma^2_{\bar{\epsilon}}(z)S(z)\int k^2(v)dv\lambda + o_p(1)$ in result (b) above. To apply central limit theorem on $\{Q_t\}_{t=1}^T$, we apply small-block and large-block technique to handle data dependence as $T\to\infty$. Let us partition time index $\{1,2,\ldots,T\}$ into a total $2q_T+1$ subsets, with large block l_T and small block s_T such that $l_T + s_T < T$ and $s_T < l_T$. Here, $q_T = \left[\frac{T}{l_T+s_T}\right]$, where $[v]$ the integer collector as in (A.4). For $0\le j\le q_T-1$, we define sums of observations within large-block as η_j, within small-block as ξ_j, and within remaining block as ζ_j such that

$$\eta_j = \sum_{t=j(l_T+s_T)+1}^{j(l_T+s_T)+l_T} Q_t, \qquad \xi_j = \sum_{t=j(l_T+s_T)+l_T+1}^{(j+1)(l_T+s_T)} Q_t, \qquad \zeta_{q_T} = \sum_{t=q_T(l_T+s_T)+1}^{T} Q_t.$$

Given that all T observations are divided into $2q_T+1$ intervals, we have

$$\frac{1}{\sqrt{T}}\sum_{t=1}^{T} Q_t = \frac{1}{\sqrt{T}}\left[\sum_{j=0}^{q_T-1}\eta_j + \sum_{j=0}^{q_T-1}\xi_j + \zeta_{q_T}\right] \equiv \frac{1}{\sqrt{T}}Q_{T_1} + \frac{1}{\sqrt{T}}Q_{T_2} + \frac{1}{\sqrt{T}}Q_{T_3}.$$

As $T\to\infty$, we show that

(c.1) $\frac{1}{T}E(Q^2_{T_2}) = \frac{1}{T}E(Q^2_{T_3}) = o_p(1)$

(c.2) $|E\left[\exp(it\sum_{j=0}^{q_T-1}\eta_j)\right] - \prod_{j=0}^{q_T-1}E\left[\exp(it\eta_j)\right]| = o_p(1)$

(c.3) $\frac{1}{T}\sum_{j=0}^{q_T-1}E(\eta_j^2) \xrightarrow{p} \lambda^\top\sigma^2_{\bar{\epsilon}}(z)S(z)\int k^2(v)dv\lambda$, and $\frac{1}{\sqrt{T}}Q_{T_1} \xrightarrow{d} \mathcal{N}\left(0, \lambda^\top\sigma^2_{\bar{\epsilon}}(z)S(z)\int k^2(v)dv\lambda\right)$.

Then by the Lyapounov central limit theorem and Cramer-Rao derive, results (i)-(iii) gives

$$\frac{1}{\sqrt{T}}\sum_{t=1}^{T} Q_T \xrightarrow{d} \mathcal{N}(0, \sigma^2_{\bar{\epsilon}}(z)S(z)\int k^2(v)dv),$$

which completes the proof of (c). Using Slutsky's Theorem, we obtain

$$\sqrt{Tb}S(z)^{-1}V_T(z) \xrightarrow{d} \mathcal{N}\left(0, S(z)^{-1}\sigma^2_{\bar{\epsilon}}(z)\int k^2(v)dv\right),$$

which implies that

$$\sqrt{Tb}(\hat{\beta}(z) - \beta(z) - b^2\mathcal{B}(z)) \xrightarrow{d} \mathcal{N}\left(0, \Omega_\beta\right),$$

with $\Omega_\beta = \int k^2(v)dv\sigma^2_{\bar{\epsilon}}(z)/E(x^2|z)f_z(z)$. Together with results in (a) and (b), the results in Theorem 2(b) follow. Note that (i) implies that $\frac{1}{\sqrt{T}}Q_{T2}$ and $\frac{1}{\sqrt{T}}Q_{T3}$ are asymptotically of smaller order and thus negligible. (ii) shows that $\{\eta_j\}_{j=0}^{q_T-1}$ are asymptotically independent. (iii) shows that the Lyapounov central limit theorem can be applied to obtain normality of $\hat{\beta}$, provided that certain sufficient condition is satisfied under our assumptions.

Proof of (i). Let l_T and s_T to grow at certain orders $l_T = \left[T^{\frac{1}{B}}\right]$ and $s_T = \left[T^{\frac{1}{B+1}}\right]$, where $B > 2$ and $B > \frac{(2+\delta)(1+\delta)}{\delta}$ as discussed in $V_{T2,2}(z)$ above. Thus, $\frac{s_T}{l_T} = O(T^{-\frac{1}{B(B+1)}}) = o(1)$. Also, $\frac{l_T}{T} = O(T^{\frac{1-B}{B}}) = o(1)$, and $\frac{s_T}{T} = \frac{s_T}{l_T} \cdot \frac{l_T}{T} = o(1)$. The variance $\frac{1}{T}EQ_{T2}^2 = \frac{1}{T}E\left[\sum_{j=0}^{q_T-1}\xi_j\right]^2 = \frac{1}{T}\sum_{j=0}^{q_T-1}V(\xi_j) + \frac{2}{T}\sum\sum_{0\le k<j\le q_T-1} Cov(\xi_k,\xi_j) \equiv J_1 + J_2$. First, by A1(2) for stationary property,

$$J_1 = \frac{q_T}{T}V(\xi_j) = \frac{q_T}{T}V\left(\sum_{t=j(l_T+s_T)+1}^{(j+1)(l_T+s_T)+l_T} Q_t\right) = \frac{q_T s_T}{T}E\left(\frac{1}{b}k\left(\frac{z_t-z}{b}\right)\lambda^\top \check{\mathcal{X}}_t(z)\check{\mathcal{X}}(z)^\top \bar{\epsilon}_t^2\lambda\right) = O\left(\frac{q_T s_T}{T}\right).$$

Since $q_T s_T/T = O(s_T/(s_T+l_T)) = O(1/(l_T/s_T+1)) = o(1)$, $J_1 = o(1)$. Second, $J_2 = \frac{2}{T}\sum\sum_{0\le k<j\le q_T-1} Cov(\xi_k,\xi_j)$. To simplify notations, define $l_j^* = j(s_T + l_T)$, and rewrite J_2 as

$$J_2 = \frac{2}{T}\sum\sum_{0\le k<j\le q_T-1} Cov\left(\sum_{t=l_k^*+l_T+1}^{l_k^*+l_T+s_T} Q_T, \sum_{t=l_j^*+l_T+1}^{l_j^*+l_T+s_T} Q_T\right) = \frac{2}{T}\sum\sum_{0\le k<j\le q_T-1}\left[\sum_{j_1=1}^{s_T}\sum_{j_2=1}^{s_T} Cov(Q_{l_k^*+l_T+j_1}, Q_{l_j^*+l_T+j_2})\right].$$

Here, notice that for any $j > k$, the indexes $|l_j^* + l_T + j_2| - |l_k^* + l_T + j_1| = |(j-k)(l_T + s_T) + (j_2 - j_1)| > |(k-j)(l_T+s_T)| > |l_T + s_T| > l_T + 1 > l_T$, where the first greater than sign holds because $\min(j_2 - j_1) = 0$, the second sign holds because $j > k$, and the last sign holds because $\min(s_T) = 1$. Thus, $|j-k| > l_T$. Given this observation, we further simplify J_2 as

$$\begin{aligned}
|J_2| &< \frac{2}{T}\sum_{k=0}^{q_T-1}\sum_{j>k}\sum_{j_1=1}^{s_T}\sum_{j_2=1}^{s_T}|Cov(Q_{l_k^*+l_T+j_1}, Q_{l_j^*+l_T+j_2})| \\
&< \frac{2}{T}\sum_{k=1}^{T-l_T}\sum_{j=k+l_T}^{T}|Cov(Q_k,Q_j)| = \frac{2(T-l_T)}{T}\sum_{j=1+l_T}^{T}|Cov(Q_1,Q_j)|, \text{ (by A1(2))} \\
&< 2\sum_{t=1}^{T}E\left(\frac{1}{b}k\left(\frac{z_1-z}{b}\right)\lambda^\top\check{\mathcal{X}}_1(z)\check{\mathcal{X}}_t(z)^\top\lambda k\left(\frac{z_t-z}{b}\right)\bar{\epsilon}_1\bar{\epsilon}_t\right) \\
&= 2b\left[\sum_{t=1}^{d_T} + \sum_{t=d_T+1}^{T}\right]E\left(\frac{1}{b^2}k\left(\frac{z_1-z}{b}\right)\lambda^\top\check{\mathcal{X}}_1(z)\check{\mathcal{X}}_t(z)^\top\lambda k\left(\frac{z_t-z}{b}\right)\bar{\epsilon}_1\bar{\epsilon}_t\right) \\
&= O(bd_T) + O\left(\left[(bd_T^2)^{\frac{1}{2+\delta}}d_T^{\frac{\delta}{2+\delta}}\right]^{-\delta}\right) \\
&= o(1)
\end{aligned}$$

under condition (A.20) with the same argument in $V_{T2,2}(z)$. So $J_2 = o(1)$ and thus $\frac{1}{\sqrt{T}}Q_{T2} = o(1)$.

