A flexible stochastic production frontier model with panel data

Taining Wang

International School of Economics and Management,
Capital University of Economics and Business,
Beijing, 100070, China.
Email: tnwang@mix.wvu.edu

Feng Yao

Department of Economics
West Virginia University
Morgantown, WV 26505, USA
email: feng.yao@mail.wvu.edu
Voice: +1 304 293 7867

Subal C. Kumbhakar

Department of Economics
Binghamton University
Binghamton, NY 13902, USA
email: kkar@binghamton.edu
Voice: + 1 607 777 4762

October 4, 2021

Abstract: We propose a flexible stochastic production frontier model with fixed effects for the panel data in which the semiparametric frontier is additive with bivariate interactions. Instead of maintaining distributional assumptions, we model the conditional mean of the inefficiency to depend on environmental variables and to be known up to a vector of parameters. We propose a difference-based estimator for parameters characterizing the conditional mean of the inefficiency term, a profile series estimator and a kernel-based one-step backfitting estimator for the frontier to facilitate inference. We establish their asymptotic properties, and show that each component in the frontier estimated by the kernel-based backfitting has the same asymptotic distribution as that estimated with the true knowledge on the other components in the frontier (i.e., the oracle property). Through a Monte Carlo study, we demonstrate that the proposed estimators perform well in finite samples. Through an application, we illustrates their applicability in estimating the frontier and obtaining the efficiency score.

Keywords: Semiparametric additive model; Technical efficiency; Profile estimation; One-step backfitting.

JEL Classifcations: C14, C22.

1 Introduction

Since the seminar work of Aigner et al. (1977) and Meeusen and van den Broeck (1977), the stochastic frontier (SF) approach as a tool to model and estimate efficiency has grown exponentially (see Greene (1993), Coelli (1995), Kumbhakar and Lovell (2000), Parmeter and Kumbhakar (2014), and Kumbhakar et al. (2015) for extensive reviews of different variants of the SF models and their applications).[1] The SF models are popular among the econometricians due to the fact that these models accommodate stochastic noise (production shocks) and can separate the noise from inefficiency. Furthermore, one can perform statistical tests on many economic hypotheses of interests. In doing so, however, restrictive assumptions are made on either the production frontier function and/or the distributional assumptions on the inefficiency and the noise terms. For example, the SF model pioneered by Aigner et al. (1977) and Meeusen and van den Broeck (1977) and extended by many in the last three decades uses a parametric frontier along with a composite error term in which the one-sided inefficiency term follows a particular distribution (half-normal, exponential, truncated normal, gamma, etc.), while the noise term follows a normal distribution. This basic structure is followed in panel models as well with some extensions on the temporal behavior of inefficiency.

With panel data, one can relax the distribution assumption on the inefficiency term at the expense of some other restrictive assumptions. For example, Schmidt and Sickles (1984) estimate a parametric frontier function by relaxing the distribution assumption in a panel data framework by assuming inefficiency to be time-invariant (parameter). Similarly, Cornwell et al. (1990) and Lee and Schmidt (1993) assume inefficiency to follow a deterministic function of time. Recently, Horrace and Parmeter (2011) propose to estimate the inefficiency distribution, while Parmeter et al. (2014) estimate the determinant function of the inefficiency, both without making distributional assumption *a priori* on the inefficiency term. However, the flexibility in modeling the frontier is still limited to some known parametric functional forms, such as the Cobb-Douglas and translog. Even with correctly specified distribution for the composite errors, incorrectly specified frontier can still lead to misleading conclusion regarding the inefficiency levels.

[1] An alternative to the SF approach is the Data Envelopment Analysis which is comprehensively summarized by Simar and Wilson (2008).

Maintaining the distribution structure in Aigner et al. (1977), Fan et al. (1996) first introduce a nonparametric frontier model and examine properties of the estimator in their simulations. Martins-Filho and Yao (2015) investigate asymptotic properties of the estimator in Fan et al. (1996), and further propose a profile-likelihood based estimator for the nonparametric frontier function where the parameters of the composite error distribution carry no asymptotic bias and are efficient in a class of semiparametric estimators defined in Severini and Wong (1992). On ther other hand, Kumbhakar et al. (2007) take a very different approach to model all parameters of the distribution of the composite error as smooth functions of the inputs. They estimate the nonparametric frontier and the distribution parameters with a local likelihood method. Their method is generalized to include discrete regessors in Park et al. (2015). A common feature of the above methods is that the frontier is fully nonparametric, although in general the rate of convergence of the proposed frontier estimator is rather slow especially when the number of inputs (conditioning variables) is large. The problem is more severe in Kumbhakar et al. (2007) since all "parameters" are local. It is the well-known curse of dimensionality problem afflicting multivariate kernel based nonparametric estimation. Since it is common to have a large number of variables in frontier models, the accuracy of the asymptotic approximation can be rather poor. Yao et al. (2019) propose a smooth coefficient frontier model for which the sample size required for estimation is not as demanding as the fully nonparametric frontier model, but the traditional composite error distribution assumption is retained.

In this paper, we propose a flexible stochastic frontier model for panel data that exhibits three distinctive features. First, we account for the individual heterogeneity with an additive fixed effect. In the frontier estimation, it is desirable to allow for the latent heterogeneity, which can be freely correlated with the inputs or environmental variables, separable from the inefficiency term (see Section 7 in Parmeter and Kumbhakar (2014)). Schmidt and Sickles (1984) incorporate the fixed effect in a parametric frontier. In the smooth coefficient frontier model as in Yao et al. (2019), a fixed effect is not present. Second, we propose a flexible semiparametric stochastic frontier, which is additive with bivariate interactions, each being a nonparametric smooth function. The additive structure allows us to alleviate the burden from the curse of dimensionality, say, relative to the fully nonparametric model, while bivariate terms accommodate the empirically much desired interactions. The smooth coefficient frontier in Yao et al. (2019) require clear distinction between the regular inputs (I), and the environmental variables (E). I affect the frontier linearly, and E enter the frontier through the smooth coefficient function which is multiplied with I. Furthermore, only E can be the determinants for the inefficiency term. Though useful, these assumptions can be restrictive in practice. For example, the extended translog production function, popular in empirical modeling, is generally not a special case of the smooth coefficient frontier model. In our proposed frontier model, I and E do not need to be pre-assigned, I's effect does not have to be linear, and the determinants of the inefficiency term can include any of I and E, or even other exogenous environment variables. Here, the extended translog production function is included in our model as a special case. Finally, we do not need to maintain the distributional assumption on the composite error. Instead, we replace it with a conditional mean specification on the one-sided inefficiency term as a positive function of environmental variables known up to certain parameters.

We propose a nonlinear difference-based estimator for the parameters of the conditional mean of the inefficiency term. For the frontier, we propose a profile series estimator and a kernel-based one-step backfitting estimator to facilitate inference. We establish the asymptotic properties of the proposed estimators. The parameter estimators converge at a rate of $\sqrt{n}$. On the other hand, the estimator of the frontier converges at a slower rate. Due to the additive structure, the convergence rate of the components in the frontier is either that of the univariate nonparametric estimation in the additive component, or that of the bivariate nonparametric estimation in the interaction term, while the dimensions of all the inputs and environmental variables do not impact the convergence rate. This alleviates the curse of dimensionality in the fully nonparametric frontier. Furthermore, the one-step backfiting estimators for the frontier function exhibit the same asymptotic distribution as the estimators assuming the true knowledge of other components in the frontier. We illustrate their finite sample performance through a Monte-Carlo study.

The rest of the article is organized as follows. In Section 2, we present our multi-step semiparametric estimation of the parameters and the additive frontier with bivariate interactions. Section 3 details their asymptotic characterization. We illustrate their finite sample performance in a Monte-Carlo study in Section 4. The empirical application is presented in Section 5. Section 6 concludes the paper. Proof of all the theorems are provided in the Appendix.

2 A multi-step semiparametric estimation

We consider the following semiparametric stochastic frontier model with panel data

$$Y_{it} = \alpha_i + m_0 + \sum_{j=1}^{d} m_j(X_{it}^j) + \sum\sum_{1\le j<l\le d} H_{jl}(X_{it}^j, X_{it}^l) + v_{it} - u_{it},\ i = 1, \cdots, n,\ t = 1, \cdots, T, \tag{1}$$

where y_{it} is the univariate random output, $X_{it} = \{X_{it}^1, \cdots, X_{it}^d\} \in \Re^d$ is a random vector of traditional input variables and exogenous environmental variables, such as $R\&D$ or human capital. The frontier is $g(X_{it}) = m_0 + \sum_{j=1}^{d} m_j(X_{it}^j) + \sum\sum_{1\le j<l\le d} H_{jl}(X_{it}^j, X_{it}^l)$, consisting of the global intercept m_0, additive components $\sum_{j=1}^{d} m_j(X_{it}^j)$, and interaction terms $\sum\sum_{1\le j<l\le d} H_{jl}(X_{it}^j, X_{it}^l)$. $m_j(\cdot)$ for $j = 1 \cdots, d$, $H_{jl}(\cdot)$ for $1 \le j < l \le d$ are unknown smooth functions of X, which allow the frontier function to vary in a completely flexible manner. Thus, different technologies for different firms and different time periods are allowed because X_{it} varies with i and t. α_i is the fixed effect that captures the latent heterogeneity, which can be freely correlated with variables of X . Accommodating α_i in the model allows us to separate the unobserved heterogeneity from the inefficiency term, the level of which would therefore not be overestimated or underestimated. The composite error terms consist of a two sided error term v_{it} representing random noise and a one-sided random term u_{it} representing inefficiency. We do not need to maintain distribution assumptions on the composite error terms, except that the conditional mean of u_{it} is a positive nonlinear function $E(u_{it}|\alpha_i, X_{it}, W_{it}) \equiv \mu(Z_{it}; \gamma_0)$, which is known up to certain parameters $\gamma_0 \in \Re^p$, where variables Z can be any of X and/or additional exogenous environmental variables W.

Obviously, the individual functions in the frontier are not identified without some identification conditions. As our frontier contains both additive and interaction terms, one can follow the kernel estimation in Sperlich et al. (2002) to impose moment conditions $E(m_j(x^j)) = 0$ and $\int H_{jl}(x^j, x^l) f_{x^j}(x^j) dx^j = \int H_{jl}(x^j, x^l) f_{x^l}(x^l) dx^l = 0$, where $f_{x^j}(x^j)$ is the marginal density of x^j, $j, l = 1, \cdots d$ and $j \neq l$. Here, we use series estimation method. It is more straightforward to follow Li (2000) and Li and Racine (2007) to impose the identification condition as

$$m_j(x^j=0)=0, \forall j=1,\cdots,d, H_{jl}(x^j=0,x^l)=H_{jl}(x^j,x^l=0)=0, \forall 1\leq j<l\leq d. \tag{2}$$

Equation (1) is not a proper regression model because the conditional mean of $v_{it}-u_{it}$ is not zero. With $\epsilon_{it}=v_{it}-(u_{it}-\mu(Z_{it};\gamma_0))$, then $E(\epsilon_{it}|\alpha_i,X_{it})=0$ using the Law of Iterated Expectation. So we consider

$$Y_{it}=\alpha_i+m_0+\sum_{j=1}^{d} m_j(X_{it}^j)+\sum\sum_{1\leq j<l\leq d} H_{jl}(X_{it}^j,X_{it}^l)-\mu(Z_{it};\gamma_0)+\epsilon_{it}. \tag{3}$$

To estimate (3), one needs to handle the unobserved fixed effect α_i properly, because its number n can be large. One strategy would be to follow Su and Ullah (2006), Sun et al. (2009), or Chen et al. (2013) to model α_i explicitly assuming $\sum_{i=1}^{n}\alpha_i=0$, use a projection matrix to "partial out" their presence, and then construct profile estimation for the terms of interest, which are the frontier and γ parameters. We do not pursue this path because with a large n, the "partial out" projection matrix has a large dimension and we further need to tease out the delicate structure on the frontier and γ parameters.

In this paper, we propose to perform a first-order difference to wipe out the fixed effect. Specifically,

$$\begin{aligned}Y_{it}-Y_{it-1}= & \sum_{j=1}^{d}(m_j(X_{it}^j)-m_j(X_{it-1}^j))+\sum\sum_{1\leq j<l\leq d}(H_{jl}(X_{it}^j,X_{it}^l)-H_{jl}(X_{it-1}^j,X_{it-1}^l)) \\ & -(\mu(Z_{it};\gamma_0)-\mu(Z_{it-1};\gamma_0))+\epsilon_{it}-\epsilon_{it-1}.\end{aligned}$$

With a short-hand notation $\Delta y_{it}=y_{it}-y_{it-1}$, we denote the above equation as

$$\Delta Y_{it}=\sum_{j=1}^{d}\Delta m_j(X_{it}^j)+\sum\sum_{1\leq j<l\leq d}\Delta H_{jl}(X_{it}^j,x_{it}^l)-\Delta\mu(Z_{it};\gamma_0)+\Delta\epsilon_{it}. \tag{4}$$

Performing a difference in the semiparametric panel data model can be convenient, because the original targets after the difference can be recovered without increasing the dimension of estimation. Utilizing the differenced data to deal with the fixed effects, Qian and Wang (2012) perform a marginal integration, and Baltagi and Li (2002) implement series estimator to estimate a partially linear model. Their setups differs from ours because, even after differencing, equation (4) still contains nonparametric additive and interaction terms, together with γ inside a nonlinear function to be estimated.