We now focus on Q_{T3}. Using similar argument,

$$\frac{1}{T}E(Q_{T3}^2) = \frac{1}{T}V\left(\sum_{t=1}^{T-q_T(l_T+s_T)} Q_T\right) = \frac{1}{T}\sum_{t=1}^{T-q_T(l_T+s_T)} V(Q_T) + \frac{2}{T}\sum\sum_{1\le t<k\le T-q_T(s_T+l_T)} Cov(Q_t, Q_s) \equiv J_3 + J_4.$$

First, $J_3 = \frac{1}{T}(T - q_T(l_T + s_T))E(Q_t^2) = \frac{1}{T}(T - \left[\frac{T}{l_T+s_T}\right](l_T + s_T))E(Q_t^2) = o(T^{-1})E(Q_t^2)$. Since $E(Q_t^2) = \lambda^\top \sigma_\epsilon^2(z)S(z)\lambda \int k^2(v)dv + o_p(1) = O(1)$, $J_3 = o(1)$. Second, $J_4 = \frac{2}{T}\sum_{t=1}^{T-q_T(s_T+l_T)}(T - q_T(s_T+l_T) - t)Cov(Q_1, Q_{t+1}) < \frac{2}{T}(T - q_t(s_T + l_T))\sum_{t=1}^{T-1} Cov(Q_1, Q_{t+1}) = o(T^{-1})(\sum_{t=1}^{d_T} + \sum_{t=d_T+1}^{T-1})Cov(Q_1, Q_{t+1}) = o(1)$ as in J_2, so $J_4 = o(1)$. This implies $\frac{1}{\sqrt{T}}Q_{T3} = o(1)$, and result (c.1) follows.

Proof of (c.2). Following the notation in Lemma 2, recall that $\mathcal{F}_i^j$ is the σ-algebra generated by random variables $\{x_t, z_t\}_{t=i}^j$, so η_j is $\mathcal{F}_{j(s_T+l_T)+1}^{j(s_T+l_T)+l_T}$ measurable. Note that the distance between any two different η_j and $\eta_{j'}$ with $j < j'$, say $j = 0$ and $j' = 1$, is separated by a small block of size $s_T + 1$. Define $V_j = \exp(it\eta_j)$ with $i = \sqrt{-1}$ be the imaginary number, we apply Lemma 2 to obtain

$$\left|E[\exp(it\sum_{j=0}^{q_T-1}\eta_j)] - \prod_{j=0}^{q_T-1} E\left[\exp(it\eta_j)\right]\right| \le 16(q_T - 1)\alpha(s_T + 1) = 16q_T\alpha(s_T + 1) - 16\alpha(s_T + 1).$$

By A1(2), $q_T\alpha(s_T) \le \left[\frac{T}{s_T+l_T}\right] s_T^{-B} = \left[\frac{T}{s_T+l_T}\right]\left[T^{-\frac{B}{B+1}}\right] = O\left(\frac{1}{T^{\frac{1}{B(B+1)}+1}}\right) = o(1)$ and $B > 2$. Thus, as $T \to \infty$, $16q_T\alpha(s_T+1) < 16\left[\frac{T}{s_T+l_T}\right](S_T+1)^{-B} < 16\left[\frac{T}{s_T+l_T}\right](S_T)^{-B} = o_p(1)$, and $16\alpha(s_T+1) < 16\alpha(s_T) = o(1)$, which comes to (c.2).

Proof of (c.3). Note that (c.1) implies $\sqrt{Tb}\lambda^\top V_T(z) = \frac{1}{\sqrt{T}}\sum_{j=0}^{q_T-1}\eta_j + o(1)$. By (c.2), $\{\eta_j\}_{j=0}^{q_T-1}$ are symptomatically independent, allowing us to apply the Lyapounov central limit theorem. Notice that

$$\frac{1}{T}V\left(\sum_{j=0}^{q_T-1}\eta_j\right) = \frac{1}{T}\sum_{j=0}^{q_T-1}V(\eta_j) + \frac{2}{T}\sum\sum_{0\le k<j\le q_T-1} Cov(\eta_k, \eta_j) \equiv J_5 + J_6.$$

First, $J_5 = O\left(\frac{l_T}{l_T+s_T}\right)V(Q_t) = \left[\frac{1}{1+\frac{s_T}{l_T}}\right]E(Q_tQ_t^\top) = \lambda^\top\sigma_\epsilon^2(z)S(z)\int k^2(v)dv\lambda + o_p(1)$ because $s_T/l_T = o(1)$. Second, $J_6 = \frac{2}{T}\sum\sum_{0\le k<j\le q_T-1} Cov(\eta_k, \eta_j) = \frac{2}{T}\sum\sum_{0\le k<j\le q_T-1}\left[\sum_{j_1=1}^{l_T}\sum_{j_2=1}^{l_T} Cov(Q_{l_k^*+j_1}, Q_{l_j^*+j_2})\right]$, with $l_j^* = j(s_T + l_T)$. As discussed above, the distance $|l_k^* + j_1| - |l_j^* + j_1| > l_T$, which implies $j - k > l_T$. Thus, $|J_6| < \frac{2}{T}\sum_{k=1}^{T-l_T}\sum_{j=k+l_T}^{T}|Cov(Q_k, Q_j)| < \frac{2}{T}(T - l_T)\sum_{j=1+l_T}^{T}|Cov(Q_1, Q_j)| < 2\sum_{t=1}^{T}|Cov(Q_1, Q_t)| = o(1)$ as shown in J_2. Now, define $Z_{jT} = \sum_{t=j(s_T+l_T)+1}^{j(s_T+l_T)+l_T}\frac{1}{\sqrt{T}}Q_t$. By Lyapounov central limit theorem (Li and Racine, 2007),

$$\sum_{j=0}^{q_T} Z_{jT} \xrightarrow{d} \mathcal{N}\left(0, \lambda^\top\int k^2(v)dv\sigma_\epsilon^2(z)S(z)\lambda\right), \tag{A.21}$$

provided that $\lim_{T\to\infty} \frac{1}{T}\sum_{j=0}^{q_T-1} E\left[Z_{jT}^2 1(|Z_{jT}| \geq \epsilon\Sigma(z)\sqrt{T})\right] = 0$, where $\Sigma(z) \equiv \lambda^\top \int k^2(v)dv\sigma_\epsilon^2(z)S(z)\lambda$. This condition is established below.

Notice first that, for any $\delta > 0$, $E|Z_{jT}|^{2+\delta} \geq E\left[|Z_{jT}|^{2+\delta} 1(|Z_{jT}| > \epsilon\Sigma(z)\sqrt{T})\right]$. Since $|Z_{jT}|^\delta > [\epsilon\Sigma(z)]^\delta T^{\frac{\delta}{2}}$, $E|Z_{jT}|^{2+\delta} \geq CT^{\frac{\delta}{2}} E\left[|Z_{jT}|^2 1(|Z_{jT}| > \epsilon\Sigma(z)\sqrt{T})\right]$, or

$$
\begin{aligned}
E\left[|Z_{jT}|^2 1(|Z_{jT}| > \epsilon\Sigma(z)\sqrt{T})\right] &\leq T^{-\frac{\delta}{2}} E\left|\sum_{t=j(l_T+s_T)+1}^{j(l_T+s_T)+l_T} Q_t\right|^{2+\delta} \\
&\leq C l_T^{1+\frac{\delta}{2}} \left[E|Q_t|^r\right]^{\frac{2+\delta}{r}},
\end{aligned}
$$

where the second inequality is obtained using Lemma 3 in Cai and Li (2008) by letting $r = 2(1+\delta)$, $p = 2+\delta$, $n = l_T$. Recall that $Q_t \equiv \frac{1}{\sqrt{b}}\lambda k\left(\frac{z_t-z}{b}\right)^\top \chi_t(z)\bar{\epsilon}_t$. Since

$$
E|Q_t|^r = b^{-\frac{r}{2}} E\left|\lambda k\left(\frac{z_t - z}{b}\right)^\top \chi_t(z)\bar{\epsilon}_t\right|^r = O_p(b^{1-\frac{r}{2}}) = O_p(b^{-\delta}),
$$

we have $E|Z_{jT}|^{2+\delta} = O(b^{-\frac{\delta(2+\delta)}{2(1+\delta)}} l_T^{1+\frac{\delta}{2}})$. As a result,

$$
\begin{aligned}
\frac{1}{T}\sum_{j=0}^{q_T-1} E\left[Z_{jT}^2 1(|Z_{jT}| \geq \epsilon\Sigma(z)\sqrt{T})\right] &\leq \frac{C}{T}\sum_{j=0}^{q_T-1} T^{-\frac{\delta}{2}} E|Z_{jT}|^{2+\delta} \\
&= C\left[\frac{q_T l_T}{T}\right] T^{-\frac{\delta}{2}} \left[T^{\frac{1}{B}}\right]^{\frac{\delta}{2}} b^{-\frac{\delta(2+\delta)}{2(1+\delta)}} \\
&\leq \left[T^{\frac{B-1}{B}} b^{\frac{2+\delta}{1+\delta}}\right]^{-\frac{\delta}{2}} \\
&= o_p(1),
\end{aligned}
$$

where the last equality holds by B3(4), and the result in (A.21) follows. □

Appendix 2: Simulation

This section presents simulation studies to evaluate the finite sample performance of the proposed pseudo-maximum likelihood estimator (pseudo-MLE) for ϑ_0 in the first step and the profile kernel estimators $(\beta_0, \beta(\cdot))$ in the second step. In Appendix 2.1, we conduct the simulation under a correct distributional assumption for the skewed normal distribution (SN) of the error. In Appendix 2.2, we repeat the exercise with incorrect distributional assumptions and provide detailed discussions on the results.

Appendix 2.1: Simulation under Correct Distributional Assumptions

We begin with a simulation study based on the correct assumption that the error in our skewed panel model follows a SN. First, we illustrate the numerical properties of the estimators with different choices of T and a fixed n, assuming $n/T \to 0$. Second, we investigate whether the estimators converge at their expected parametric or nonparametric rates, as outlined in Theorems 1 and 2. We consider the following data generating process (DGP):

$$y_{i,t+1} = \beta_0 + x_t\beta(z_t) + y_{i,t}\delta_{0y} + w_{i,t}\delta_{0w} + \alpha_{0i} + e_{i,t}, \tag{A.22}$$

where for $i = 1, \ldots, n$ and $t = 1, \ldots, T$, we specify $\omega_{i,t} = [y_{i,t}, w_{i,t}]^\top$ to feature a dynamic structure through a one-period lag variable $y_{i,t}$, along with a univariate $w_{i,t}$. We generate a SN error

$$e_{i,t} \sim SN(0, \sigma(x_t, z_t; \gamma_{0\sigma}), \lambda(x_t, z_t; \gamma_{0\lambda})), \tag{A.23}$$

where $\sigma(x_t, z_t; \gamma_{0\sigma}) = \exp(x_t z_t \gamma_{0\sigma})$ and $\lambda(x_t, z_t; \gamma_{0\lambda}) = x_t z_t \gamma_{0\lambda}$. To introduce time dependence, we generate $z_t = z_t^0 + \zeta_t^z$, where $z_t^0 \sim U(0,1)$ and $\zeta_t^z = 0.25\zeta_{t-1}^z + \xi_t^z$ following an AR(1) process with $\xi_t^z \sim \mathcal{N}(0, 0.25^2)$. To be consistent with the range of $z_t \in [0,1]$ in our empirical dataset, we re-scale z_t into the range of $[0,1]$. Similarly, we obtain $w_{i,t} = w_{i,t}^0 + \xi_{i,t}^w$, where $w_{i,t}^0 \sim U(1,4)$ and $\zeta_{i,t}^w = 0.5\zeta_{i,t-1}^w + \xi_{i,t}^w$ with $\xi_{i,t}^w \sim \mathcal{N}(0, 0.5^2)$. We simulate $x_t = 0.5z_t + \zeta_t^x$ to allow fairly strong correlation between x_t and its effect modifier z_t, where $\zeta_t^x = 0.75\zeta_{t-1}^x + \xi_t^x$ and $\xi_t^x \sim \mathcal{N}(0, 0.65^2)$. We follow the convention in the literature to set $y_{i,1} = \beta_0 + \alpha_{0i} + e_{i,0}$ as the initial condition. Here, $\alpha_{0i} = \frac{1}{T}\sum_{t=1}^T c_0(w_{i,t} + y_{i,t}) + \xi_i$ with $\xi_i \sim \mathcal{N}(0, 0.5)$ and $c_0 \neq 0$ represents a fixed effect model. We set $c_0 = 1$ throughout the experiment, and impose the identification condition for fixed effects by setting $\alpha_{01} = -\sum_{i=2}^n \alpha_{0i}$.

We consider two different $DGPs$ with different specifications for parameters $\vartheta_0 = (\gamma_0^\top, \theta_0^\top, \alpha_{0,-1}^\top)$, intercept β_0, and unknown function $\beta(z)$, where in this case, $\gamma_0 = (\gamma_{0\sigma}, \gamma_{0\lambda})$, $\theta_0 = (\theta_{0y}, \delta_{0w})$, and $\alpha_{0,-1} = (\alpha_{02}, \ldots, \alpha_{0n})$. In DGP_1, we set $\gamma_0 = (0.5, 2)$, $\theta_0 = (0.5, 1.5)$, $\beta_0 = 2$, and $\beta(z) = 1 - z$. In DGP_2, we set

$\gamma_0 = (0.4, -2)$, $\theta_0 = (0.25, -1.5)$, $\beta_0 = 1.5$, and $\beta(z) = 0.5 + z - 2z^2$. It is clear that the performance of function estimator $\hat{\beta}(\cdot)$ is evaluated for a linear structure in DGP_1 while a nonlinear (quadratic) structure in DGP_2.

In our first step pseudo-MLE, we implement quasi-Newton method for maximization algorithm, from which we set the starting values as one unit above the true values ϑ_0 to reduce computational burden. In the second step kernel estimation, we use a second-order Gaussian kernel function with a rule-of-thumb bandwidth $b = \hat{\sigma}_z T^{-1/5}$ satisfying B3(2) (see Appendix 1), where $\hat{\sigma}_z$ is the standard deviation of $\{z_t\}_{t=1}^T$. Throughout the experiment, we fix $n = 10$ or $n = 30$, select $T = (50, 100, 200)$, and conduct 500 repetitions. We evaluate the performance of all parameter estimators $\hat{\vartheta}$ and $\hat{\beta}_0$ by root mean squared error (RMSE), absolute bias (BIAS), and standard deviation (SD). We evaluate the performance of kernel estimator $\hat{\beta}(\cdot)$ via root average MSE (RAMSE), average BIAS (ABIAS), and average SD (ASD).

Table A.1 summarizes the simulation results for our two-step estimator in DGP_1 (upper panel) and DGP_2 (lower panel). In each DGP, the results for the estimators are summarized in the first step (Step 1) and in the second step (Step 2). Conditioning on each sample, the parametric estimator $\hat{\theta} = (\hat{\gamma}_\sigma, \hat{\gamma}_\lambda, \hat{\delta}_y, \hat{\delta}_w)$ uniformly outperforms the nonparametric estimator $\hat{\beta}(\cdot)$ with different DGPs. This is expected, because $\hat{\theta}$ converges at a rate of $1/\sqrt{nT}$ (see Theorem 1(b)) faster than $\hat{\beta}(\cdot)$ at $1/\sqrt{Tb}$ (see Theorem 2(b)). $\hat{\theta}$ is also superior to $\hat{\beta}_0$ because the latter involves kernel smoothing on an unknown coefficient function, which introduces an additional approximation error that worsens β_0 estimation. Due to the large dimension of fixed effects $\alpha_{0,-1}$, the statistic measures associated with $\hat{\alpha}_{-1}$ are averaged over each element in $\hat{\alpha}_{-1}$ ($\hat{\alpha}_{-1}$: Ave). Since $\hat{\alpha}_i$ is only $\sqrt{T}$-consistent, it performs less well compared to other parameter estimates. We observe that a larger n improves all estimators for any fixed T. The main reason is that increasing n effectively reduces the approximation error for $\frac{1}{n}\sum_{i=1}^n e_{i,t}$ using its probability limit $\mu(x_t, z_t; \gamma_0)$ for the pseudo-MLE estimator $\hat{\vartheta}$. This improvement is in turn transmitted to the second-step estimators because of their tight dependence. In particular, increasing n barely affects the performance of $\hat{\alpha}_{-1}$ because the fixed effect is only $\sqrt{T}$ consistent.