In the first step, we approximate the nonlinear functions in the frontier with series and estimate them together with γ parameters. Specifically, with the basis function $\phi^\kappa(x^j)=\{\phi_1(x^j),\cdots,\phi_\kappa(x^j)\}'$, $\theta^j\in\Re^\kappa$,

$m_j(X_{it}^j) - m_j(X_{it-1}^j) \approx (\phi^\kappa(X_{it}^j) - \phi^\kappa(X_{it-1}^j))'\theta^j = \Delta\phi^\kappa(X_{it}^j)'\theta^j$. Similarly, we approximate the differenced bivariate interaction $H_{jl}(X_{it}^j, X_{it}^l) - H_{jl}(X_{it-1}^j, X_{it-1}^l) \approx (\phi^{\kappa^2}(X_{it}^j, X_{it}^l) - \phi^{\kappa^2}(X_{it-1}^j, X_{it-1}^l))'\theta^{jl} = \Delta\phi^{\kappa^2}(X_{it}^j, X_{it}^l)'\theta^{jl}$, where

$$\phi^{\kappa^2}(x^j, x^l) = \{ \begin{array}{l} \phi_1(x^j)\phi_1(x^l), \cdots, \phi_1(x^j)\phi_\kappa(x^l), \\ \phi_2(x^j)\phi_1(x^l), \cdots, \phi_2(x^j)\phi_\kappa(x^l), \\ \vdots \\ \phi_\kappa(x^j)\phi_1(x^l), \cdots, \phi_\kappa(x^j)\phi_\kappa(x^l)\}' \end{array}$$

is the bivariate basis function constructed as the tensor product of the univariate basis functions and $\theta^{jl} \in \Re^{\kappa^2}$. Define $\Delta Y = \{\Delta Y_{it}\}_{t=2}^{T}{}_{i=1}^{n}$, $\Delta\mu(\gamma) = \{\Delta\mu(z_{it};\gamma)\}_{t=2}^{T}{}_{i=1}^{n}$, $\theta^m = (\theta^{1\,\prime}, \cdots, \theta^{d\,\prime})'$, $\theta^H = (\theta^{12\,\prime}, \cdots, \theta^{1d\,\prime}, \theta^{23\,\prime} \cdots,$ $\theta^{2d\,\prime} \cdots, \theta^{d-1d\,\prime})'$. Also, let $\Delta\Phi^{d\kappa}(x) = (\Delta\phi^\kappa(x^1)', \cdots, \Delta\phi^\kappa(x^d)')'$, and with $d_1 = d(d-1)/2$, define

$$\Delta\Phi^{d_1\kappa^2}(x) = (\Delta\phi^{\kappa^2}(x^1, x^2)', \cdots, \Delta\phi^{\kappa^2}(x^1, x^d)', \Delta\phi^{\kappa^2}(x^2, x^3)', \cdots, \Delta\phi^{\kappa^2}(x^2, x^d)', \cdots, \Delta\phi^{\kappa^2}(x^{d-1}, x^d)')'.$$

Furthermore, with $\theta = (\theta^{m\,\prime}, \theta^{H\,\prime})'$, $\Delta\Phi(x) = (\Delta\Phi^{d\kappa}(x)', \Delta\Phi^{d_1\kappa^2}(x)')'$, we define $\Delta P_\kappa(\theta, x) = \Delta\Phi(x)'\theta$ and $\Delta P_\kappa(\theta) = \Delta\Phi\theta$, where $\Delta\Phi = \{\{\Delta\Phi(X_{it})'\}_{t=2}^{T}\}_{i=1}^{n}$. Minimizing the sum of squared residuals, we estimate θ and γ with

$$(\hat{\theta}, \hat{\gamma}) = \underset{\{\theta,\gamma\}}{argmin}\ [\Delta Y - \Delta\Phi\theta + \Delta\mu(\gamma)]'[\Delta Y - \Delta\Phi\theta + \Delta\mu(\gamma)].$$

Note that θ and γ are additively separable. For a given γ, we obtain

$$\hat{\theta}(\gamma) = [\Delta\Phi'\Delta\Phi]^{-1}\Delta\Phi'(\Delta Y + \Delta\mu(\gamma)). \tag{5}$$

Defining $N = n(T-1)$, $M_{\Delta\Phi} = I_N - P_{\Delta\Phi} = I_N - \Delta\Phi(\Delta\Phi'\Delta\Phi)^{-1}\Delta\Phi'$ for I_N being the $N \times N$ identity matrix, we estimate γ with a nonlinear weighted least square as

$$\hat{\gamma} = \underset{\gamma}{argmin}\ [\Delta Y + \Delta\mu(\gamma)]'M_{\Delta\Phi}[\Delta Y + \Delta\mu(\gamma)]. \tag{6}$$

From (5) and (6) , we obtain $\hat{\theta} = \hat{\theta}(\hat{\gamma})$, and the profile based estimators are $\hat{m}_j(x^j) = \phi^\kappa(x^j)'\hat{\theta}^j$ for $j = 1, \cdots, d$, $\hat{H}_{jl}(x^j, x^l) = \phi^{\kappa^2}(x^j, x^l)'\hat{\theta}^{jl}$ for $j < l$, and $j, l \in \{1, \cdots, d\}$.

With the pilot estimation of the frontier performed through series, a kernel-based estimation does offer convenient asymptotic characterization. So we propose an one-step backfitting for $m_j(\cdot)$ and $H_{jl}(\cdot)$ using kernel estimation to facilitate inference in the second step.

Based on (4), we collect all the estimated differenced $\mu(Z_{it};\gamma_0)$ and frontier, except $m_j(X_{it}^j)$, to be the dependent variable. Specifically, defining

$$\hat{Q}_{it,-j} = \Delta Y_{it} + \hat{m}_j(X_{it-1}^j) - \sum_{l=1,l\neq j}^{d} \Delta m_l(X_{it}^l) - \sum\sum_{1\leq j<l\leq d} \Delta\hat{H}_{jl}(X_{it}^j, x_{it}^l) + \Delta\mu(Z_{it};\hat{\gamma}),$$

and using $\hat{Q}_{it,-j}$ as the dependent variable, we obtain the backfitting estimator for $m_j(\cdot)$ with a local linear regression of $\hat{Q}_{it,-j}$ on X_{it}^j as $\tilde{m}_j(x^j) = \tilde{a}$, where for a kernel function $K(\cdot)$ and a bandwidth h,

$$(\tilde{a}, \tilde{b}) = argmin_{a,b} \sum_{i=1}^{n}\sum_{t=1}^{T} [\hat{Q}_{it,-j} - a - b(X_{it}^j - x^j)]^2 K(\frac{X_{it}^j - x^j}{h}). \quad (7)$$

To estimate $H_{jl}(\cdot)$, we define

$$\hat{Q}_{it,-jl} = \Delta Y_{it} + \hat{H}_{jl}(X_{it-1}^j, X_{it-1}^l) - \sum_{l=1}^{d} \Delta m_l(X_{it}^l) - \sum\sum_{1\leq j'<l'\leq d,(j',l')\neq(j,l)} \Delta\hat{H}_{j'l'}(X_{it}^{j'}, x_{it}^{l'}) + \Delta\mu(Z_{it};\hat{\gamma}),$$

and use $\hat{Q}_{it,-jl}$ as the dependent variable. Again, with a local linear regression, we obtain the backfitting estimator for $H_{jl}(x^j, x^l)$ as $\tilde{H}_{jl}(x^j, x^l) = \tilde{\alpha}$, where for a different bandwidth h_1,

$$(\tilde{\alpha}, \tilde{\beta}_1, \tilde{\beta}_2) = argmin_{\alpha,\beta_1,\beta_2} \sum_{i=1}^{n}\sum_{t=1}^{T} [\hat{Q}_{it,-jl} - \alpha - \beta_1(X_{it}^j - x^j) - \beta_2(X_{it}^l - x^l)]^2 K(\frac{X_{it}^j - x^j}{h_1}) K(\frac{X_{it}^l - x^l}{h_1}). \quad (8)$$

Note that in the first step, the use of series approximation in (4) is fairly convenient. Series estimates easily allow for imposing the additive and interaction structures in the frontier. Because the frontier is simply differenced, we just need to difference the series estimator to estimate it, avoiding the need for higher dimensional approximation if a marginal integration is used in the initial step. Our proposal of combining series in the first step and kernel in the second step also provides an easy alternative to the pure kernel method in Sperlich et al. (2002) to estimate the additive model with interactions. Our approach avoids the full dimensional kernel smoothing that proceeds the marginal integration in Sperlich et al. (2002).

3 Asymptotic charaterization

We start with assumptions that are used to prove the main results. The following notation is used. For any matrix A, define the norm $||A|| = [trace(A'A)]^{1/2}$. Below we denote a generic constant by C, the magnitude of which is inconsequential for the asymptotic analysis, and can vary from one place to another. Let's denote a generic function $\zeta(w) \in C^j$ if $\zeta(w)$ and all of its partial derivatives of order $\leq j$ are continuous and uniformly bounded on its support. $\zeta(w)$ satisfies the Cramer's condition if $E|\zeta(w)|^j \leq C^{j-2} j! E(\zeta^2(w)) < \infty$ for all $j \geq 2$.

Definition: a function $\zeta(x, y) : \Re^d \times \Re^d \to \Re$ is said to belong to an additive class of functions $\mathcal{G}$, or $\zeta \in \mathcal{G}$, if $\zeta(x, y) = \sum_{j=1}^d \zeta_j(x^j) + \sum\sum_{1 \leq j < l \leq d} \zeta_{jl}(x^j, x^l) - \sum_{j=1}^d \zeta_j(y^j) - \sum\sum_{1 \leq j < l \leq d} \zeta_{jl}(y^j, y^l)$, where $\zeta_j(\cdot), \zeta_{jl}(\cdot)$ are twice differentiable in the interior of their compact support, $E\zeta_j^2(\cdot) < \infty$ and $E\zeta_{jl}^2(\cdot) < \infty$.

For a function $\zeta(x, y)$, we use $E_{\mathcal{G}}(\zeta(x, y))$ to denote the projection of $\zeta(\cdot)$ onto the additive functional space $\mathcal{G}$ under the L_2 norm. $E_{\mathcal{G}}(\zeta(x, y))$ is in $\mathcal{G}$ and it is the closest function to $\zeta(x, y)$ among all functions in $\mathcal{G}$ in the sense that $E(\zeta(x, y) - E_{\mathcal{G}}(\zeta(x, y)))^2 = \inf_{\xi \in \mathcal{G}} E(\zeta(x, y) - \xi(x, y))^2$.

Assumptions:

A1 (1) $\{(Y_{it}, X_{it}, Z_{it}) : i = 1, \cdots, n, t = 1, \cdots, T\}$ is i.i.d. (identically and independently distributed) across $i = 1, \cdots, n$. T is finite. (2) The fixed effect is α_i and $E(\alpha_i) = 0$. (3) $v_{it} \sim i.i.d.(0, \sigma_v^2)$ across both i and t. (4) u_{it} is i.i.d. across $i = 1, \cdots, n$, with $E(u_{it}|\alpha_i, Z_{it}) = \mu(Z_{it}; \gamma_0) \equiv \mu_{it}(\gamma_0)$, $E(u_{it}^2|\alpha_i, Z_i) < C$ for all i, t and $Z_i = \{Z_{it}\}_{t=2}^T$. (5) For $\tilde{u}_{it} = u_{it} - \mu(Z_{it}; \gamma_0)$, $E(\tilde{u}_{it} - \tilde{u}_{it-1}|Z_{it}, Z_{it-1}) = 0$.

A2 (1) For Γ, a compact subset of $\Re^p$, γ_0 is contained in the interior of Γ. (2) For $\gamma \in \Gamma$, define $Q(\gamma) = \frac{1}{T}\sum_{t=1}^T E[\Delta\mu(Z_{it}; \gamma) - h(X_{it}, X_{it-1}; \gamma) - (\Delta\mu(Z_{it}; \gamma_0) - h(X_{it}, X_{it-1}; \gamma_0))]^2$, where $h(X_{it}, X_{it-1}; \gamma) = E_G(E(\Delta\mu(Z_{it}; \gamma)|X_{it}, X_{it-1}))$. $Q(\gamma) = 0$ only when $\gamma = \gamma_0$. (3) $\mu(Z_{it}; \gamma)$ is continuous in $\gamma \in \Gamma$ uniformly such that $|\mu(Z_{it}; \gamma) - \mu(Z_{it}; \gamma')| \leq B_\mu(Z_{it})||\gamma - \gamma'||$ with $EB_\mu^2(Z_{it}) < \infty$, $\forall\ \gamma, \gamma' \in \Gamma$. (4) $(\Delta\mu(Z_{it}; \gamma) - \Delta\mu(Z_{it}; \gamma_0))^2$, $\Delta\epsilon_{it}\Delta\mu(Z_{it}; \gamma)$, $\Delta_1\mu_{it}(\gamma)$, $(\Delta_1\mu_{it}(\gamma))^2$, $\Delta_1\mu_{it}(\gamma)\Delta_1\mu_{it'}(\gamma)$ and $\Delta_1\mu_{it}(\gamma)\Delta_1\mu_{it}(\gamma_0)$ satisfy the Cramer's condition, where $\Delta_1\mu_{it}(\gamma) = \Delta\mu(Z_{it}; \gamma) - h(X_{it}, X_{it-1}; \gamma)$. (5)

$E\sup_{\gamma\in\Gamma}|\mu(Z_{it};\gamma)|^2<\infty$.

A3 (1) $\exists$ some $\delta_i>0$, $i=1,2$, and B_h, a $S(\kappa)\times 1$ real vector, where $S(\kappa)=d\kappa+d_1\kappa^2$, such that

$\sup_{\gamma\in\Gamma}\sup_{X_{it}\in[-1,1]^d,X_{it-1}\in[-1,1]^d}|h(X_{it},X_{it-1};\gamma)-\Delta\Phi(X_{it})'B_h|=O(d\kappa^{-\delta_1}+d_1\kappa^{-2\delta_2})$.

(2) $\exists$ some $\delta_i>0$, $i=1,2$, B_{m_j}, a $\kappa\times 1$ real vector, and $B_{H_{jl}}$, a $\kappa^2\times 1$ real vector, such that

$$\sup_{x^j\in[-1,1]}|m_j(x^j)-(\phi^\kappa(x^j))'B_{m_j}|=O(\kappa^{-\delta_1}),\quad \sup_{x^j\in[-1,1],x^l\in[-1,1]}|H_{jl}(x^j,x^l)-(\phi^{\kappa^2}(x^j,x^l))'B_{H_{jl}}|=O(\kappa^{-2\delta_2}).$$

(3) $\sqrt{n}(d\kappa^{-\delta_1}+d_1\kappa^{-2\delta_2})\to 0$ as $n\to\infty$.

(4) For some nonstochastic sequence $\xi_{S(\kappa)}$, $\sup_{x^j\in[-1,1]}||\phi^\kappa(x^j)||\le\xi_{S(\kappa)}$, $\sup_{x^j\in[-1,1],x^l\in[-1,1]}||\phi^{\kappa^2}(x^j,x^l)||\le$ $\xi_{S(\kappa)}$ for $1\le j<l\le d$. Furthermore, as $n\to\infty$, $\xi^2_{S(\kappa)}S(\kappa)/n\to 0$, and $\xi^2_{S(\kappa)}(Lnn)^{1/2}/n\to 0$.

(5) The smallest eigenvalue of $E(\Delta\Phi(X_{it})\Delta\Phi(X_{it})')$ is bounded away from zero for all t.

B1 For $\partial_j\mu_{it}(\gamma)=\frac{\partial\mu(Z_{it};\gamma)}{\partial\gamma_j}$, $\partial_{jl}\mu_{it}(\gamma)=\frac{\partial^2\mu(Z_{it};\gamma)}{\partial\gamma_j\partial\gamma_l}$, $|\partial_j\mu_{it}(\gamma)-\partial_j\mu_{it}(\gamma')|\le B_{\partial_j\mu}(Z_{it})||\gamma-\gamma'||$, and $|\partial_{jl}\mu_{it}(\gamma)-\partial_{jl}\mu_{it}(\gamma')|\le B_{\partial_{jl}\mu}(Z_{it})||\gamma-\gamma'||$, for all γ, $\gamma'\in\Gamma$, $t=1,\cdots,T$, j or $l=1,\cdots,p$ with $EB^2_{\partial_j\mu}(Z_{it})<\infty$ and $EB^2_{\partial_{jl}\mu}(Z_{it})<\infty$.