However, both parameter and function estimators are consistent in that all measures decay to zero as T doubles for each n, regardless of which DGP is considered. More importantly, all estimators converge at their expected rates based on Theorems 1 and 2. To clearly demonstrate the convergence rates, we re-estimate our model in (A.22) by fixing $n = 30$, raising $T = (50, 100, 150, 200, 250)$, and computing RMSE and RAMSE given each sample size. Figure 1(a) plots the RMSE of the parameter in the skew function $\hat{\gamma}_\lambda$ (dashed line with o in DGP_1 and + in DGP_2), and the coefficient of the lag variable $\hat{\delta}_y$ (dot line with $\triangle$ in DGP_1 and

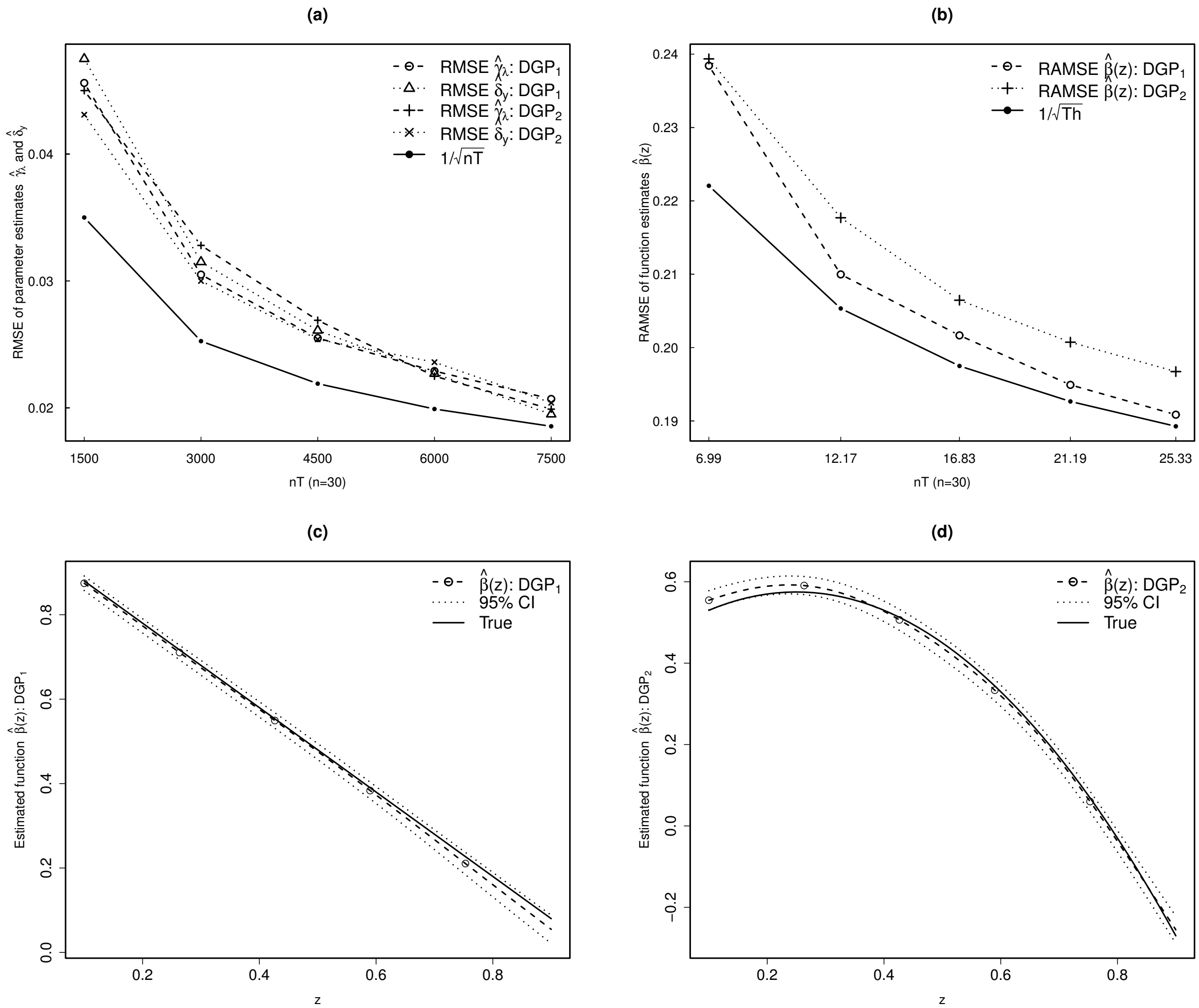
Table A.1: Simulation Results for Two-Step Estimator under Correct Distribution

			$n=10$			$n=30$		
DGP_1	**Step 1**		$T=50$	$T=100$	T=200	$T=50$	$T=100$	T=200
	$\hat{\gamma}_\lambda$	RMSE	0.0392	0.0267	0.0219	0.0244	0.0159	0.0118
		BIAS	0.0307	0.0210	0.0183	0.0193	0.0129	0.0096
		SD	0.0243	0.0165	0.0122	0.0150	0.0094	0.0070
	$\hat{\gamma}_\sigma$	RMSE	0.0380	0.0221	0.0164	0.0230	0.0155	0.0104
		BIAS	0.0293	0.0173	0.0132	0.0182	0.0123	0.0083
		SD	0.0241	0.0138	0.0098	0.0142	0.0094	0.0064
	$\hat{\delta}_y$	RMSE	0.0238	0.0155	0.0103	0.0173	0.0105	0.0076
		BIAS	0.0188	0.0123	0.0081	0.0140	0.0082	0.0060
		SD	0.0145	0.0094	0.0063	0.0103	0.0066	0.0046
	$\hat{\delta}_w$	RMSE	0.0508	0.0346	0.0241	0.0354	0.0236	0.0172
		BIAS	0.0404	0.0282	0.0193	0.0285	0.0185	0.0136
		SD	0.0308	0.0201	0.0146	0.0210	0.0146	0.0106
	$\hat{\alpha}_{-1}$: Ave	RMSE	0.1021	0.0694	0.0500	0.0940	0.0692	0.0487
		BIAS	0.0798	0.0544	0.0393	0.0743	0.0541	0.0392
		SD	0.0638	0.0432	0.0309	0.0577	0.0432	0.0290
	Step 2							
	$\hat{\beta}_0$	RMSE	0.0939	0.0706	0.0509	0.0632	0.0458	0.0336
		BIAS	0.0730	0.0554	0.0400	0.0501	0.0356	0.0261
		SD	0.0894	0.0638	0.0435	0.0612	0.0433	0.0304
	$\hat{\beta}(z)$	RAMSE	0.1565	0.1036	0.0697	0.1025	0.0714	0.0544
		ABIAS	0.1215	0.0828	0.0558	0.0811	0.0570	0.0440
		ASD	0.0988	0.0624	0.0418	0.0628	0.0431	0.0320
DGP_2	**Step 1**							
	$\hat{\gamma}_\lambda$	RMSE	0.0431	0.0283	0.0230	0.0198	0.0161	0.0124
		BIAS	0.0330	0.0226	0.0196	0.0151	0.0125	0.0103
		SD	0.0279	0.0172	0.0121	0.0129	0.0101	0.0069
	$\hat{\gamma}_\sigma$	RMSE	0.0334	0.0228	0.0152	0.0233	0.0146	0.0124
		BIAS	0.0257	0.0170	0.0119	0.0187	0.0110	0.0099
		SD	0.0214	0.0152	0.0095	0.0140	0.0096	0.0075
	$\hat{\delta}_y$	RMSE	0.0229	0.0188	0.0109	0.0182	0.0109	0.0070
		BIAS	0.0189	0.0149	0.0091	0.0141	0.0090	0.0056
		SD	0.0130	0.0116	0.0060	0.0116	0.0061	0.0042
	$\hat{\delta}_w$	RMSE	0.0438	0.0327	0.0233	0.0355	0.0214	0.0164
		BIAS	0.0357	0.0256	0.0186	0.0267	0.0168	0.0122
		SD	0.0255	0.0203	0.0141	0.0235	0.0133	0.0110
	$\hat{\alpha}_{-1}$: Ave	RMSE	0.0855	0.0580	0.0410	0.0829	0.0535	0.0366
		BIAS	0.0675	0.0465	0.0339	0.0655	0.0434	0.0282
		SD	0.0528	0.0349	0.0233	0.0512	0.0314	0.0235
	Step 2							
	$\hat{\beta}_0$	RMSE	0.0743	0.0522	0.0393	0.0541	0.0382	0.0306
		BIAS	0.0584	0.0408	0.0310	0.0430	0.0305	0.0246
		SD	0.0711	0.0491	0.0356	0.0528	0.0370	0.0273
	$\hat{\beta}(z)$	RAMSE	0.1868	0.1511	0.0929	0.1617	0.1090	0.0731
		ABIAS	0.1497	0.1193	0.0736	0.1264	0.0876	0.0568
		ASD	0.1123	0.0932	0.0569	0.1013	0.0652	0.0462

$\times$ in DGP_2), against its parametric rate of $1/\sqrt{nT}$. Clearly, each parameter estimate decays at a rate quite close to the theoretical rate. Similarly, Figure 1(b) plots the RAMSE of the function estimator $\hat{\beta}(\cdot)$ in DGP_1 (dash line with o) and DGP_2 (dot line with $+$), which closely resembles against its nonparametric rate of $1/\sqrt{Tb}$. Therefore, the numerical properties of our estimator support our theoretical arguments on the $\sqrt{nT}$ ($\sqrt{Tb}$)-consistent estimator for our interested unknowns θ_0 ($\beta(\cdot)$).

To clearly demonstrate the performance of the function estimator, Figure 1 plots kernel estimates $\hat{\beta}(\cdot)$

Figure 1: Convergence Rates (First Row) and Kernel Estimation of $\beta(z)$ (Second Row)



with a sample $(n, T) = (30, 97)$, the same sample size in our empirical study, in Panels (c)-(d) for DGP_1 and DGP_2, respectively. Each function estimate (dot line with o) is plotted against its point-wise 95% confidence interval (CI) based on Theorem 2(b) (dashed line) and true function (solid line). Consistent with Table A.1, the estimates sufficiently reveal the shape of the unknown coefficient function across all DGPs. We note that the bias is smaller for functions with a lower degree of curvature (i.e., DGP_1) compared to that with a larger degree of curvature (i.e., DGP_2). Overall, our proposed estimator demonstrates appealing numerical performance in our semiparametric dynamic skewed panel model with fixed effects and smooth coefficient.

Appendix 2.2: Simulation under Incorrect Distributional Assumptions

We now focus on cases where the skew normal distribution (SN) is misspecified. The true distribution may deviate from the assumed SN in terms of skewness and kurtosis. In addition to SN, there are a few alternative asymmetric distributions allowing for either positive or negative skewness. We select skewed t distribution (ST) and asymmetric Laplace distribution (AL) as two such alternatives, each exhibiting different tail behaviors compared to SN. To facilitate our discussion, we denote the mean of the SN error e as $\mu_t^{SN} \equiv E(e_{i,t})$ from

$$\mu_t^{SN} = \sqrt{\frac{2}{\pi}} \frac{\sigma_t \lambda_t}{1+\lambda_t^2}, \tag{A.24}$$

where $\sigma_t \equiv \sigma(x_t, z_t; \gamma_{0\sigma})$ and $\lambda_t \equiv \lambda(x_t, z_t; \gamma_{0\lambda})$, respectively, are parametric conditional scale and skewness functions.

Skew t Distribution. In our semiparametric dynamic panel model, the error follows a ST if

$$e_{i,t} \sim ST(0, \sigma_t, \lambda_t, \nu),$$

which, compared to SN in (A.23), is characterized by one additional parameter $\nu > 1$, or the degree of freedom that regulates the tail sickness (Azzalini and Capitanio, 2003). As $\nu \to \infty$, the ST becomes $SN(0, \sigma_t, \lambda_t)$. The density function of $e_{i,t}$ is given by

$$f(e_{i,t}) = \frac{2}{\sigma_t} \phi_{ST}\left(\frac{e_{i,t}}{\sigma_t}; \nu\right) \Phi_{ST}\left(\frac{\lambda_t e_{i,t}(\nu+1)}{\nu + (e_{i,t}/\sigma_t)^2}; \nu+1\right), \tag{A.25}$$

where ϕ_{ST} and Φ_{ST} refer to the PDF and CDF of the standard ST distribution, respectively. It can be shown that the mean of $e_{i,t}$ under ST, denoted as $\mu_t^{ST} \equiv E(e_{i,t})$, is

$$\mu_t^{ST} = b(\nu) \frac{\sigma_t \lambda_t}{1+\lambda_t^2}, \tag{A.26}$$

where $b(\nu) = \sqrt{\nu/\pi}\Gamma(0.5(\nu-1))\Gamma(\nu/2)^{-1}$ is a decreasing function of ν and $\Gamma(r) = \int e^t t^{r-1} dt$ is gamma function. It can be shown that $\lim_{\nu \to 1} b(\nu) = \infty$ and $\lim_{\nu \to \infty} b(\nu) = \sqrt{2/\pi}$. Denote the difference in the mean functions between ST and SN as

$$D_t(\nu) = \mu_t^{ST} - \mu_t^{SN} = \left(b(\nu) - \sqrt{\frac{2}{\pi}}\right) \frac{\sigma_t \lambda_t}{1+\lambda_t^2},$$

which is a decreasing function of ν given σ_t and λ_t. Since $|\lambda_t/(1+\lambda_t^2)| < 1$, $|D_t(\nu)| < |b(\nu) - \sqrt{2/\pi}|\sigma_t$, implying that $\lim_{\nu \to \infty} |D_t(\nu)| = 0$ and $\lim_{\nu \to 1} |D_t(\nu)| = \infty$.

Denote $\tilde{\zeta}_t = \zeta_{i,t} - \frac{1}{n}\sum_{i=1}^n \zeta_{i,t}$ for any variable $\zeta_{i,t}$. Based on the within-transformed model in our first step (see equation (4) of the paper), $e_{i,t} = \tilde{y}_{i,t} - \tilde{\omega}_{i,t}^\top \delta_0 - d_{i,-1}^\top \alpha_{0,-1} + \mu_t^{ST} + O_p(1/\sqrt{n}) \xrightarrow{d} ST(0, \sigma_t, \lambda_t, \nu)$.

Table A.2: Simulation Results for Two-Step Estimator under Incorrect Distribution: **ST** ($n = 30$)

		$\hat{\gamma}_\lambda$	$\hat{\gamma}_\sigma$	$\hat{\delta}_\omega$	$\hat{b}(z)$
		RMSE	RMSE	RMSE	RAMSE
$SN(0, \sigma_t, \lambda_t)$	$T = 50$	0.2070	0.0207	0.0250	0.0569
	$T = 100$	0.1218	0.0124	0.0219	0.0366
	$T = 200$	0.0933	0.0088	0.0152	0.0265
$ST(0, \sigma_t, \lambda_t, \nu = 5)$	$T = 50$	0.2688	0.1257	0.0337	0.2058
	$T = 100$	0.2470	0.1240	0.0241	0.2001
	$T = 200$	0.2293	0.1230	0.0151	0.1971
$ST(0, \sigma_t, \lambda_t, \nu = 15)$	$T = 50$	0.2062	0.0340	0.0284	0.0603
	$T = 100$	0.1443	0.0335	0.0197	0.0488
	$T = 200$	0.1331	0.0292	0.0136	0.0418
$ST(0, \sigma_t, \lambda_t, \nu = 35)$	$T = 50$	0.1812	0.0172	0.0254	0.0481
	$T = 100$	0.1331	0.0129	0.0205	0.0359
	$T = 200$	0.0983	0.0093	0.0132	0.0264

However, since the distributional assumption is violated, the true distribution of e is ST but misspecified as SN. This implies that $e_{i,t} - D_t(\nu) = \tilde{y}_{i,t} - \tilde{\omega}_{i,t}^\top \delta_0 - d_{i,-1}^\top \alpha_{0,-1} + \mu_t^{SN} + O_p(1/\sqrt{n})$, or equivalently,

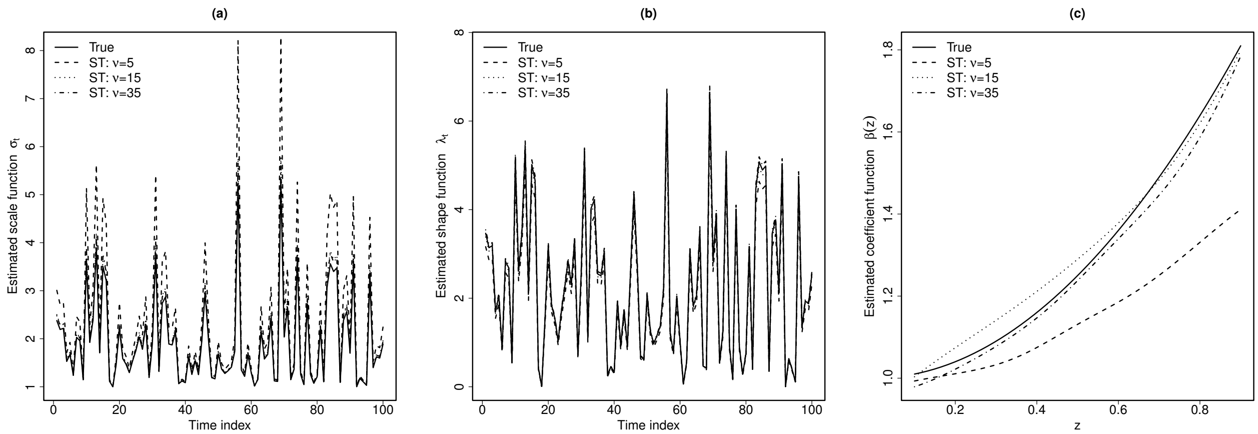
$$e_{i,t}^c = \epsilon_{i,t}^*(\vartheta) + O_p(1/\sqrt{n}) \xrightarrow{d} ST(-D_t(\nu), \sigma_t, \lambda_t, \nu),$$

where $e_{i,t}^c = e_{i,t} - D_t(\nu)$ and $\epsilon_{i,t}^*(\vartheta) = \tilde{y}_{i,t} - \tilde{\omega}_{i,t}^\top \delta_0 - d_{i,-1}^\top \alpha_{0,-1} + \mu_t^{SN}$ is our pseudo-residual constructed for pseudo-MLE. Notice that 1) μ_t^{SN} on the right-hand-side is misspecified because $\mu_t^{SN} \neq \mu_t^{ST}$ in general; and 2) $e_{i,t}^c$ is now a mean-adjusted error as $E(e_{i,t}^c) = \mu_t^{SN}$, the same to the mean of SN.