B2 For $\partial_jh(X_{it},X_{it-1};\gamma)=E_G(E(\partial_j\Delta\mu_{it}(\gamma)|X_{it},X_{it-1}))$, $\partial_{jl}h(X_{it},X_{it-1};\gamma)=E_G(E(\partial_{jl}\Delta\mu_{it}(\gamma)|X_{it},X_{it-1}))$, where $\partial_j\Delta\mu_{it}(\gamma)=\frac{\partial\Delta\mu(Z_{it};\gamma)}{\partial\gamma_j}$, $\partial_{jl}\Delta\mu_{it}(\gamma)=\frac{\partial^2\Delta\mu(Z_{it};\gamma)}{\partial\gamma_j\partial\gamma_l}$, $\exists$ some $\delta_i>0$, $i=1,2$, B_{∂_jh} and $B_{\partial_{jl}h}$, both $S(\kappa)\times 1$ real vectors, such that

$\sup_{\gamma\in\Gamma}\sup_{X_{it}\in[-1,1]^d,X_{it-1}\in[-1,1]^d}|\partial_jh(X_{it},X_{it-1};\gamma)-\Delta\Phi(X_{it})'B_{\partial_jh}|=O(d\kappa^{-\delta_1}+d_1\kappa^{-2\delta_2})$ and

$\sup_{\gamma\in\Gamma}\sup_{X_{it}\in[-1,1]^d,X_{it-1}\in[-1,1]^d}|\partial_{jl}h(X_{it},X_{it-1};\gamma)-\Delta\Phi(X_{it})'B_{\partial_{jl}h}|=O(d\kappa^{-\delta_1}+d_1\kappa^{-2\delta_2})$.

B3 For all $t=2,\cdots,T$, $j,l\in\{1,\cdots,p\}$, $E(\partial_j\Delta\mu_{it}(\gamma)-\partial_jh(X_{it},X_{it-1};\gamma))(\partial_l\Delta\mu_{it}(\gamma)-\partial_lh(X_{it},X_{it-1};\gamma))$, $E(\partial_{jl}\Delta\mu_{it}(\gamma)-\partial_{jl}h(X_{it},X_{it-1};\gamma))^2$ and $E(\Delta_1\mu_{it}(\gamma)-\Delta_1\mu_{it}(\gamma_0))\partial_{jl}\Delta\mu_{it}(\gamma)$ are finite $\forall\ \gamma\in\Gamma$ and continuous at $\gamma_0\in\Gamma$.

B4 $\Sigma(\gamma_0)$ is a nonstochastic invertible matrix with its $(jl)th$ element as $\Sigma_{jl}(\gamma_0)=\frac{2}{T-1}\sum_{t=1}^{T}E(\partial_j\Delta\mu_{it}(\gamma_0)-\partial_jh(X_{it},X_{it-1};\gamma_0))(\partial_l\Delta\mu_{it}(\gamma_0)-\partial_lh(X_{it},X_{it-1};\gamma_0))$.

B5 $E\sup_{\gamma\in\Gamma}|\partial_j\mu_{it}(\gamma)|^2 < C$, $E\sup_{\gamma\in\Gamma}|\partial_{jl}\mu_{it}(\gamma)|^2 < C$, $E[\sup_{\gamma\in\Gamma}|\partial_j\theta(X_{it}, X_{it-1};\gamma)|]^2 < C$, and $E[\sup_{\gamma\in\Gamma}|\partial_{jl}\theta(X_{it}, X_{it-1};\gamma)|]^2 < C$, where $\partial_j\theta(X_{it}, X_{it-1};\gamma) = E_G(E(\partial_j\mu_{it}(\gamma)|X_{it}, X_{it-1}))$, and $\partial_{jl}\theta(X_{it}, X_{it-1};\gamma) = E_G(E(\partial_{jl}\mu_{it}(\gamma)|X_{it}, X_{it-1}))$.

B6 For all $j, l \in \{1, \cdots, p\}$, $t, t' \in \{2, \cdots, T\}$, $\partial_j\mu_{it}(\gamma)\partial_l\mu_{it'}(\gamma)$, $\partial_j\Delta_1\mu_{it}(\gamma)$, $\partial_j\Delta_1\mu_{it}(\gamma)\partial_j\Delta_1\mu_{it'}(\gamma)$, $\partial_j\theta(X_{it}, X_{it-1};\gamma)\partial_l\theta(X_{it}, X_{it-1};\gamma)$, $\partial_j h(X_{it}, X_{it-1};\gamma)\partial_l h(X_{it}, X_{it-1};\gamma)$; $(\partial_{jl}\mu_{it}(\gamma))^2$, $(\partial_{jl}\theta(X_{it}, X_{it-1};\gamma))^2$, $(\partial_{jl}h(X_{it}, X_{it-1};\gamma))^2$,$\partial_{jl}\Delta_1\mu_{it}(\gamma)$, $\partial_{jl}\Delta_1\mu_{it}(\gamma)\partial_{jl}\Delta_1\mu_{it'}(\gamma)$, and $\Delta\epsilon_{it}\partial_{jl}\Delta\mu_{it}(\gamma)$ satisfy the Cramer's condition, where $\partial_j\Delta_1\mu_{it}(\gamma) = \partial_j\Delta\mu_{it}(\gamma) - \partial_j h(X_{it}, X_{it-1};\gamma)$ and $\partial_{jl}\Delta_1\mu_{it}(\gamma) = \partial_{jl}\Delta\mu_{it}(\gamma) - \partial_{jl}h(X_{it}, X_{it-1};\gamma)$.

C1 The kernel function $K(\psi_j) : \Re \to \Re$ is symmetric and satisfies (1) $\left|K(\psi)\psi^j\right| \le C$ for all $\psi \in \Re$ with $j = 0, 1, \cdots, 3$; (2) $\int |\psi^j K(\psi)|d\psi \le C$ for $j = 0, 1, \cdots, 3$; (3) $\int K(\psi)d\psi = 1$, $\int \psi K(\psi)d\psi = 0$; (4) $K(\psi)$ is continuously differentiable on $\Re$ with $|\psi^j \frac{d}{d\psi}K(\psi)| \le C$ for all $\psi \in \Re$ and $j = 0, 1, \cdots, 3$.

C2 (1) $\forall\ j, l \in \{1, \cdots, d\}$, $j < l$, $m_j(\cdot) \in C^2$, and $H_{jl}(\cdot) \in C^2$.

(2) For some $\delta > 0$, $E|v_{it}|^{2+\delta} < C$, $E|u_{it}|^{2+\delta} < C$, and $E|\Delta\epsilon_{it}|^{2+\delta}|X_{it}^j, X_{it}^l) < C$.

(3) $E(\Delta\epsilon_{it}^2|X_{it}^j = x^j) = \sigma^2_{\Delta\epsilon_{it}}(x^j)$, $E(\Delta\epsilon_{it}^2|X_{it}^j = x^j, X_{it}^l = x^l) = \sigma^2_{\Delta\epsilon_{it}}(x^j, x^l)$ are continuous at x^j, x^l.

C3 (1) The conditional density of X_{it}^j, X_{it}^l given $\tilde{u}_{it}, \tilde{u}_{it-1}$ is $f_{X_t^j, X_t^l|\tilde{u}_t, \tilde{u}_{t-1}}(X_{it}^j, X_{it}^l) < C$.

(2) For G, a compact subset of $[-1, 1]$, $\inf_{X_{it}^j\in G, X_{it}^l\in G, X_{i\tau}^m\in G, X_{i\tau}^q\in G} f_{X_t^j, X_t^l, X_\tau^m, X_\tau^q}(X_{it}^j, X_{it}^l, X_{i\tau}^m, X_{i\tau}^q) > 0$, where $f_{X_t^j, X_t^l, X_\tau^m, X_\tau^q}(X_{it}^j, X_{it}^l, X_{i\tau}^m, X_{i\tau}^q)$ is the joint density of $X_{it}^j, X_{it}^l, X_{i\tau}^m, X_{i\tau}^q$, $\tau = t$ or $t-1$, $\forall$ $j < l$, $m < q$, $(j, l, m, q) \in \{1, \cdots, d\}$.

(3) Denote the joint density of X_{it} by $f_{X_t}(X_{it})$, and $f_{X_t}(X_{it}) \in C^2$.

(4) The joint density of X_{it}^j, X_{it}^l and X_{it-1} is $f_{X_t^j, X_t^l, X_{t-1}}(X_{it}^j, X_{it}^l, X_{it-1})$, and it is in C^2.

(5) The joint density of Z_{it} is $f_{Z_t}(Z_{it})$. The joint density of X_{it}^j, X_{it}^l and Z_{it-1} is $f_{X_t^j, X_t^l, Z_{t-1}}(X_{it}^j, X_{it}^l, Z_{it-1})$. Both are continuous.

C4 (1) $\phi_k(x^j) \in C^2$. (2) $h = O(n^{-1/5})$, $h_1 = O(n^{-1/6})$, $\xi^2_{S(\kappa)} = O(S(\kappa))$, and $\frac{\kappa^5}{n} = o(1)$.

(3) $B_\mu(Z_{it})$ is continuous.

The challenge of analyzing $\hat{\gamma}$ in (6) lies in the presence of $M_{\Delta\Phi}$ and nonlinearity of $\mu(Z_{it};\gamma)$. Assumption A facilitates the consistency argument for $\hat{\gamma}$. We focus in A1(1) on the panel data with a possible large n and fixed T, so appropriate handling of the fixed effects is important due to the incidental parameter problem. In A1(2), we allow the fixed effect α_i to be correlated with Z_{it} and its zero unconditional mean allows us to recover m_0. We assume in A1(4) that the conditional mean of inefficiency term u_{it} is known up to a vector of parameters γ_0. A1(4) and (5) allow us to treat (4) as a proper regression model after differencing, and A1(5) is satisfied if $E(u_{it}|Z_{it}, Z_{i\tau}) = \mu(Z_{it};\gamma_0)$ for $\tau = t-1$ or $t+1$. The consistency of $\hat{\gamma}$ calls for A2(2), which is necessary due to the potential nonlinearity of $\mu(Z_{it};\gamma_0)$ in γ_0, and is similar in spirit to the classical assumption D of Theorem 4.3.1 in the consistency of nonlinear least squares in Amemiya (1985). We further require uniform continuity of $\mu(Z_{it};\gamma)$ in A2(3) and uniform boundedness on the second moment of its projection onto the additive space G in A2(5). The Cramer's condition on centered $\mu(Z_{it};\gamma)$ in A2(4) allows us to apply uniform convergence argument in Lemma 6. Assumptions A3(1) and (2) insure that the errors in the series approximations to $m_j(\cdot)$'s, $H_{jl}(\cdot)$'s, and $h(X_{it}, X_{it-1};\gamma)$ converge to zero sufficiently fast as $\kappa \to \infty$. A3(4) bounds the magnitudes of the basis functions, and together with A3(3), they state the rates at which $\kappa \to \infty$ as $n \to \infty$ such that the asymptotic bias and variance generated in estimating the frontier are sufficiently small to achieve the $\sqrt{n}$ convergence rate of $\hat{\gamma}$. A3(5) insures the nonsingularity of the covariance matrix of the asymptotic form of the differenced frontier estimator.

Assumption B gives the conditions to obtain asymptotic normality of $\hat{\gamma}$, which is based on the Taylor's expansion of the criterion function in (6). As can be expected, we need the first and second order derivatives of $\mu_{it}(\gamma)$ with respect to γ to be smooth in B1. We need their projections onto the additive space G to be well approximated by the basis function in B2. We further assume that the derivatives and their projections to be continuous, uniformly bounded in γ and satisfy the Cramer's condition in B3, B5 and B6. B4 insures the nonsingularity of the asymptotic covariance matrix of $\hat{\gamma}$.

Assumptions A and B are sufficient for the uniform convergence of the first stage series estimator for the frontier, as we detail in Theorem 2 below. For the second stage backfitting estimation, we need further Assumption C. C1 gives conditions on kernel functions, which are frequently used in kernel-based estimation, see Martins-Filho et al. (2018). Specifically, a Gaussian kernel can be used in the backfitting. C2 restricts m_j and H_{jl} to be sufficiently smooth for Taylor expansion to be performed, conditional moment of differenced ϵ_{it} and moment of v_{it} and u_{it} of order slightly larger than 2 are bounded, which allow us to check the Liapunov'c condition in the central limit theorem. C3 assumes that the density or conditional density of X_{it}, X_{it} and X_{it-1} are smooth. C4 requires further that the basis function is smooth and proper orders for κ and bandwidth h and h_1 need to be satisfied. We note that the rate optimal bandwidths for regression can be used. However, for the backfitting estimators' asymptotic normality statement, we need to adopt κ in the first stage to go to infinity at a rate slower than the κ that maximizes the L_2 rate of convergence of a series estimator for m_j, i.e., $\kappa \propto n^{1/5}$. Thus, C4 requires the first stage estimator to be oversmoothed. We note that this assumption is different from Horowitz and Mammen (2004) and Ozabaci et al. (2014), and the difference arises from the fact that we need to properly handle the left hand side variable in our backfitting step , which contains one variable that is also being smoothed on the right hand side.

Due to the differencing, we do not provide an estimate for m_0[2]. Since the shape of the frontier and ranking of firms' inefficiency are not impacted by m_0, below we only focus on the asymptotic characterization of $m_j(\cdot)$, $H_{jl}(\cdot)$ and γ.

The nonlinear nature of $\mu(z;\gamma)$ together with the presence of fixed effects in a semiparmetric frontier estimated with series makes the characterization of $\hat{\gamma}$ in (6) a bit involved, which is given in Theorem 1.

Theorem 1. *(a) With Assumption A, we have* $\hat{\gamma} \xrightarrow{p} \gamma_0$.

(b) With Assumptions A and B, we have $\sqrt{n}(\hat{\gamma}-\gamma_0) \xrightarrow{d} N(0, \Sigma(\gamma_0)^{-1}\Omega_T\Sigma(\gamma_0)^{-1})$.

Here, we outline the main steps involved for consistency for illustration. From (6), we let $L_N(\gamma) = \frac{1}{N}(\Delta Y + \Delta\mu(\gamma))' M_{\Delta\Phi}(\Delta Y + \Delta\mu(\gamma))$, $\Delta g = \{\Delta g(X_{it})\}_{t=2}^{t},{}_{i=1}^{n}$, $\Delta\epsilon = \{\Delta\epsilon_{it}\}_{t=2}^{t},{}_{i=1}^{n}$. Then $\Delta Y = \Delta g - \Delta\mu(\gamma_0) + \Delta\epsilon$, and

$$\begin{aligned} L_N(\gamma) = \;& \tfrac{1}{N}(\Delta\mu(\gamma)-\Delta\mu(\gamma_0))' M_{\Delta\Phi}(\Delta\mu(\gamma)-\Delta\mu(\gamma_0)) + \tfrac{2}{N}(\Delta\mu(\gamma)-\Delta\mu(\gamma_0))' M_{\Delta\Phi}(\Delta g + \Delta\epsilon) \\ & + \tfrac{1}{N}(\Delta g + \Delta\epsilon)' M_{\Delta\Phi}(\Delta g + \Delta\epsilon) \\ = \;& L_{1N}(\gamma) + L_{2N}(\gamma) + L_{3N}. \end{aligned}$$

We can show that $\sup_{\gamma\in\Gamma} L_{1N}(\gamma) = Q(\gamma) + o_p(1)$, $\sup_{\gamma\in\Gamma} L_{2N}(\gamma) = o_p(1)$ and $L_{3N} = \frac{1}{T-1}\sum_{t=1}^{T} E(\Delta\epsilon_{it})^2 +$

[2] With an additional assumption that the fixed effect has a zero mean, we can easily consider a consistent estimator for m_0.

$o_p(1)$. Letting $L_0(\gamma) = \frac{1}{T-1}\sum_{t=1}^{T} E(\Delta\epsilon_{it})^2 + Q(\gamma)$, we have $\sup_{\gamma\in\Gamma}|L_N(\gamma) - L_0(\gamma)| = o_p(1)$. Given Assumption A2(2), $L_0(\gamma)$ is minimized at $\gamma = \gamma_0$. The standard argument for consistency of extreme estimators (i.e., Theorem 4.1.1 in Amemiya (1985)) leads to $\hat{\gamma} \xrightarrow{p} \gamma_0$, which we detail in the proof.

Based on Theorem 1, Theorem 2 gives the asymptotic behavior of the first-stage series estimators for the frontier components.

Theorem 2. *Let Assumption A hold. Then*

(a) $||\hat{\theta}(\hat{\gamma}) - \theta|| = O_p((\frac{S(\kappa)}{n})^{1/2} + d\kappa^{-\delta_1} + d_1\kappa^{-2\delta_2})$.

(b) $\sup_{x^j\in[-1,1]}|\hat{m}_j(x^j) - m_j(x^j)| = O_p(\xi_{S(\kappa)}[(\frac{S(\kappa)}{n})^{1/2} + d\kappa^{-\delta_1} + d_1\kappa^{-2\delta_2}])$.

(c) $\sup_{x^j\in[-1,1],x^l\in[-1,1]}|\hat{H}_{jl}(x^j, x^l) - H_{jl}(x^j, x^l)| = O_p(\xi_{S(\kappa)}[(\frac{S(\kappa)}{n})^{1/2} + d\kappa^{-\delta_1} + d_1\kappa^{-2\delta_2}])$.

Theorem 3 gives the asymptotic properties of the second-stage estimator for the frontier components.

Theorem 3. *Let Assumptions A-C hold. Then*

(a) $\sqrt{nh}(\tilde{m}_j(x^j) - m_j(x^j) - \frac{h^2}{2}m_j^{(2)}(x^j)\mu_{k,2} + o_p(h^2)) \xrightarrow{d} N(0, \sigma^2_{m_j}(x^j))$.

(b) For $l \neq j$, $\sqrt{nh}(\tilde{m}_j(x^j) - m_j(x^j) - \frac{h^2}{2}m_j^{(2)}(x^j)\mu_{k,2} + o_p(h^2))$ *and* $\sqrt{nh}(\tilde{m}_l(x^l) - m_l(x^l) - \frac{h^2}{2}m_l^{(2)}(x^l)\mu_{k,2} + o_p(h^2))$ *are asymptotically independently normally distributed.*

(c) $\sqrt{nh_1^2}(\tilde{H}_{jl}(x^j, x^l) - H_{jl}(x^j, x^l) - \frac{h_1^2}{2}\mu_{k,2}(H^{(2)}_{jl,j}(x^j, x^l) + H^{(2)}_{jl,l}(x^j, x^l)) + o_p(h_1^2)) \xrightarrow{d} N(0, \sigma^2_{H_{jl}}(x^j, x^l))$.

(d) For $(j, l) \neq (m, p)$, $\sqrt{nh_1^2}(\tilde{H}_{jl}(x^j, x^l) - H_{jl}(x^j, x^l) - \frac{h_1^2}{2}\mu_{k,2}(H^{(2)}_{jl,j}(x^j, x^l) + H^{(2)}_{jl,l}(x^j, x^l)) + o_p(h_1^2))$ *and* $\sqrt{nh_1^2}(\tilde{H}_{mp}(x^m, x^p) - H_{mp}(x^m, x^p) - \frac{h_1^2}{2}\mu_{k,2}(H^{(2)}_{mp,m}(x^m, x^p) + H^{(2)}_{mp,p}(x^m, x^p)) + o_p(h_1^2))$ *are asymptotically independently normally distributed.*

Theorem 3 part 1 implies that asymptotically, $\sqrt{nh}(\tilde{m}_j(x^j) - m_j(x^j) - \frac{h^2}{2}m_j^{(2)}(x^j)\mu_{k,2} + o_p(h^2))$ is not affected by random sampling errors in the first stage. The second stage estimator of $m_j(x^j)$ actually has the same asymptotic distribution that it would have if all $m_l(\cdot)$'s for $l \neq j$ and all $H_{jl}(\cdot)$'s were known and the local linear estimation were used to estimate $m_j(x^j)$ directly. It is in this sense that our estimator $\tilde{m}_j(x^j)$

has an oracle property. Parts 1 and 2 of Theorem 3 imply that the estimators of $m_1(x^1), \cdots, m_d(x^d)$ are asymptotically independently distributed. Similarly, part 3 shows that our estimator $\tilde{H}_{jl}(x^j, x^l)$ enjoys the oracle property. Parts 3 and 4 of Theorem 3 imply that the estimators of $H_{12}(x^1, x^2), \cdots, H_{d-1d}(x^d, x^{d-1})$ are asymptotically independently distributed.

4 Monte Carlo Simulation

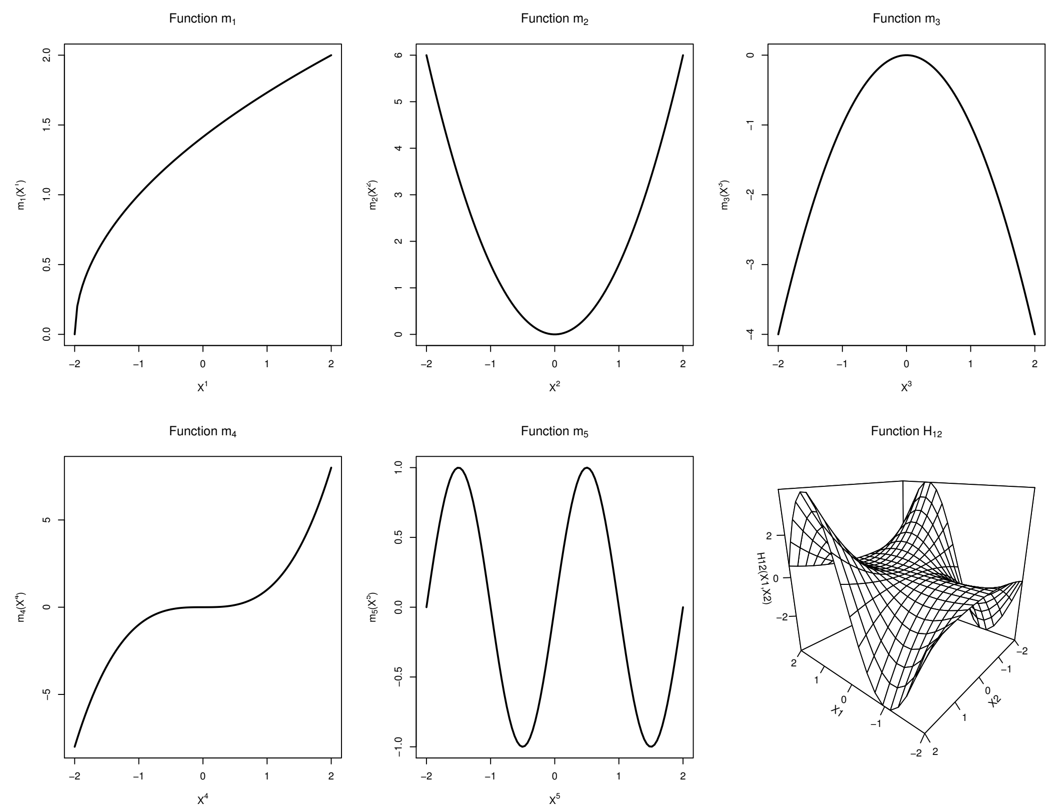
In this section, we investigate the finite sample performance of our estimators of parameters in inefficiency mean function and functions in frontier. To be more specific, we evaluate 1) the consistency of the proposed estimators $\hat{\gamma}$ in (6) and $\hat{\beta}(\hat{\gamma})$ in (5) with different choice of n and fixed T; 2) the effect of increase in dimension (d) on the performance of estimators; and 3) the impact of different specifications in inefficiency term u_{it} on the parameter, functions, and frontier estimation. We consider the following data generating processes (DGPs):

$$y_{it} = \alpha_i + m_0 + \sum_{j=1}^{d} m_j(X_{it}^j) + \sum_{1 \leq j < l \leq d}\sum H_{jl}(X_{it}^j, X_{it}^l) + v_{it} - u_{it}, \tag{9}$$

where for $i = 1, .., n$ and $t = 1, ..., T$, we set $d = 3$ and $d = 5$ to generate a total of 6 and 15 smooth functions of $m_j(\cdot)$ and $H_{jl}(\cdot, \cdot)$, respectively. In DGP_1, we set $m_1(v) = \sqrt{v+2}$, $m_2(v) = \frac{3}{2}v^2$, $m_3(v) = -v^2$, $m_4(v) = v^3$, $m_5(v) = \sin(v\pi)$, and $H_{jl}(v^j, v^l) = sin(1.5v^j)(v^l - 0.5(v^l)^3)$ for all $1 \leq j < l \leq d$; In DGP_2, $m_1(v) = \exp(v) - 1$, $m_2(v) = -v^2$, $m_3(v) = 1.5v^3$, $m_4(v) = 2v + 1$, $m_5(v) = cos(v) - 1$, and $H_{jl}(v^j, v^l) = 0.5(v^j v^l)$ for all $1 \leq j < l \leq d$. To reflect possible correlations across variables as a common feature in practice, we generate $\{X_{it}^j\}_{j=1}^d$ from a multivariate normal distribution $N(\mu_0, \Sigma_0)$, where $\mu_0 = [4, 6, 2]'$ with $d = 3$ and $\mu_0 = [4, 6, 2, 0, -2]'$ with $d = 5$, and Σ_0 is the covariance matrix with its $(j, l)^{th}$ element $\Sigma_{jl} = 0.5^{|j-l|}$ for $j, l = 1, ..., d$. We set $\delta = 0.5$ to allow regressors to be fairly correlated, and rescale each variable into $[-2, 2]$ to satisfy our assumption on the compact support of each variable X^j.

To provide a clear picture of smooth functions in our simulations, Figure 1 plots the simulated functions $\{m_j(\cdot)\}_{j=1}^5$ and $H_{12}(\cdot, \cdot)$ in DGP_1. The DGP_1 displays clear nonlinearities of inputs and environment variables, as well as their interactive effects that enters the frontier model beyond a simple linear fashion. DGP_2 retains simple linear structure in functions m_4 and product of variables in H_{12}, both of which are

Figure 1: Functions Plot in DGP_1



commonly specified in applications.

We consider three specifications of the inefficiency term $u_{it} > 0$: 1) (Scaling) we generate u_{it} with a scaling property $u_{it} = (\sqrt{2\sigma_u^2/\pi})^{-1} u_i^* \mu_1(Z_{it};\gamma_0)$, where $u_i^* \sim |N(0,\sigma_u^2)|$ follows a half-normal distribution with $E(u_i^*|z_{it}) = \sqrt{2\sigma_u^2/\pi}$, and $\mu_1(Z_{it};\gamma_0) = \exp(Z_{it}'\gamma_0)$; 2) (Probit) we introduce a log of probit function in $u_{it} = \mu_2(Z_{it},\gamma_0)$, where $\mu_2(Z_{it},\gamma_0) = -\ln(\Phi(Z_{it}'\gamma_0))$; and 3) (Log-normal) we specify u_{it} as a log-normal random variable $u_{it} \sim logN(0,\sigma^2(Z_{it}'\gamma_0))$, with $\mu_3(Z_{it};\gamma_0) = \exp(0.5Z_{it}'\gamma_0)$. In the case of Scaling, we set $\sigma_u^2 = 1$. In all three specifications, we draw $v_{it} \sim N(0,\sigma_v^2)$ with $\sigma_v^2 = 0.5$, and specify $Z_{it}'\gamma_0 = W_{it}\gamma_{01} + X_{it}^1\gamma_{02}$ in $\mu_1(\cdot) - \mu_3(\cdot)$, where $W_{it} \sim U(0,1)$ represents an exogenous environment variable not appearing on the frontier, and $(\gamma_{01},\gamma_{02}) = (1.5,-0.5)$. We construct the fixed effect $\alpha_i = \frac{1}{T}\sum_{t=1}^{T} c_0(\sum_{j=1}^{d} X_{it}^j + W_{it}) + \xi_i$, where $\xi_i \sim N(0,1)$ and $c_0 \neq 0$ reflects a fixed effect model. We set $c_0 = 1$, and empirically center α_i to be consistent with our assumption $E(\alpha_i) = 0$. Finally, we employ cubic B-spline estimator in (6), and 2nd-order Gaussian kernel function in the last step of backfitting. We impose the identification condition by empirically centering estimated functions around zero.