The results indicate that when the ST distribution is misspecified as SN, the ST is location-adjusted by a magnitude of $|D_t(\nu)|$. In other words, the true ST with a mean of μ_t^{ST} is shifted to match the mean of the SN and approximated by the SN through our pseudo-MLE. Therefore, a larger $D_t(\nu)$ results in a greater discrepancy between the true ST and misspecified SN in terms of location and tail thickness. Given that the ST covers the SN as a special case when $\nu \to \infty$, our pseudo-MLE $\hat{\vartheta}$ behaves poorly and becomes inconsistent as $\nu \to 1$. In contrast, as ν increases, $\hat{\vartheta}$ performs better and achieves consistency as $\nu \to \infty$. Our second step estimator $(\hat{\beta}_0, \hat{\beta}(z))$ exhibits the same pattern, as their consistency relies on consistent parameter estimates from the first step.

We investigate the numerical properties of our proposed two-step estimator when ST is misspecified as SN. We conduct a similar simulation study from Section 2.1, except that $e_{i,t} \sim ST(0, \sigma_t, \lambda_t, \nu)$ now follows a ST, and $\beta(z) = z + z^2$ highlights nonlinear estimation performance. Given fixed σ_t and λ_t, we choose three degrees of freedom: $\nu = 5$ for heavy tails, $\nu = 15$ for moderate tails, and $\nu = 35$ for thin tails, which

Figure 2: Estimated Scale and Shape Function under ST with Different Specifications



is fairly close to that of SN. For a focused presentation on estimation consistency, we report the RMSE of three selected parameter estimates $(\hat{\gamma}_\lambda, \hat{\gamma}_\sigma, \hat{\delta}_\omega)$ and the RAMSE of function estimates $\hat{\beta}(z)$ in Table A.2 for fixed $n = 30$ and $T = (50, 100, 200)$. The top panel presents the results under the correct specification of SN for comparison purposes. Figure A.2 provides a vivid illustration by reporting: 1) scale function estimates $\hat{\sigma}_t = x_t z_t \hat{\gamma}_\sigma$ in panel (a); 2) shape function estimates $\hat{\lambda}_t = x_t z_t \hat{\gamma}_\lambda$ in panel (b); and 3) coefficient function estimates $\hat{\beta}(z)$ in panel (c).

The results are consistent with our conjecture. For $\nu = 5$ where the ST exhibits heavy tails, the performance of the estimators in both steps is poor and inconsistent, indicated by the large and non-decreasing RMSE or RAMSE. In Figure 2, we observe the largest deviation of $\hat{\sigma}_t$ and $\hat{\lambda}_t$ under ST with $\nu = 5$ (dashed line) from the true function (solid line), followed by those with $\nu = 15$ (dotted line) and $\nu = 35$ (dot-dash line). A similar pattern is observed for $\hat{\beta}(z)$. However, with a larger $\nu = 35$, $D_t(\nu)$ approaches zero, and the ST resembles the SN more closely in terms of mean and thin tails. In this case, the two-step estimator shows significant improvement and become consistent.

The results suggest that one may need to consider ST if the conditional distribution of the dependent variable is deemed to have heavy tails based on economic theories or empirical kurtosis. In this case, one needs to construct log-likelihood functions of ST based on (A.25), where the CDF of ST does not permit a closed form solution and needs to be numerically evaluated. Also, $\mu(x_t, z_t; \gamma_0)$ in our pseudo-MLE should be replaced with $\mu_t^{ST} = b(\nu) \frac{\sigma_t \lambda_t}{1+\lambda_t^2}$, where $\nu > 1$ is an additional parameter to be estimated along with the other parameters. Alternatively, one can estimate $b(\nu) \equiv c$ as a constant and back out ν by numerically approximating the inverse function $b^{-1}(c)$. Notice that for identification purposes, in either case the multiplicative structure in μ_t^{ST} prevents one from adding a constant $c_{0\sigma}$ in σ_t or $c_{0\lambda}$ in λ_t.

Asymmetric Laplace Distribution. The error term follows a AL if

$$e_{i,t} \sim AL(0, \sigma_t, \lambda_t).$$

Similar to SN in (A.23), the AL is characterized by a location parameter at zero, a scale function $\sigma_t > 0$, and a shape function λ_t regulating the skewness, with $\lambda_t = 0$ corresponding to a symmetric Laplace distribution (Kotz et al., 2012). Unlike ST, the AL and SN do not nest within one another as special cases, and we adopt the same notations for scale and shape functions only for notation simplicity. The density function of AL is given by

$$f(e_{i,t}) = \frac{2}{\sigma_t} \frac{\kappa_t}{1+\kappa_t^2} \exp\left(-\frac{\sqrt{2}}{\sigma_t} \kappa_t^{\text{sign}(e_{i,t})} |e_{i,t}| \right), \tag{A.27}$$

where $\kappa_t = \sqrt{2}\sigma_t/(\lambda_t + \sqrt{2\sigma_t^2 + \lambda_t^2})$ and $\text{sign}(e) = 1$ ($\text{sign}(e) = -1$) if $e \geq 0$ ($e < 0$). Notably, the AL exhibits a peak (i.e., non-differentiable) point at $e = 0$, therefore different from SN or ST which are continuously differentiable everywhere. As a graphical illustration, we simplify $\sigma_t = \sigma = \exp(1)$ and $\lambda_t = \lambda$ for all t, and plot the density of $AL(0, \sigma, \lambda = 2)$ (dash line), $AL(0, \sigma, \lambda = 1)$ (dot line), and $AL(0, \sigma, \lambda = 0)$ (dot-dash line) in Figure 3 (a), against $SN(0, \sigma, \lambda = 0)$ (solid line). Notice that, compared to SN, the observations under AL are highly condensed around zero regardless of λ, thus exhibiting larger kurtosis than SN.

The mean of $e_{i,t}$ under AL, denoted as $\mu_t^{AL} \equiv E(e_{i,t})$, is

$$\mu_t^{AL} = \lambda_t, \tag{A.28}$$

which is equivalent to the degree of skewness. Given a fixed σ_t, denote the difference in the mean functions between AL and SN as

$$D(\lambda_t) = \mu_t^{AL} - \mu_t^{SN} = \lambda_t - \sqrt{\frac{2}{\pi}} \frac{\sigma_t \lambda_t}{1+\lambda_t^2}.$$

Clearly, $|D(\lambda_t)| < |\lambda_t| + \sqrt{2/\pi}\sigma_t$, and $D(0) = 0$ for all σ_t when both AL and SN are symmetric.

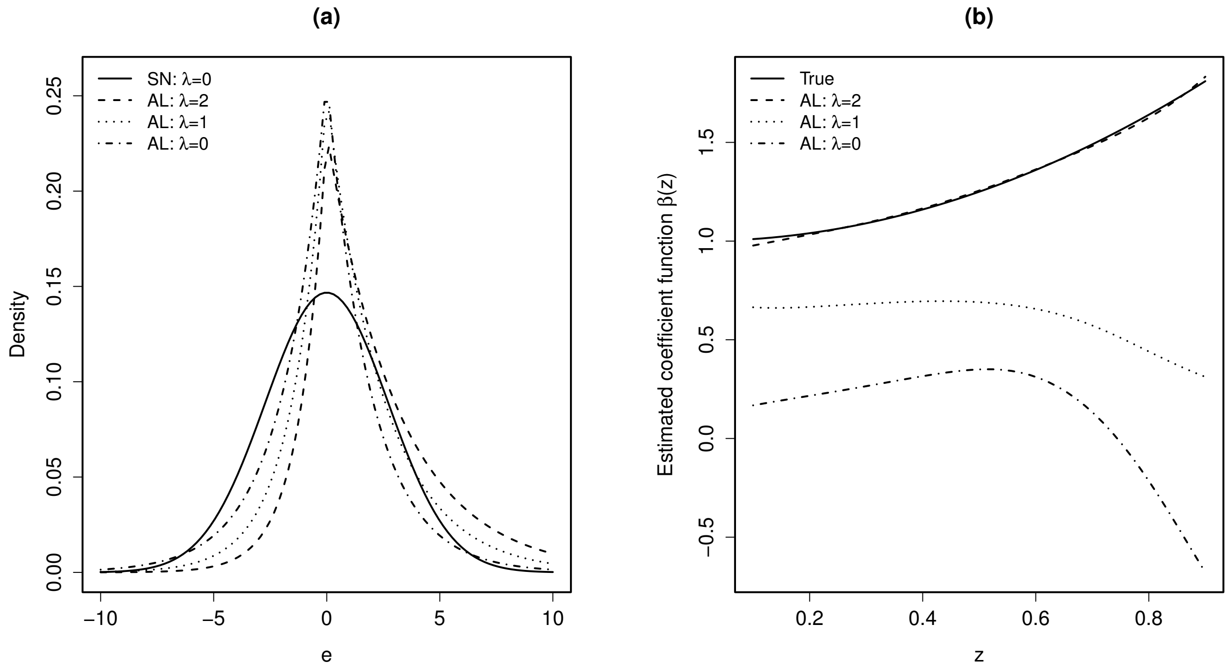
Under a misspecified distributional assumption, the true distribution of e is AL but misspecified as SN. Following similar arguments in the case of ST, we see that

$$e_{i,t}^c = \epsilon_{i,t}^*(\vartheta) + O_p(1/\sqrt{n}) \xrightarrow{d} AL(-D(\lambda_t), \sigma_t, \lambda_t),$$

where $e_{i,t}^c = e_{i,t} - D(\lambda_t)$. Again, μ_t^{SN} on the right-hand-side is misspecified because $\mu_t^{SN} \neq \mu_t^{AL}$, and $e_{i,t}^c$ is a mean-adjusted error such that $E(e_{i,t}^c) = \mu_t^{SN}$.