Throughout the experiment, we choose n from $(100,200,400)$, fix $T = 3$ or $T = 6$, and perform $R = 1000$ repetitions. To evaluate the performance of parameters $\hat{\gamma} = (\hat{\gamma}_1,\hat{\gamma}_2)$, for $s = 1,2$ we report the root mean squared error (RMSE) $\frac{1}{R}\sum_{r=1}^{R}\sqrt{(\hat{\gamma}_{s,r} - \gamma_{0s})^2}$; the bias (BIAS) $\frac{1}{R}\sum_{r=1}^{R}(\hat{\gamma}_{s,r} - \gamma_{0s})$; and the empirical standard deviation (STD) $sd(\{\hat{\gamma}_{s,r}\}_{r=1}^{R})$. To evaluate the performance of function estimates $(\hat{m}_j(\cdot),\hat{H}_{jl}(\cdot,\cdot))$, for each function, say $\hat{m}_j(\cdot)$, we report the root averaged MSE (RAMSE) $\frac{1}{R}\sum_{r=1}^{R}\left[\frac{1}{nT}\sum_{i=1}^{n}\sum_{t=1}^{T}\left(\hat{m}_{j,r}(X_{it}^j) - m_j(X_{it}^j)\right)^2\right]^{\frac{1}{2}}$;
the averaged BIAS (ABIAS) $\frac{1}{R}\sum_{r=1}^{R}\left[\frac{1}{nT}\sum_{i=1}^{n}\sum_{t=1}^{T}\left(\hat{m}_{j,r}(X_{it}^j) - m_j(X_{it}^j)\right)\right]$;
and averaged STD (ASTD) $\frac{1}{nT}\sum_{i=1}^{n}\sum_{t=1}^{T} sd(\hat{m}_{j,1}(X_{it}^j),...,\hat{m}_{j,R}(X_{it}^j))$.

We first report the simulation results of $\hat{\gamma}$ in Table 1, where *Panel A* displays the results for DGP_{1-2} with $T = 3$, and *Panel B* with $T = 6$. For comparison purpose, in each panel we compare the performance of our estimator in (6), denoted as FD, with the pool estimator that ignores the presence of fixed effect α_i, denoted as Pool.[3] In each panel, we record the measures of estimator's performance in the case of $d = 3$ and

[3]In this case, the Pool estimator of γ_0 is obtained by replacing $\Delta\Phi$ with Φ in (6). Similarly, the Pool estimator of functions replaces $\Delta\Phi$ with Φ in (5), which is then backfitted through kernel estimator.

Table 1: Simulation Results of Parameters Estimates $\hat{\gamma}$ ($c_0 = 1$)

Panel A	$T = 3$	$n = 100$		200		400	
DGP_1	**FD**	$d = 3$	$d = 5$	$d = 3$	$d = 5$	$d = 3$	$d = 5$
$\hat{\gamma}_1$	RMSE	0.0568	0.0641	0.0276	0.0299	0.0169	0.0172
	ABIAS	0.0038	-0.0022	0.0047	0.0018	0.0008	0.0029
	ASD	0.0570	0.0644	0.0274	0.0292	0.0170	0.0170
$\hat{\gamma}_2$	RMSE	0.0368	0.0403	0.0180	0.0190	0.0113	0.0112
	ABIAS	0.0021	-0.0026	0.0035	0.0049	0.0003	0.0013
	ASD	0.0369	0.0404	0.0177	0.0185	0.0114	0.0112
	Pool						
$\hat{\gamma}_1$	RMSE	0.0744	0.1031	0.0840	0.0989	0.0359	0.1468
	ABIAS	-0.0457	-0.0402	-0.0559	-0.0336	-0.0270	-0.0657
	ASD	0.0590	0.0954	0.0405	0.0487	0.0238	0.0401
$\hat{\gamma}_2$	RMSE	0.0651	0.0788	0.0620	0.0643	0.0637	0.0679
	ABIAS	-0.0515	-0.0501	-0.0440	-0.0624	-0.0377	-0.0628
	ASD	0.0401	0.0611	0.0279	0.0341	0.0155	0.0217
DGP_2	**FD**						
$\hat{\gamma}_1$	RMSE	0.0495	0.0672	0.0273	0.0303	0.0171	0.0196
	ABIAS	0.0042	0.0029	0.0019	-0.0011	0.0003	0.0052
	ASD	0.0496	0.0675	0.0274	0.0304	0.0171	0.0190
$\hat{\gamma}_2$	RMSE	0.0316	0.0419	0.0179	0.0206	0.0114	0.0125
	ABIAS	0.0025	0.0038	0.0008	-0.0009	0.0003	0.0029
	ASD	0.0317	0.0420	0.0179	0.0207	0.0114	0.0123
	Pool						
$\hat{\gamma}_1$	RMSE	0.1115	0.1232	0.0874	0.1134	0.0706	0.1143
	ABIAS	-0.0886	-0.0787	-0.0770	-0.0754	-0.0651	-0.0767
	ASD	0.0680	0.0953	0.0417	0.0521	0.0275	0.0304
$\hat{\gamma}_2$	RMSE	0.0917	0.0976	0.0792	0.0852	0.0674	0.0923
	ABIAS	-0.0796	-0.0748	-0.0736	-0.0665	-0.0647	-0.0688
	ASD	0.0458	0.0631	0.0294	0.0353	0.0191	0.0206
Panel B	$T = 6$	$n = 100$		200		400	
DGP_1	**FD**	$d = 3$	$d = 5$	$d = 3$	$d = 5$	$d = 3$	$d = 5$
$\hat{\gamma}_1$	RMSE	0.0287	0.0277	0.0198	0.0177	0.0141	0.0106
	ABIAS	0.0048	0.0029	0.0013	-0.0005	-0.0020	0.0000
	ASD	0.0284	0.0277	0.0198	0.0178	0.0140	0.0106
$\hat{\gamma}_2$	RMSE	0.0185	0.0176	0.0129	0.0115	0.0089	0.0068
	ABIAS	0.0022	0.0024	0.0012	-0.0002	-0.0011	-0.0002
	ASD	0.0185	0.0175	0.0129	0.0115	0.0089	0.0068
	Pool						
$\hat{\gamma}_1$	RMSE	0.0487	0.0452	0.0410	0.0386	0.0363	0.0411
	ABIAS	-0.0338	-0.0179	-0.0319	-0.0171	-0.0327	-0.0405
	ASD	0.0352	0.0417	0.0259	0.0238	0.0157	0.0162
$\hat{\gamma}_2$	RMSE	0.0434	0.0344	0.0363	0.0457	0.0356	0.0509
	ABIAS	-0.0355	-0.0215	-0.0321	-0.0253	-0.0339	-0.0376
	ASD	0.0250	0.0270	0.0170	0.0159	0.0109	0.0113
DGP_2	**FD**						
$\hat{\gamma}_1$	RMSE	0.0305	0.0287	0.0194	0.0174	0.0117	0.0114
	ABIAS	-0.0020	-0.0096	0.0002	-0.0007	-0.0007	0.0009
	ASD	0.0306	0.0271	0.0195	0.0175	0.0118	0.0115
$\hat{\gamma}_2$	RMSE	0.0193	0.0189	0.0126	0.0114	0.0079	0.0072
	ABIAS	-0.0007	-0.0059	0.0005	-0.0004	-0.0005	0.0004
	ASD	0.0194	0.0180	0.0126	0.0115	0.0079	0.0072
	Pool						
$\hat{\gamma}_1$	RMSE	0.0504	0.0618	0.0419	0.0665	0.0345	0.0462
	ABIAS	-0.0365	-0.0433	-0.0342	-0.0488	-0.0311	-0.0430
	ASD	0.0350	0.0443	0.0242	0.0258	0.0151	0.0151
$\hat{\gamma}_2$	RMSE	0.0424	0.0479	0.0384	0.0495	0.0345	0.0441
	ABIAS	-0.0357	-0.0383	-0.0347	-0.0395	-0.0330	-0.0355
	ASD	0.0231	0.0289	0.0165	0.0174	0.0101	0.0103

$d = 5$. As sample size n doubles, we observe a clear improvement in FD estimator $\hat{\gamma}$ in terms of a declining trend in RMSE, BIAS, and STD toward zero. This observation holds across different $DGPs$ and dimensions. The FD estimator also disclose well performance regardless of either n doubles with T fixed, or T doubles with n fixed, indicating the consistency of $\hat{\gamma}$. By contrast, the Pool estimator, ignoring the unobserved cross-sectional heterogeneities, exhibits all three measures uniformly larger in magnitude than those of the FD estimator, and is inconsistent particularly in the direction of n (e.g., $\hat{\gamma}_2$ in DGP_1 with $T = 3$).

We next evaluate the performance of kernel-backfitting estimator for functions in Table 2 for DGP_1, and in Table 3 for DGP_2. In each table, we report the results with $n = (100, 200, 400)$ when $T = 3$ in *Panel A*, and $T = 6$ in *Panel B*. For brevity, we report the results of the first two functions $\hat{m}_1(\cdot)$ and $\hat{m}_1(\cdot)$, and one interaction function $\hat{H}_{12}(\cdot, \cdot)$. We then report the overall performance of function estimator by averaging RAMSE, ABIAS, and ASTD over all function estimators for FD (Average: FE) and Pool (Average: Pool). We further evaluate the performance of the frontier estimation $\hat{m}_0 + \sum_{j=1}^{d} \hat{m}_j(X_{it}^j) + \sum\sum_{1 \leq j < l \leq d} \hat{H}_{jl}(X_{it}^j, X_{it}^l)$ by FD (Frontier: FD) and Pool estimator (Frontier: Pool). We observe a qualitatively similar pattern of the performance FD and Pool estimator to those in Table 1. Across two $DGPs$, the FD estimator is consistent as indicated by the decreasing RAMSE, ABIAS, and ASTD as n rises, which holds true with either $T = 3$ or $T = 6$. We observe that the FD estimator exhibits a slightly poor performance with $d = 5$ compared to $d = 3$, although the discrepancies diminishes quickly as sample size increases. The FD estimator outperforms its Pool counterpart in terms of the estimation of function on average and on the frontier. By contrast, the Pool estimator is inconsistent as expected. We provide a clear picture on the estimation of interaction function H_{12} through kernel-backfitting in Figure 2, which plots the function estimates $\hat{H}_{12}$ (kernel-FD, gray surface with dotted line) against the true function (true, white surface with solid line) in DGP_1 and DGP_2, respectively, with $(n, T) = (400, 6)$. The results show that the approximation on the interaction function is performed fairly well through our FD estimator, and is robust to the number of functions considered, consistent with our arguments that the *curse-of-dimensionality* problem is invariant to the dimension of X.

Finally, Table 4 investigates the impact of different specifications in u_{it} on FD estimator. We report the results estimation on two parameters $(\hat{\gamma}_1, \hat{\gamma}_2)$, three functions $(\hat{m}_1, \hat{m}_2, \hat{H}_{12})$, and the frontier. In each

Table 2: Simulation Results of Function Estimates: A Fixed Effect Model in DGP_1 ($c_0 = 1$)

Panel A	$T = 3$	$n = 100$		200		400	
		$d = 3$	$d = 5$	$d = 3$	$d = 5$	$d = 3$	$d = 5$
$\hat{m}_1$	RAMSE	0.5393	0.8421	0.3707	0.3770	0.2145	0.2362
	ABIAS	0.0583	-0.0101	-0.0179	-0.0097	-0.0187	0.0017
	ASD	0.4198	0.6321	0.2873	0.2908	0.1760	0.1938
m2	RAMSE	0.4352	0.5965	0.2815	0.3097	0.2081	0.2138
	ABIAS	-0.0353	-0.0285	-0.0213	-0.0189	-0.0296	-0.0233
	ASD	0.3312	0.4810	0.2076	0.2384	0.1499	0.1552
$\hat{h}_{12}$	RAMSE	0.5125	0.7234	0.3144	0.3603	0.2328	0.2357
	ABIAS	0.0151	-0.0712	-0.0168	-0.0073	-0.0097	-0.0017
	ASD	0.3908	0.5406	0.2477	0.2790	0.1917	0.1944
Average: FE	RAMSE	0.5057	0.7195	0.3154	0.3426	0.2240	0.2359
	ABIAS	-0.0075	-0.0322	-0.0189	-0.0175	-0.0141	-0.0104
	ASD	0.3814	0.5416	0.2398	0.2566	0.1730	0.1790
Average: Pool	RAMSE	0.6837	1.2254	0.4689	1.0634	0.3667	1.0931
	ABIAS	-0.0616	-0.0219	-0.0477	-0.0991	-0.0434	-0.1197
	ASD	0.4811	0.6866	0.3151	0.4043	0.2273	0.4072
Frontier: FD	RAMSE	0.6374	1.4127	0.5010	0.7892	0.4004	0.5861
	ABIAS	-0.0575	-0.2495	-0.0661	-0.1699	-0.0670	-0.1340
	ASD	0.4494	1.1012	0.3377	0.6430	0.2700	0.4693
Frontier: Pool	RAMSE	4.1232	4.0996	4.1881	4.4006	4.2582	4.6649
	ABIAS	-3.9374	-3.3611	-4.0563	-3.8495	-4.1351	-4.2034
	ASD	0.6716	1.2693	0.4232	0.9162	0.3629	0.6542

Panel B	$T = 6$	$n = 100$		200		400	
		$d = 3$	$d = 5$	$d = 3$	$d = 5$	$d = 3$	$d = 5$
$\hat{m}_1$	RAMSE	0.3824	0.3430	0.2632	0.2407	0.1869	0.1504
	ABIAS	0.0032	0.0295	-0.0011	-0.0139	-0.0038	-0.0135
	ASD	0.3049	0.2722	0.2148	0.1950	0.1490	0.1252
$\hat{m}_2$	RAMSE	0.3006	0.3104	0.2316	0.2046	0.1583	0.1489
	ABIAS	-0.0262	-0.0106	-0.0370	-0.0354	-0.0039	-0.0029
	ASD	0.2305	0.2347	0.1687	0.1481	0.1095	0.1080
$\hat{h}_{12}$	RAMSE	0.3364	0.7234	0.2410	0.3603	0.1685	0.2357
	ABIAS	0.0031	-0.0712	-0.0106	-0.0073	-0.0066	-0.0017
	ASD	0.2709	0.5406	0.1978	0.2790	0.1420	0.1944
Average: FE	RAMSE	0.3429	0.3287	0.2423	0.2235	0.1687	0.1614
	ABIAS	-0.0104	-0.0097	-0.0143	-0.0126	-0.0027	-0.0075
	ASD	0.2650	0.2499	0.1893	0.1683	0.1307	0.1201
Average: Pool	RAMSE	0.4133	0.9949	0.4157	0.9928	0.4414	1.0195
	ABIAS	-0.0338	-0.0860	-0.0342	-0.1135	-0.0274	-0.1294
	ASD	0.3086	0.3469	0.2969	0.2398	0.2963	0.1761
Frontier: FD	RAMSE	1.0262	0.6820	0.6899	0.5159	0.5306	0.3894
	ABIAS	-0.1914	-0.1516	-0.1069	-0.1266	-0.0871	-0.0745
	ASD	0.7135	0.5612	0.4716	0.4133	0.3575	0.3237
Frontier: Pool	RAMSE	4.0807	4.0151	4.3112	4.2603	4.3283	4.4047
	ABIAS	-3.7122	-3.7974	-4.0375	-4.0651	-4.0969	-4.2212
	ASD	0.9293	0.7145	0.6752	0.5714	0.4919	0.4741