The results indicate that when the AL is misspecified as SN, the AL is location-adjusted by a magnitude of $|D(\lambda_t)|$. Similar to the case of ST, the true AL with a mean of μ_t^{AL} is shifted to match the mean of the

Figure 3: Plots of AL (a) and Estimated Coefficient Functions (b) under Different Skewness



SN and is approximated by the SN through the pseudo-MLE. Therefore, a larger $D(\lambda_t)$ results in a greater discrepancy between the AL and SN. We conjecture that our two-step estimator may perform best in terms of bias when both SN and AL distributions are symmetric, i.e., $D(\lambda_t = 0) = 0$. In cases where $\lambda_t \neq 0$, our estimator becomes more biased, as the SN poorly approximates the AL due to its fatter tails. However, in either situation, our estimator is inconsistent because the likelihood function, misspecified a SN, does not converge to AL. This contrasts with the ST, whose likelihood function converges to that of the SN as the degrees of freedom approach infinity, thereby improving our estimator's performance.

We investigate the numerical properties of our proposed two-step estimator where AL is misspecified as SN. We repeat the experiment from Section 2.1, with the adjustment that $e_{i,t} \sim AL(0, \sigma_t, \lambda_t)$ and $\beta(z) = z + z^2$. We found that specifying time-varying scale and shape functions in the AL often renders our first-step estimator infeasible. This occurs because using a SN to approximate an AL with fat tails results in notably large residuals, which in turn makes the skew normal likelihood function undefined. For demonstration purposes, we simplify both functions to constants, setting $\sigma_t = \sigma = \exp(c_{0\sigma})$ and $\lambda_t = \lambda$ for all t. We fix $\gamma_{0\sigma} = 1$ and select $\lambda = (0, 1, 2)$, corresponding to zero skewness, moderate skewness, and large skewness, respectively.

Given the AL with three different skewness levels, Table A.3 reports the results for $n = 30$ with $T =$

Table A.3: Simulation Results for Two-Step Estimator under Incorrect Distribution of AL ($n = 30$)

		$\hat{\lambda}$	$\hat{c}_\sigma$	$\hat{\delta}_\omega$	$\hat{b}(z)$
		RMSE	RMSE	RMSE	RAMSE
$SN(0, \exp(1), \lambda = 0)$	$T = 50$	0.0980	0.0654	0.0562	0.0848
	$T = 100$	0.0868	0.0502	0.0394	0.0629
	$T = 200$	0.0669	0.0407	0.0243	0.0463
$AL(0, \exp(1), \lambda = 2)$	$T = 50$	1.3547	0.5717	0.0666	0.1200
	$T = 100$	1.3331	0.5766	0.0423	0.0855
	$T = 200$	1.3019	0.5774	0.0310	0.0580
$AL(0, \exp(1), \lambda = 1)$	$T = 50$	0.9613	0.3780	0.0670	0.0898
	$T = 100$	0.9694	0.3708	0.0447	0.0770
	$T = 200$	0.9453	0.3728	0.0323	0.0443
$AL(0, \exp(1), \lambda = 0)$	$T = 50$	0.7099	0.1235	0.0651	0.0820
	$T = 100$	0.7148	0.1260	0.0402	0.0658
	$T = 200$	0.6988	0.1187	0.0299	0.0444

Table A.4: Estimated Scale and Shape Functions under Incorrect Distribution of AL: $(n, T) = (30, 100)$

	$AL(0, \exp(1), \lambda = 2)$		$AL(0, \exp(1), \lambda = 1)$		$AL(0, \exp(1), \lambda = 0)$	
	λ	σ	λ	σ	λ	σ
True	2.0000	1.6487	1.0000	1.6487	0.0000	1.6487
Estimates	2.1065	2.4393	1.1100	2.1210	0.0005	1.6213

$(50, 100, 200)$, where the results for a correctly specified SN are shown in the top panel. Figure 3 (b) depicts the estimated coefficient functions $\hat{\beta}(z)$. Consistent with our conjecture, an AL with heavy tails ($\lambda = 2$) is challenging to be approximated by a SN, leading to the poorest performance of our estimator compared to the other two cases. In particular, the first-step parameters $(\lambda, \gamma_{0\sigma})$ are estimated with significant bias, and the function estimates $\hat{\beta}(z)$ deviate considerably from the true function in both magnitude and shape. As λ decreases to one or zero, $D(\lambda) \to 0$ and the tails become thinner, resulting in a relative improvement in the two-step estimator.

Table 3 compares the true and estimated scale parameter $\sigma = \exp(c_{0\sigma})$ and shape parameter λ given a sample size of $(n, T) = (30, 100)$. Across all three cases, the bias of $(\hat{\lambda}, \hat{c}_\sigma)$ shrinks as $\lambda \to 0$, although the bias *given each* λ does not rapidly decay toward zero, as indicated by non-decreasing RMSE in Table 2. The performance of $\hat{\delta}_w$ and $\hat{\beta}(z)$ is relatively less affected by the incorrect distribution.

The simulation results suggest that our estimator may provide relatively reasonable results only when AL is more symmetric. If the dependent variable is distributed with heavy tails or a notable peak, AL is more appropriate to accommodate those tail behaviors. In this case, one needs to construct log-likelihood functions based on (A.27), and replaces $\mu(x_t, z_t; \gamma_0)$ in our pseudo-MLE with $\mu_t^{AL} = \frac{\sigma_t}{\sqrt{2}} \left(\kappa_t^{-1} - \kappa_t\right)$, where $\kappa_t = \frac{\sqrt{2}\sigma_t}{\lambda_t + \sqrt{2\sigma_t^2 + \lambda_t^2}} \neq 1$.

Appendix 3: MCMC Algorithm for the Skew Normal Stochastic Volatility Model

Recall that we model return series in equations (13)-(15) in Section 4.1 with a simple time-varying parameter model that deviates from the symmetric distributional assumption as

$$r_{j,t} = \mu_t^j + \sigma_t^j \eta_{j,t}, \quad \eta_{j,t} \sim SN(0, 1, \alpha^j), \tag{A.29}$$

where $j \in \{mkt, max, min\}$. Concerning the time-varying parameters, we assume the mean return follows a driftless random walk process:

$$\mu_t = \mu_{t-1} + \epsilon_t^\mu, \quad \epsilon_t^\mu \sim N(0, \psi^\mu). \tag{A.30}$$

Equation (A.30) captures potential secular trend of each series. Our primary focus is on the volatility of each series. We assume that they comove substantially but allow for non-synchronized movements. Specifically, we assume that

$$\sigma_t^{max} = \sigma_t^{mkt} \sigma_t^+, \quad \sigma_t^{min} = \sigma_t^{mkt} \sigma_t^-. \tag{A.31}$$

The volatility of the market return, represented by σ_t^{mkt}, determines the general level of volatility of the stock market, but there are additional factors, σ_t^+ and σ_t^-, influencing the volatility of the maximum and the minimum returns. Let $\mathbf{h}_t = [\ln(\sigma_t^{mkt})^2, \ln(\sigma_t^+)^2, \ln(\sigma_t^-)^2]'$, the evolving process of $\mathbf{h}_t$ is assumed to follow a driftless random walk:

$$\mathbf{h}_t = \mathbf{h}_{t-1} + \epsilon_t. \tag{A.32}$$

Equations (A.29) - (A.32) form a widely used local level model with stochastic volatility, except that the distribution of $\eta_{j,t}$ is extended to allow for asymmetric properties. We term this model the Skew Normal Stochastic Volatility (SNSV) model. Estimation is carried out using the Bayesian MCMC approach proposed by Huang and Luo (2020) with minor adjustments due to restrictions (A.31).

The nonlinear relationships listed in (A.31) using the algorithm proposed by Huang and Luo (2020). We therefore adopt a simple two-step strategy to deal with this difficulty. In the first step, we estimate σ_t^{mkt} using the algorithm proposed by Huang and Luo (2020). In the second step, re-scale r_t^{max} and r_t^{min} by $\hat{\sigma}_t^{mkt}$ obtained from the first step, and implement the algorithm proposed by Huang and Luo (2020) to draw samples from the posterior distributions of σ_t^+ and σ_t^-.