Table 3: Simulation Results of Function Estimates: A Fixed Effect Model in DGP_2 ($c_0 = 1$)

Panel A	$T = 3$	$n = 100$		200		400	
		$d = 3$	$d = 5$	$d = 3$	$d = 5$	$d = 3$	$d = 5$
$\hat{m}_1$	RAMSE	0.5934	0.8431	0.3381	0.3875	0.2283	0.2576
	ABIAS	-0.0075	-0.0220	0.0001	0.0116	-0.0070	0.0291
	ASD	0.4533	0.6370	0.2665	0.3005	0.1843	0.2002
$\hat{m}_2$	RAMSE	0.4236	0.6062	0.2426	0.2577	0.1712	0.1836
	ABIAS	0.0950	0.0580	0.0263	0.0226	0.0158	0.0169
	ASD	0.3447	0.5136	0.2128	0.2245	0.1488	0.1609
$\hat{h}_{12}$	RAMSE	0.5870	0.7825	0.3973	0.3674	0.2692	0.2791
	ABIAS	-0.0411	0.0097	0.0096	-0.0027	-0.0077	0.0002
	ASD	0.4350	0.5651	0.2681	0.2662	0.1932	0.1950
Average: FE	RAMSE	0.6049	0.8068	0.4016	0.4101	0.2934	0.3024
	ABIAS	0.0197	0.0191	0.0249	0.0189	0.0157	0.0182
	ASD	0.4092	0.5621	0.2548	0.2641	0.1791	0.1853
Average: Pool	RAMSE	0.7319	1.0922	0.7286	1.0875	0.7424	1.0709
	ABIAS	-0.0151	-0.0047	-0.0162	0.0186	-0.0197	0.0556
	ASD	0.4893	0.6631	0.3202	0.4005	0.3211	0.2745
Frontier: FD	RAMSE	0.7855	1.4139	0.5911	0.8674	0.4814	0.6014
	ABIAS	0.1198	0.1884	0.0764	0.1306	0.1077	0.1006
	ASD	0.4599	1.0929	0.3685	0.6750	0.2755	0.4806
Frontier: Pool	RAMSE	2.6286	3.0419	2.7065	2.9958	2.7461	3.1649
	ABIAS	2.4275	1.9130	2.5603	2.2436	2.6267	2.5067
	ASD	0.5742	1.2518	0.4214	0.8205	0.3274	0.6718
Panel B	$T = 6$	$n = 100$		200		400	
		$d = 3$	$d = 5$	$d = 3$	$d = 5$	$d = 3$	$d = 5$
$\hat{m}_1$	RAMSE	0.3968	0.3766	0.2605	0.2359	0.1733	0.1704
	ABIAS	-0.0217	-0.0538	0.0124	-0.0026	0.0013	0.0285
	ASD	0.3039	0.2785	0.1963	0.1842	0.1393	0.1297
$\hat{m}_2$	RAMSE	0.2481	0.2518	0.1796	0.1784	0.1335	0.1236
	ABIAS	0.0105	0.0435	0.0195	0.0148	0.0050	0.0232
	ASD	0.2145	0.2151	0.1526	0.1582	0.1201	0.1061
$\hat{h}_{12}$	RAMSE	0.3667	0.3586	0.2666	0.2487	0.2093	0.1863
	ABIAS	0.0042	0.0170	0.0090	0.0108	-0.0193	-0.0090
	ASD	0.2524	0.2559	0.1989	0.1793	0.1520	0.1354
Average: FE	RAMSE	0.4092	0.3921	0.2947	0.2799	0.2268	0.2094
	ABIAS	0.0119	0.0139	0.0223	0.0142	0.0080	0.0146
	ASD	0.2645	0.2509	0.1854	0.1722	0.1407	0.1245
Average: Pool	RAMSE	0.5006	0.8061	0.4705	0.7617	0.4965	0.8140
	ABIAS	-0.0021	0.0470	-0.0100	0.0577	-0.0149	0.0716
	ASD	0.3268	0.3410	0.2214	0.2359	0.2046	0.1712
Frontier: FD	RAMSE	1.2001	0.7719	0.8197	0.5563	0.6544	0.4486
	ABIAS	0.1815	0.1736	0.1174	0.1222	0.0962	0.1245
	ASD	0.7896	0.6208	0.5019	0.4417	0.4139	0.3531
Frontier: Pool	RAMSE	2.7687	2.7505	2.6708	2.7699	2.7951	2.8639
	ABIAS	2.2934	2.4290	2.3434	2.5317	2.5150	2.6681
	ASD	0.8787	0.8143	0.6514	0.5851	0.4885	0.4257

Figure 2: Functions Plot in DGP_1 (upper panel) and DGP_2 (lower panel)

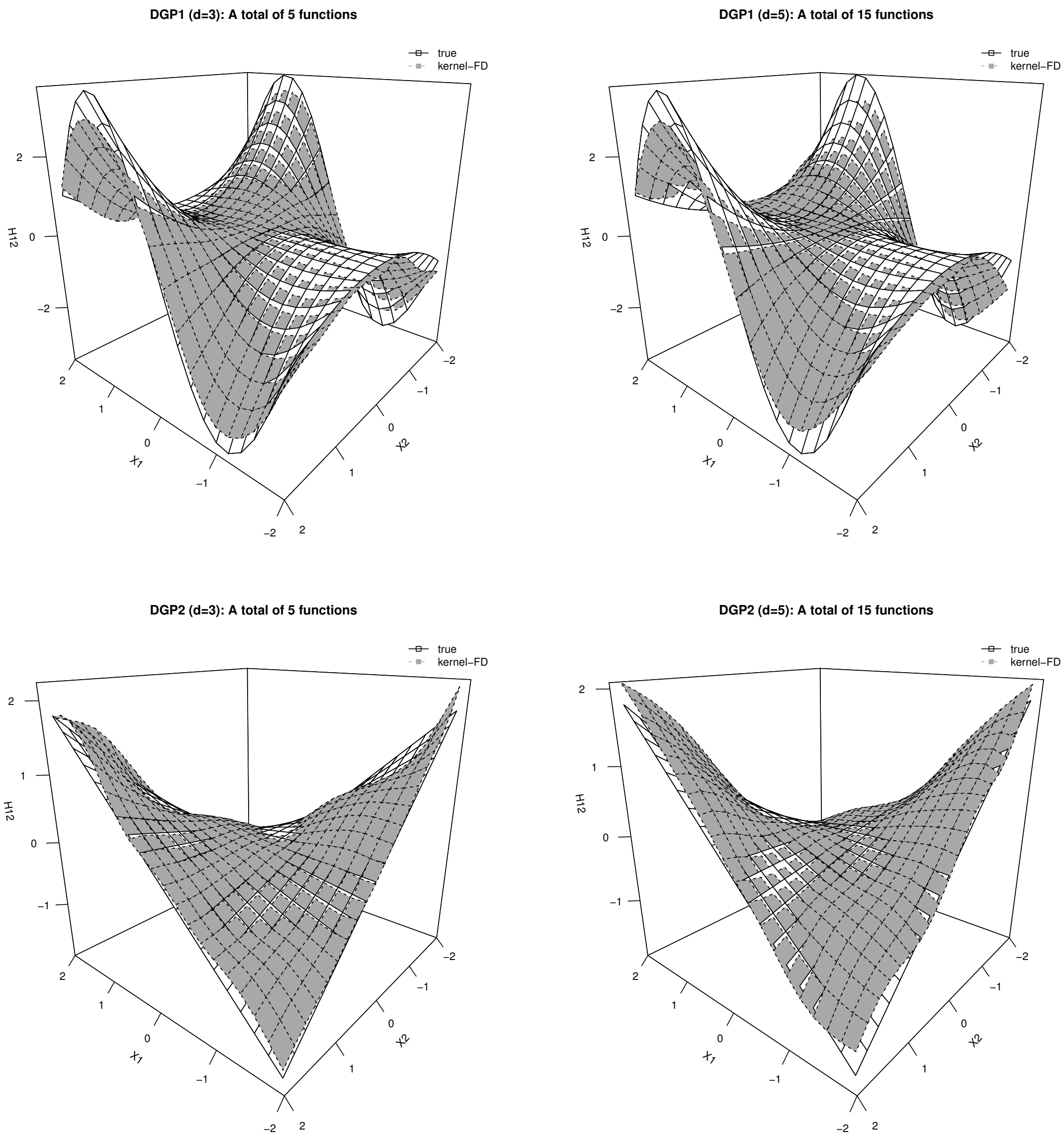


Table 4: Results of Estimation with Fixed Effects: Different Distribution of Inefficiency in DGP_1 ($d = 5$)

Panel A	$T = 3$		$n = 100$			200			400	
		Scaling	Probit	Log-normal	Scaling	Probit	Log-normal	Scaling	Probit	Log-normal
$\hat{\gamma}_1$	RMSE	0.0793	0.0716	0.0777	0.0298	0.0328	0.0316	0.0194	0.0174	0.0168
	BIAS	0.0019	0.0030	0.0084	-0.0016	-0.0028	-0.0038	0.0011	0.0013	0.0008
	STD	0.0797	0.0719	0.0777	0.0295	0.0328	0.0315	0.0192	0.0175	0.0168
$\hat{\gamma}_2$	RMSE	0.0474	0.0482	0.0505	0.0203	0.0212	0.0208	0.0124	0.0117	0.0112
	BIAS	0.0005	0.0007	0.0025	-0.0004	-0.0004	-0.0021	0.0005	0.0004	0.0003
	STD	0.0477	0.0484	0.0507	0.0200	0.0212	0.0207	0.0123	0.0118	0.0112
$\hat{m}_1$	RAMSE	0.9855	0.8227	0.9028	0.3836	0.4293	0.4078	0.2403	0.2391	0.2170
	ABIAS	0.0198	0.0223	0.0510	-0.0159	-0.0152	-0.0464	0.0057	-0.0117	-0.0027
	ASTD	0.7175	0.6277	0.6874	0.3014	0.3286	0.3122	0.1957	0.1977	0.1779
$\hat{m}_2$	RAMSE	0.6240	0.6099	0.6393	0.2960	0.2916	0.3156	0.2245	0.2207	0.2176
	ABIAS	-0.0834	-0.0993	-0.0807	-0.0342	-0.0124	-0.0501	-0.0331	-0.0104	-0.0225
	ASTD	0.4831	0.4704	0.5236	0.2258	0.2292	0.2298	0.1552	0.1613	0.1556
$\hat{H}_{12}$	RAMSE	0.7494	0.7474	0.7895	0.3637	0.3497	0.3476	0.2396	0.2281	0.2397
	ABIAS	-0.0163	0.0234	-0.0430	0.0070	-0.0093	0.0173	0.0019	-0.0073	-0.0092
	ASTD	0.5613	0.5478	0.6037	0.2833	0.2754	0.2750	0.1981	0.1856	0.1993
Frontier-FD	RAMSE	1.4066	1.3613	0.6571	0.7974	0.7734	0.6017	0.6059	0.5989	0.4911
	ABIAS	-0.1726	-0.1110	-0.1298	-0.2144	-0.1447	-0.1546	-0.1099	-0.1204	-0.1366
	ASTD	1.1323	1.0853	0.5950	0.6244	0.6345	0.5006	0.5035	0.4759	0.4633
Panel B	$T = 6$		$n = 100$			200			400	
		Scaling	Probit	Log-normal	Scaling	Probit	Log-normal	Scaling	Probit	Log-normal
$\hat{\gamma}_1$	RMSE	0.0261	0.0263	0.0293	0.0158	0.0180	0.0168	0.0097	0.0106	0.0108
	BIAS	-0.0022	0.0008	-0.0030	-0.0003	-0.0012	-0.0018	0.0009	0.0015	0.0006
	STD	0.0262	0.0264	0.0293	0.0159	0.0180	0.0168	0.0097	0.0106	0.0108
$\hat{\gamma}_2$	RMSE	0.0175	0.0172	0.0186	0.0101	0.0122	0.0105	0.0067	0.0072	0.0074
	BIAS	-0.0014	0.0006	-0.0020	-0.0001	-0.0007	-0.0010	0.0006	0.0004	0.0001
	STD	0.0176	0.0173	0.0186	0.0102	0.0123	0.0105	0.0068	0.0072	0.0075
$\hat{m}_1$	RAMSE	0.3483	0.3444	0.3519	0.2345	0.2398	0.2273	0.1478	0.1438	0.1610
	ABIAS	-0.0287	0.0075	-0.0410	-0.0024	-0.0201	-0.0263	0.0031	-0.0096	0.0004
	ASTD	0.2769	0.2730	0.2720	0.1912	0.1923	0.1806	0.1241	0.1226	0.1354
$\hat{m}_2$	RAMSE	0.3057	0.2667	0.2993	0.1959	0.2134	0.2022	0.1538	0.1501	0.1493
	ABIAS	-0.0361	-0.0357	-0.0537	-0.0155	-0.0075	-0.0181	-0.0066	-0.0061	-0.0168
	ASTD	0.2274	0.1980	0.2305	0.1478	0.1556	0.1461	0.1136	0.1096	0.1053
$\hat{H}_{12}$	RAMSE	0.3106	0.3070	0.3131	0.2193	0.2216	0.2092	0.1612	0.1568	0.1583
	ABIAS	-0.0091	-0.0222	0.0184	-0.0056	-0.0296	-0.0063	-0.0132	-0.0039	-0.0013
	ASTD	0.2462	0.2440	0.2459	0.1802	0.1802	0.1748	0.1369	0.1354	0.1336
Frontier-FD	RAMSE	0.7060	0.6904	0.7462	0.5188	0.5364	0.5366	0.3907	0.4039	0.3952
	ABIAS	-0.1824	-0.1235	-0.1403	-0.1367	-0.1364	-0.1511	-0.0842	-0.1363	-0.1188
	ASTD	0.5609	0.5707	0.6077	0.4120	0.4270	0.4249	0.3265	0.3233	0.3197

panel, we report the results with three specifications in u_{it}, Scaling, Probit, and log-normal, against $n = 100, 200, 400$. Regardless of the three cases considered, the FD estimator works fairly well and demonstrates its consistency with large n or T. Overall, our FD estimator demonstrates its appealing performance on estimating unknown parameters and functions in the frontier, which is robust to the dimension of variables and different specifications of inefficiency term.

5 Conclusion

In this paper, we propose a semiparametric additive stochastic frontier model with three features that adds to the literature. First, we model the frontier flexibly as an additive unknown functions of transitional inputs, environment variables, as well as their possible interactions, all of which are nonparametrically estimated to capture the potential nonlinearities of interests. Second, we eschew the distribution assumption on composite error terms, and allow the conditional mean function of inefficiency to be influenced by relevant determinants that may or may not appear on the frontier model. Finally, we adopt a panel data structure by separating out the time-invariant fixed effect from inefficiency. We estimate the model through series estimator combined with nonlinear least square estimator, and further employ one-step kernel backfitting estimator to facilitate inference. Our estimator does not suffer from the incidental parameter problem, and the *curse of dimensionality* does not link to the dimension of input or environment variables. Thus, our model can be potentially applied in various applications where the variables in frontier and inefficiency are of high dimension with panel structure. Under mild conditions, we show the consistency of our proposed estimator. We also demonstrate through simulation studies the finite sample performance of estimators, which perform reasonably well with fairly correlated regressors. The simulation results remain robust to the dimension of variables in the frontier and different specifications of inefficiency term.

Appendix

Theorem 1.

Proof. (a) Define $E_{\mathcal{G}}(E(\Delta\mu(\gamma)|X)) = \{h(X_{it}, X_{it-1};\gamma)\}_{t=2,}^{T}{}_{i=1}^{n}$, and we write

$$\begin{aligned} L_{1N}(\gamma) = & \frac{1}{N}[\Delta\mu(\gamma) - \Delta\mu(\gamma_0) - E_{\mathcal{G}}(E(\Delta\mu(\gamma) - \Delta\mu(\gamma_0)|X))]' M_{\Delta\Phi} \\ & \times[\Delta\mu(\gamma) - \Delta\mu(\gamma_0) - E_{\mathcal{G}}(E(\Delta\mu(\gamma) - \Delta\mu(\gamma_0)|X))] \\ & + \frac{2}{N}[\Delta\mu(\gamma) - \Delta\mu(\gamma_0) - E_{\mathcal{G}}(E(\Delta\mu(\gamma) - \Delta\mu(\gamma_0)|X))]' M_{\Delta\Phi} E_{\mathcal{G}}(E(\Delta\mu(\gamma) - \Delta\mu(\gamma_0)|X)) \\ & + \frac{1}{N} E_{\mathcal{G}}(E(\Delta\mu(\gamma) - \Delta\mu(\gamma_0)|X))' M_{\Delta\Phi} E_{\mathcal{G}}(E(\Delta\mu(\gamma) - \Delta\mu(\gamma_0)|X)) \\ = & L_{11N}(\gamma) + L_{12N}(\gamma) + L_{13N}(\gamma). \end{aligned}$$

We claim that the following results hold:

(1) $\sup_{\gamma\in\Gamma} |L_{11N}(\gamma) - Q(\gamma)| = o_p(1)$.

(2) $\sup_{\gamma\in\Gamma} L_{13N}(\gamma) = o_p(1)$.

(3) $\sup_{\gamma\in\Gamma} L_{2N}(\gamma) = o_p(1)$.

(4) $L_{3N} = \frac{1}{N}\Delta\epsilon'\Delta\epsilon + o_p(1) = \frac{1}{T-1}\sum_{t=1}^{T} E(\Delta\epsilon_{it})^2 + o_p(1)$.

With the Cauchy-Schawartz inequality and the fact that $M_{\Delta\Phi}$ is symmetric and idemptotent, we have $L_{12N}(\gamma) \leq 2(L_{11N}(\gamma))^{1/2}(L_{13N}(\gamma))^{1/2}$. Using (1) and (2) above, we obtain $L_{1N} = Q(\gamma) + o_p(1)$ uniformly in $\gamma \in \Gamma$. Together with (3) and (4), we have $\sup_{\gamma\in\Gamma} |L_N(\gamma) - L_0(\gamma)| = o_p(1)$. Given Assumption A2(2), $L_0(\gamma)$ is minimized at $\gamma = \gamma_0$. We argue that $\hat{\gamma} \xrightarrow{p} \gamma_0$.

By the definition of $\hat{\gamma}$ in (6), we have

$$L_N(\hat{\gamma}) \leq L_N(\gamma_0). \tag{10}$$

Let $\mathcal{N}$ be an open neighborhood in $\Re^p$ containing γ_0. Then $\mathcal{N}^c \cap \Gamma$ is compact, where $\mathcal{N}^c$ is the complement of $\mathcal{N}$ in $\Re^p$ and Γ compact. So $\min_{\gamma\in\mathcal{N}^c\cap\Gamma} L_0(\gamma)$ exists by A2(1) and (3).

Because $L_0(\gamma)$ is globally minimized at $\gamma_0 \in \Gamma$, we define $\varepsilon = \min_{\gamma\in\mathcal{N}^c\cap\Gamma} L_0(\gamma) - L_0(\gamma_0) > 0$. Define the event ω as $\{\omega : |L_N(\gamma) - L_0(\gamma)| < \frac{\varepsilon}{2}, , \forall\gamma \in \Gamma\}$. Then the event ω implies $|L_N(\hat{\gamma}) - L_0(\hat{\gamma})| < \frac{\varepsilon}{2}$ and thus $L_0(\hat{\gamma}) < L_N(\hat{\gamma}) + \frac{\varepsilon}{2}$. By (10), $L_0(\hat{\gamma}) < L_N(\hat{\gamma}) + \frac{\varepsilon}{2}$. The event ω also implies $|L_N(\gamma_0) - L_0(\gamma_0)| < \frac{\varepsilon}{2}$, and thus, $L_N(\gamma_0) < L_0(\gamma_0) + \frac{\varepsilon}{2}$.

So adding them up, the event ω implies that $L_0(\hat{\gamma}) < L_0(\gamma_0) + \varepsilon = \min_{\gamma\in\mathcal{N}^c\cap\Gamma} L_0(\gamma)$. So event ω implies that $\hat{\gamma} \in \mathcal{N}$, and thus $P(\omega) \leq P(\hat{\gamma} \in \mathcal{N})$. Since by $\sup_{\gamma\in\Gamma} |L_N(\gamma) - L_0(\gamma)| = o_p(1)$, $\lim_{N\to\infty} P(\omega) = 1$, and thus $\lim_{N\to\infty} P(\hat{\gamma} \in \mathcal{N}) = 1$. With $\mathcal{N}$ being any neighborhood of γ_0, we conclude that $\hat{\gamma} \xrightarrow{p} \gamma_0$. We only need to show the claims (1)-(4) above.

(2)

$$\begin{aligned} L_{13N}(\gamma) = & \ \tfrac{1}{N}E_{\mathcal{G}}(E(\Delta\mu(\gamma)|X))'M_{\Delta\Phi}E_{\mathcal{G}}(E(\Delta\mu(\gamma)|X)) - \tfrac{2}{N}E_{\mathcal{G}}(E(\Delta\mu(\gamma_0)|X))'M_{\Delta\Phi}E_{\mathcal{G}}(E(\Delta\mu(\gamma)|X)) \\ & + \tfrac{1}{N}E_{\mathcal{G}}(E(\Delta\mu(\gamma_0)|X))'M_{\Delta\Phi}E_{\mathcal{G}}(E(\Delta\mu(\gamma_0)|X)) \\ = & \ L_{131N}(\gamma) + L_{132N}(\gamma) + L_{133N}. \end{aligned}$$

In Lemma 3, we show that $\sup_{\gamma\in\Gamma} L_{131N}(\gamma) = O_p((d\kappa^{-\delta_1} + d_1\kappa^{-2\delta_2})^2) = o_p(1)$. Similarly, $L_{133N} = o_p(1)$.

By the Cauchy-Schwartz inequality, $|L_{132N}(\gamma)| \leq 2|L_{131N}(\gamma)|^{1/2}|L_{133N}|^{1/2}$. So $\sup_{\gamma\in\Gamma} L_{13N}(\gamma) = o_p(1)$.

(1) $L_{11N}(\gamma)$

$$\begin{aligned} = & \ \tfrac{1}{N}[\Delta\mu(\gamma) - \Delta\mu(\gamma_0) - E_{\mathcal{G}}(E(\Delta\mu(\gamma) - \Delta\mu(\gamma_0)|X))]'[\Delta\mu(\gamma) - \Delta\mu(\gamma_0) - E_{\mathcal{G}}(E(\Delta\mu(\gamma) - \Delta\mu(\gamma_0)|X))] \\ & - \tfrac{1}{N}[\Delta\mu(\gamma) - \Delta\mu(\gamma_0) - E_{\mathcal{G}}(E(\Delta\mu(\gamma) - \Delta\mu(\gamma_0)|X))]'P_{\Delta\Phi}[\Delta\mu(\gamma) - \Delta\mu(\gamma_0) - E_{\mathcal{G}}(E(\Delta\mu(\gamma) - \Delta\mu(\gamma_0)|X))] \\ = & \ L_{111N}(\gamma) - L_{112N}(\gamma). \end{aligned}$$

We apply Lemma 4(2) to obtain that $\sup_{\gamma\in\Gamma} L_{112N}(\gamma) = o_p(1)$.

We show below that $\sup_{\gamma\in\Gamma}|L_{112N}(\gamma)-Q(\gamma)|=o_p(1)$. Recall the definition of $E_{\mathcal{G}}(E(\Delta\mu(\gamma)|X))$,

$$\begin{aligned}
&\sup_{\gamma\in\Gamma}|L_{112N}(\gamma)-Q(\gamma)| \\
\leq\ & \tfrac{1}{T-1}\textstyle\sum_{T=2}^{T}\sup_{\gamma\in\Gamma}|\tfrac{1}{n}\sum_{i=1}^{n}[\Delta\mu(Z_{it};\gamma)-h(X_{it},X_{it-1};\gamma)-(\Delta\mu(Z_{it};\gamma_0)-h(X_{it},X_{it-1};\gamma_0))]^2 \\
&-E[\Delta\mu(Z_{it};\gamma)-h(X_{it},X_{it-1};\gamma)-(\Delta\mu(Z_{it};\gamma_0)-h(X_{it},X_{it-1};\gamma_0))]^2| \\
=\ & \tfrac{1}{T-1}\textstyle\sum_{T=2}^{T}\sup_{\gamma\in\Gamma}|\tfrac{1}{n}\sum_{i=1}^{n}[\Delta_1\mu_{it}(\gamma)-\Delta_1\mu_{it}(\gamma_0)]^2-E[\Delta_1\mu_{it}(\gamma)-\Delta_1\mu_{it}(\gamma_0)]^2|,
\end{aligned}$$

by the definition of $\Delta_1\mu_{it}(\gamma)$ in A2. Observe that $[\Delta_1\mu_{it}(\gamma)-\Delta_1\mu_{it}(\gamma_0)]^2=(\Delta_1\mu_{it}(\gamma))^2-2\Delta_1\mu_{it}(\gamma)\Delta_1\mu_{it}(\gamma_0)+(\Delta_1\mu_{it}(\gamma_0))^2$. We easily obtain that $\frac{1}{n}\sum_{i=1}^{n}(\Delta_1\mu_{it}(\gamma_0))^2=E(\Delta_1\mu_{it}(\gamma_0))^2=o_p(1)$ by the weak law of large numbers. The claim $\sup_{\gamma\in\Gamma}|L_{112N}(\gamma)-Q(\gamma)|=o_p(1)$ follows the results (A) and (B) below.

(A) We apply Lemma 5 to show that $\sup_{\gamma\in\Gamma}|\frac{1}{n}\sum_{i=1}^{n}(\Delta_1\mu_{it}(\gamma))^2-E(\Delta_1\mu_{it}(\gamma))^2|=o_p(1)$.

(i) For all $\gamma\neq\gamma'$,

$$\begin{aligned}
&|(\Delta_1\mu_{it}(\gamma))^2-(\Delta_1\mu_{it}(\gamma'))^2 \\
\leq\ & |\Delta\mu(Z_{it};\gamma)-\Delta\mu(Z_{it};\gamma')-(h(X_{it},X_{it-1};\gamma)-h(X_{it},X_{it-1};\gamma_0))| \\
&\times|\Delta\mu(Z_{it};\gamma)+\Delta\mu(Z_{it};\gamma')-(h(X_{it},X_{it-1};\gamma)+h(X_{it},X_{it-1};\gamma_0))| \\
\leq\ & ||\gamma-\gamma'||[B_\mu(Z_{it})+B_\mu(Z_{it-1})+E_{\mathcal{G}}(B_\mu(Z_{it})|X_{it},X_{it-1})+E_{\mathcal{G}}(B_\mu(Z_{it-1})|X_{it},X_{it-1})] \\
&\times|\Delta\mu(Z_{it};\gamma)+\Delta\mu(Z_{it};\gamma')-(h(X_{it},X_{it-1};\gamma)+h(X_{it},X_{it-1};\gamma_0))|.
\end{aligned}$$

Given that $EB_\mu^2(Z_{it})<C$ in A2(3), $E(u_{it}^2|\alpha_1 Z_2)<C$ in A1(4), and $E[\sup_{\gamma\in\Gamma}|E_G(E(\mu(Z_{it};\gamma)|X_{it},X_{it-1}))|]^2<\infty$ in A2(5), we have $|(\Delta_1\mu_{it}(\gamma))^2-(\Delta_1\mu_{it}(\gamma'))^2\leq||\gamma-\gamma'||B_1(Z_{it},Z_{it-1})$ for some $B_1(Z_{it},Z_{it-1})$ such that $EB_1(Z_{it},Z_{it-1})<C$.

(ii) Clearly, $(\Delta_1\mu_{it}(\gamma))^2$ is i.i.d. across i. (iii) $(\Delta_1\mu_{it}(\gamma))^2$ satisfies the Cramer's condition by A2(4). So all three conditions in Lemma 5 are satisfied and we conclude that $\sup_{\gamma\in\Gamma}|\frac{1}{n}\sum_{i=1}^{n}(\Delta_1\mu_{it}(\gamma))^2-E(\Delta_1\mu_{it}(\gamma))^2|=o_p(1)$.

(B) We claim further that $\sup_{\gamma\in\Gamma}|\frac{1}{n}\sum_{i=1}^{n}\Delta_1\mu_{it}(\gamma)\Delta_1\mu_{it}(\gamma_0)-E(\Delta_1\mu_{it}(\gamma)\Delta_1\mu_{it}(\gamma_0))|=o_p(1)$.

(i) For all $\gamma\neq\gamma'$,

$$\begin{aligned}
&|\Delta_1\mu_{it}(\gamma)\Delta_1\mu_{it}(\gamma_0)-\Delta_1\mu_{it}(\gamma')\Delta_1\mu_{it}(\gamma_0)| \\
\leq\ & |\Delta\mu(Z_{it};\gamma)-\Delta\mu(Z_{it};\gamma')-(h(X_{it},X_{it-1};\gamma)-h(X_{it},X_{it-1};\gamma_0))| \\
&\times|\Delta\mu(Z_{it};\gamma_0)-(h(X_{it},X_{it-1};\gamma_0)|| \\
\leq\ & ||\gamma-\gamma_0||B_2(Z_{it},Z_{it-1}),
\end{aligned}$$

where $EB_2(Z_{it},Z_{it-1})<C$ with similar arguments by assumptions A1(4),A2(3), and A2(5).