The mean and quantiles of the posterior are approximated by MCMC draws. The details are as follows. We first introduce the background of stochastic representation of skew normality. A skew normal random variable, $Z \sim SN(0,1,\alpha)$, can be represented as

$$Z = \delta V + \sqrt{1-\delta^2}U, \tag{A.33}$$

where $V \sim N^{+}(0,1)$, $U \sim N(0,1)$, and $\delta = sign(\alpha)\sqrt{\alpha^2/(1+\alpha^2)}$ (Azzalini, 2013). Based on this stochastic representation, the disturbance term for the log return can be written as

$$\eta_t = \exp(h_t)\eta_t^* = \exp(h_t)(\delta v_t + \sqrt{1-\delta^2}u_t) = \beta_t v_t + q_t, \quad q_t \sim N(0, \sigma_q^2). \tag{A.34}$$

where $\beta_t = \exp(h_t)\delta$ and $\sigma_q = \exp(h_t)\sqrt{1-\delta^2}$.

Let Y denote the observed data $\{r_t\}_{t=1}^T$. Let θ denote the vector of unknown random variables for which Gibbs draws will be taken. In this model, $\theta = (\theta_1, \theta_2, \theta_3, \theta_4, \theta_5)$ where

$$\theta_1 = \{\mu_1 \ldots \mu_T\}, \quad \theta_2 = \{h_1 \ldots, h_T\}, \quad \theta_3 = \alpha, \quad \theta_4 = \{\psi^\mu, \psi^h\}, \quad \theta_5 = \{v_1, \ldots, v_T\}.$$

In the following, we outline the procedure for Gibbs draws.

1. Draw θ_1 from $f(\theta_1|Y, \theta_2, \theta_3, \theta_4, \theta_5)$. Conditioning on a draw of v_t, Equation (A.29) - (A.30) form a conditionally linear Gaussian unobserved component model. Conditional moments can be computed from Kalman smoother and draws of μ_t can be conveniently obtained using the insights in Carter and Kohn (1994).

2. Draw θ_2 from $f(\theta_2|Y, \theta_1, \theta_3, \theta_4, \theta_5)$. Jacquier et al. (1994) proposed a single-move Metropolis-Hasting algorithm for drawing latent stochastic volatility from models with any distributional assumption. Conditional on θ_1, the error terms $\eta_t = \sigma_t \eta_t^*$ are observed. Note that

$$f(h_t|h_{t-1}, h_{t+1}, q_t, \Theta) \propto f(h_t|h_{t-1}, h_{t+1}, \Theta) f(\eta_t|h_t, \Theta),$$

   We use a random walk Metropolis-Hastings sampler to draw from the above density.

3. Draw θ_3 from $f(\theta_3|Y, \theta_1, \theta_2, \theta_4, \theta_5)$. Conditional on θ_1 and θ_2, the error terms $\eta_t^* = \eta_t/\sigma_t$ are observed. Given a normal prior $\alpha \sim N(\alpha_0, V_{\alpha_0})$, the posterior kernel density is

$$f(\theta_3|Y, \theta_1, \theta_2, \theta_4, \theta_5) \propto f_N(\alpha|\alpha_0, V_{\alpha_0}) \prod_{t=1}^{T} f(\eta_t^*|\alpha),$$

   We use a Metropolis-within-Gibbs algorithm to draw a random sample from the above kernel density.

4. Draw θ_4 from $f(\theta_4|Y, \theta_1, \theta_2, \theta_3, \theta_5)$. Conditional on θ_2, ϵ_t^μ and ϵ_t^h are observed. Assume that the prior for the inverse-Gamma ψ^μ and ψ^h with hyperparameters a_0 and b_0, the posterior distribution of ψ^μ and ψ^h is also inverse-Gamma with parameters $a_0 + T/2$ and $b_0 + \sum_{t=1}^T \epsilon_t^i/2$ for $i = \mu, h$.

5. Draw θ_5 from $f(\theta_5|Y, \theta_1, \theta_2, \theta_3, \theta_4)$. Conditional on μ_t and the data, we observe η_t. Following Frühwirth-Schnatter and Pyne (2010), we can show that

$$v_t|\eta_t \sim N^+\left(\frac{A_t\beta_t}{\sigma_{qt}^2}\eta_t, A_t\right), \quad A_t = \left(\frac{\beta_t^2}{\sigma_{qt}^2} + 1\right)^{-1}, \tag{A.35}$$

where N^+ is a truncated normal with a lower truncation at 0. Equation (A.35) can be applied to generate posterior samples of $\{v_1, \ldots, v_T\}$.

The priors are set as follows: $\mu_0 \sim N(0, 100)$, $h_0 \sim N(0, 100)$, $\alpha \sim N(0, 10)$, $a_0 = 1$, $b_0 = 0.05$, $v_t \sim N^+(0, 1)$.

References

Amemiya, Takeshi, *Advanced Econometrics*, Harvard university press, 1985.

Arellano, Manuel, Jinyong Hahn et al., "Understanding bias in nonlinear panel models: Some recent developments," *Econometric Society Monographs*, 2007, *43*, 381.

Azzalini, A and A Capitanio, "Distributions generated by pertubation of symmetry with emphasis on a multivariate skew t distribution," *Journal of the Royal Statistical Society Series B*, 2003, *65*, 367–389.

Azzalini, Adelchi, *The Skew-Normal and Related Families*, Vol. 3, Cambridge University Press, 2013.

Bosq, D, *Nonparametric statistics for stochastic processes: estimation and prediction. No. 110 in Lecture notes in Statistics.* number 110. In 'Lecture Notes in Statistics.', Berlin: Springer Verlag, 1996.

Cai, Z. and Q. Li, "Nonparametric estimation of varying coefficient dynamic panel data models," *Econometric Thoery*, 2008, *24*, 1321–1342.

Carter, Chris K and Robert Kohn, "On Gibbs sampling for state space models," *Biometrika*, 1994, *81* (3), 541–553.

Chen, J., J. Gao, and D. Li, "Estimation in partially linear single-index panel data models with fixed effects," *Journal of Business & Economic Statistics*, 2013, *31*, 315–330.

Davidson, J, *Stochastic Limit Theory*, New York: Oxford University Press, 1994.

Fan, J. and I. Gijbels, *Local Polynomial Modeling and its Applications*, London: Chapman and Hall, 1996.

Fan, J and Q Yao, *Nonlinear Time Series: Nonparametric and Parametric Methods*, Springer Verlag, 2003.

Fan, Jianqing and Tao Huang, "Profile likelihood inferences on semiparametric varying-coefficient partially linear models," *Bernoulli*, 2005, *11* (6), 1031–1057.

Frühwirth-Schnatter, Sylvia and Saumyadipta Pyne, "Bayesian Inference for Finite Mixtures of Univariate and Multivariate Skew-Normal and Skew-T Distributions," *Biostatistics*, 2010, *11* (2), 317–336.

Graves, Lawrance M, "Riemann integration and Taylor's theorem in general analysis," *Transactions of the American Mathematical Society*, 1927, *29*, 163–177.

Hahn, Jinyong and Hyungsik Roger Moon, "Reducing bias of MLE in a dynamic panel model," *Econometric Theory*, 2006, *22* (3), 499–512.

_ **and Whitney Newey**, "Jackknife and analytical bias reduction for nonlinear panel models," *Econometrica*, 2004, *72* (4), 1295–1319.

Hall, P. and C. C. Heyde, *Martingale Limit Theory and Its Applications*, Academic Press, 1980.

Huang, Yu-Fan and Sui Luo, "Can Stock Volatility Be Benign? New Measurements and Macroeconomic Implications," *Journal of Money, Credit and Banking*, 2020, *52* (4), 933–950.

Jacquier, Eric, Nicholas G. Polson, and Peter E. Rossi, "Bayesian Analysis of Stochastic Volatility Models," *Journal of Business & Economic Statistics*, October 1994, *12* (4), 371–89.

Kotz, Samuel, Tomasz Kozubowski, and Krzystof Podgorski, *The Laplace distribution and generalizations: a revisit with applications to communications, economics, engineering, and finance*, Springer Science & Business Media, 2012.

Li, Q and J Racine, *Nonparametric Econometrics: Theory and Practice*, Princeton, NJ: Princeton University Press, 2007.

Martins-Filho, C and F Yao, "Nonparametric regression estimation with general parametric error covariance," *Journal of Multivariate Analysis*, 2009, *100*, 309–333.

Newey, W and D McFadden, "Large sample estimation and hypothesis testing," in R F Engle and D L McFadden, eds., *Handbook of Econometrics*, Vol. 4, Amsterdam: Elsevier Science B.V., 1994.

Rilstone, Paul, Virendra K Srivastava, and Aman Ullah, "The second-order bias and mean squared error of nonlinear estimators," *Journal of Econometrics*, 1996, *75* (2), 369–395.

Su, L. and A. Ullah, "Profile likelihood estimation of partially linear panel data models with fixed effects," *Economics Letters*, 2006, *92*, 75–81.

Sun, Yiguo, Raymond J Carroll, and Dingding Li, "Semiparametric estimation of fixed-effects panel data varying coefficient models," in "Nonparametric econometric methods," Emerald Group Publishing Limited, 2009, pp. 101–129.

Wooldridge, Jeffrey M, *Econometric Analysis of Cross Section and Panel Data*, MIT press, 2010.