(ii) Clearly, $\Delta_1\mu_{it}(\gamma)\Delta_1\mu_{it}(\gamma_0)$ is i.i.d. across i. (iii) $\Delta_1\mu_{it}(\gamma)\Delta_1\mu_{it}(\gamma_0)$ satisfies the Cramer's condition by A2(4). So we apply Lemma 5 to obtain that $\sup_{\gamma\in\Gamma}|\frac{1}{n}\sum_{i=1}^{n}\Delta_1\mu_{it}(\gamma)\Delta_1\mu_{it}(\gamma_0)-E(\Delta_1\mu_{it}(\gamma)\Delta_1\mu_{it}(\gamma_0))|=o_p(1)$.

(3) We write

$$\begin{aligned} L_{2N}(\gamma) = &\ \tfrac{2}{N}(\Delta\mu(\gamma)-\Delta\mu(\gamma_0))'(I-P_{\Delta\Phi})\Delta g + \tfrac{2}{N}(\Delta\mu(\gamma)-\Delta\mu(\gamma_0))'\Delta\epsilon - \tfrac{2}{N}(\Delta\mu(\gamma)-\Delta\mu(\gamma_0))'P_{\Delta\Phi}\Delta\epsilon \\ = &\ L_{21N}(\gamma)+L_{22N}(\gamma)-L_{23N}(\gamma). \end{aligned}$$

$L_{21N}(\gamma) \le 2(L_{1N}(\gamma))^{\frac{1}{2}}(L_{31N})^{\frac{1}{2}}$. By results (1) and (2) above, $\sup_{\gamma\in\Gamma}|L_{1N}(\gamma)-Q(\gamma)| = o_p(1)$, so $\sup_{\gamma\in\Gamma} L_{1N}(\gamma) = O_p(1)$. As shown in (4), $L_{31N} = O_p((d\kappa^{-\delta_1}+d_1\kappa^{2\delta_2})^2)$. So we obtain $\sup_{\gamma\in\Gamma} L_{21N}(\gamma) = o_p(1)$.

$L_{23N}(\gamma)$: we observe that $\frac{1}{N}\Delta\epsilon' P_{\Delta\Phi}\Delta\epsilon = o_p(1)$ by Lemma 4(1), then by Cauchy-Schwartz inequality, $L_{23N}(\gamma) \le 2[\frac{1}{N}(\Delta\mu(\gamma)-\Delta\mu(\gamma_0))'P_{\Delta\Phi}(\Delta\mu(\gamma)-\Delta\mu(\gamma_0))]^{\frac{1}{2}}[\frac{1}{N}\Delta\epsilon' P_{\Delta\Phi}\Delta\epsilon]^{\frac{1}{2}} = 2[L_{231N}(\gamma)]^{\frac{1}{2}}o_p(1)$. Defining $L_{232N}(\gamma) = \frac{1}{N}(\Delta\mu(\gamma)-\Delta\mu(\gamma_0))'(\Delta\mu(\gamma)-\Delta\mu(\gamma_0))$ and $L_{2320}(\gamma) = \frac{1}{T-1}\sum_{t=2}^{T} E(\Delta\mu(Z_{it};\gamma)-\Delta\mu(Z_{it};\gamma_0))^2$, we claim that $\sup_{\gamma\in\Gamma}|L_{232N}(\gamma)-L_{2320}(\gamma)| = o_p(1)$. Note that $L_{1N}(\gamma) = L_{232N}(\gamma)-L_{231N}(\gamma) = Q(\gamma)+o_p(1)$ uniformly over $\gamma\in\Gamma$. The claimed result leads to $L_{231N}(\gamma) = L_{2320}(\gamma)-Q(\gamma)+o_p(1) = O_p(1)$ uniformly over $\gamma\in\Gamma$. So we show the claim below. $L_{232N}(\gamma) = \frac{1}{n}\sum_{i=1}^{n}\frac{1}{T-1}\sum_{t=2}^{T}(\Delta\mu(Z_{it};\gamma)-\Delta\mu(Z_{it};\gamma_0))^2 = \frac{1}{n}\sum_{i=1}^{n}\frac{1}{T-1}\sum_{t=2}^{T}G(Z_{it},Z_{it-1};\gamma)$. Then $\sup_{\gamma\in\Gamma}|L_{232N}(\gamma)-L_{2320}(\gamma)| \le \frac{1}{T-1}\sum_{t=2}^{T}\sup_{\gamma\in\Gamma}|\frac{1}{n}\sum_{i=1}^{n}G(Z_{it},Z_{it-1};\gamma) - EG(Z_{it},Z_{it-1};\gamma)|$. (i) For all $\gamma\neq\gamma'$, by A2(3),

$$\begin{aligned} & |G(Z_{it},Z_{it-1};\gamma)-G(Z_{it},Z_{it-1};\gamma')| \\ = & |(\Delta\mu(Z_{it};\gamma)-\Delta\mu(Z_{it};\gamma'))(\Delta\mu(Z_{it};\gamma)+\Delta\mu(Z_{it};\gamma')-2\Delta\mu(Z_{it};\gamma_0))| \\ \le & ||\gamma-\gamma'||[B_\mu(Z_{it})+B_\mu(Z_{it-1})]|\Delta\mu(Z_{it};\gamma)+\Delta\mu(Z_{it};\gamma')-2\Delta\mu(Z_{it};\gamma_0)| = ||\gamma-\gamma'||B_3(Z_{it},Z_{it-1};\gamma). \end{aligned}$$

With assumption A2(3), and A2(5), we have $E\sup_{\gamma\in\Gamma}B_3(Z_{it},Z_{it-1};\gamma) < C$. (ii) (Z_{it},Z_{it-1}) is i.i.d. across i. (iii) The Cramer's condition on $G(Z_{it},Z_{it-1};\gamma)$ is assumed in A2(4). So we apply Lemma 5 to obtain the claim that $\sup_{\gamma\in\Gamma}|L_{232N}(\gamma)-L_{2320}(\gamma)| = o_p(1)$. It implies that $\sup_{\gamma\in\Gamma}L_{23N}(\gamma) = o_p(1)$.

$L_{22N}(\gamma) = \frac{2}{N}\sum_{i=1}^{n}\sum_{t=2}^{T}\Delta\mu(Z_{it};\gamma)\Delta\epsilon_{it} - \frac{2}{N}\sum_{i=1}^{n}\sum_{t=2}^{T}\Delta\mu(Z_{it};\gamma_0)\Delta\epsilon_{it} = L_{221N}(\gamma)-L_{222N}$. For L_{222N} : by assumptions A1(3) and (5), we have $E(\Delta\epsilon_{it}|Z_{it},Z_{it-1}) = E(v_{it}-\tilde{u}_{it}-(v_{it-1}-\tilde{u}_{it-1})|Z_{it},Z_{it-1}) = 0$. So by the law of iterated expectation, $EL_{222N} = 0$. $V(L_{222N}) = \frac{4}{n(T-1)^2}\sum_{t=1}^{T}\sum_{\tau=1}^{T}E(\Delta\epsilon_{it}\Delta\epsilon_{i\tau}\Delta\mu(Z_{it};\gamma_0)\Delta\mu(Z_{i\tau};\gamma_0)) \le [E((\Delta\epsilon_{it})^2(\Delta\mu(Z_{it};\gamma_0))^2)]^{\frac{1}{2}}[E((\Delta\epsilon_{i\tau})^2(\Delta\mu(Z_{i\tau};\gamma_0))^2)]^{\frac{1}{2}}$. Note that $E(\Delta\epsilon_{it}^2|Z_{it},Z_{it-1}) = 2\sigma_v^2+E(\tilde{u}_{it}^2|Z_{it},Z_{it-1})+E(\tilde{u}_{it-1}^2|Z_{it},Z_{it-1})-2E(\tilde{u}_{it}\tilde{u}_{it-1}|Z_{it},Z_{it-1}) < C$ by assumption A1(4). Thus, $E((\Delta\epsilon_{it})^2(\Delta\mu(Z_{it};\gamma_0))^2) \le CE((\Delta\mu(Z_{it};\gamma_0))^2) \le C[E\mu^2(Z_{it};\gamma_0)+E\mu^2(Z_{it-1};\gamma_0)] < C$ by assumption A1(4), where the second to the last inequality results from the $c-r$ inequality. Then for given T, we have $V(L_{222N}) = \frac{1}{N}$. So $L_{222N} = o_p(1)$. For $L_{221N} = \frac{2}{T-1}\sum_{t=2}^{T}\frac{1}{n}\sum_{i=1}^{n}\Delta\mu(Z_{it};\gamma)\Delta\epsilon_{it}$, we note that $E(\Delta\mu(Z_{it};\gamma)\Delta\epsilon_{it}) = 0$, $E(\Delta\epsilon_{it})^2 \le C[E(v_{it}^2)+E(v_{it-1}^2)+E\tilde{u}_{it}^2+E\tilde{u}_{it-1}^2] < C$ by the $c-r$ inequality, assumptions A1(3) and A1(4). Following similar arguments as in $L_{232N}(\gamma)$ above, we use assumptions A2(3), A2(4), and apply Lemma 5 to obtain $\sup_{\gamma\in\Gamma}|L_{221N}(\gamma)| = o_p(1)$. These two results imply $\sup_{\gamma\in\Gamma}|L_{22N}(\gamma)| = o_p(1)$.

The results on $L_{21N}(\gamma)$, $L_{22N}(\gamma)$ and $L_{23N}(\gamma)$ imply $\sup_{\gamma\in\Gamma}|L_{2N}(\gamma)| = o_p(1)$.

(4) $L_{3N} = \frac{1}{N}\Delta g'(I-P_{\Delta\Phi})\Delta g + \frac{2}{N}\Delta g'(I-P_{\Delta\Phi})\Delta\epsilon + \frac{1}{N}\Delta\epsilon'(I-P_{\Delta\Phi})\Delta\epsilon = L_{31N} + L_{32N} + L_{33N}$.
For $g_1 = \{g(X_{it})\}_{t=1}^{T-1}{}_{,i=1}^{n}$, $g_2 = \{g(X_{it})\}_{t=2}^{T}{}_{,i=1}^{n}$, we write

$$L_{31N} = \frac{1}{N}\Delta g_2'(I-P_{\Delta\Phi})\Delta g_2 - \frac{2}{N}\Delta g_2'(I-P_{\Delta\Phi})\Delta g_1 + \frac{1}{N}\Delta g_1'(I-P_{\Delta\Phi})\Delta g_1 = L_{311N} - 2L_{312N} + L_{313N}.$$

By Lemma 3, we obtain $L_{311N} = O_p((d\kappa^{-\delta_1}+d_1\kappa^{-2\delta_2})^2)$, and $L_{313N} = O_p((d\kappa^{-\delta_1}+d_1\kappa^{-2\delta_2})^2)$. Furthermore, by Cauchy-Schwartz inequality, $L_{312N} \leq L_{311N}^{\frac{1}{2}}L_{313N}^{\frac{1}{2}}$, so we obtain $L_{31N} = O_p((d\kappa^{-\delta_1}+d_1\kappa^{-2\delta_2})^2)$.

By Cauchy-Schwartz inequality again, we have $L_{32N} \leq 2L_{31N}^{\frac{1}{2}}L_{33N}^{\frac{1}{2}}$. By Lemma 4(1), $L_{33N} = \frac{1}{N}\Delta\epsilon'\Delta\epsilon - \frac{1}{N}\Delta\epsilon' P_{\Delta\Phi}\Delta\epsilon = \frac{1}{N}\Delta\epsilon'\Delta\epsilon + o_p(1)$. Combing the results, we have $L_{3N} = \frac{1}{N}\Delta\epsilon'\Delta\epsilon + o_p(1)$.

$\frac{1}{N}\Delta\epsilon'\Delta\epsilon = \frac{1}{T-1}\sum_{t=2}^{T}\frac{1}{n}\sum_{i=1}^{n}\Delta\epsilon_{it}^2 = \frac{1}{T-1}\sum_{t=2}^{T}\frac{1}{n}\sum_{i=1}^{n}[\epsilon_{it}^2 - 2\epsilon_{it}\epsilon_{it-1} + \epsilon_{it-1}^2]$.

By assumption A1(3) and A1(4), $E\epsilon_{it}^2 = Ev_{it}^2 + E(u_{it}^2 - \mu(Z_{it};\gamma_0))^2 = \sigma_v^2 + Eu_{it}^2 - E\mu(Z_{it};\gamma_0)^2 < C$. Similarly, $E\epsilon_{it-1}^2 < C$. By Cauchy-Schwartz inequality, $E\epsilon_{it}\epsilon_{it-1} < C$. So $E\Delta\epsilon_{it}^2 < C$. Thus, by the law of large numbers, we obtain the claim in (4) that $L_{3N} = \frac{1}{N}\Delta\epsilon'\Delta\epsilon + o_p(1) = \frac{1}{T-1}\sum_{t=1}^{T}E(\Delta\epsilon_{it})^2 + o_p(1)$.

The claim of part (a) follow from results (1)-(4) above.

(b) By Taylor expansion, for γ^* between $\hat\gamma$ and γ_0, because $\hat\gamma$ minimizes $L_N(\gamma)$, we have

$$\left.\frac{\partial L_N(\gamma)}{\partial\gamma}\right|_{\hat\gamma} = \left.\frac{\partial L_N(\gamma)}{\partial\gamma}\right|_{\gamma_0} + \left.\frac{\partial^2 L_N(\gamma)}{\partial\gamma\partial\gamma'}\right|_{\gamma^*}(\hat\gamma-\gamma_0), \text{ and } \sqrt{n}(\hat\gamma-\gamma_0) = -\left[\left.\frac{\partial^2 L_N(\gamma)}{\partial\gamma\partial\gamma'}\right|_{\gamma^*}\right]^{-1}\sqrt{n}\left.\frac{\partial L_N(\gamma)}{\partial\gamma}\right|_{\gamma_0} 1(\cdot)$$

where $1(\cdot)$ is an indicator for the event that $\left[\left.\frac{\partial^2 L_N(\gamma)}{\partial\gamma\partial\gamma'}\right|_{\gamma^*}\right]^{-1}$ exists. Given $\frac{\partial\Delta\mu(\gamma)}{\partial\gamma_j} = \{\frac{\partial\Delta\mu(Z_{it};\gamma)}{\partial\gamma_j}\}_{t=2}^{T}{}_{,i=1}^{n}$, $\frac{\partial\Delta\mu(\gamma)}{\partial\gamma} = [(\frac{\partial\Delta\mu(\gamma)}{\partial\gamma_1})', \cdots, (\frac{\partial\Delta\mu(\gamma)}{\partial\gamma_p})']'$, $\frac{\partial^2\Delta\mu(\gamma)}{\partial\gamma_j\partial\gamma_l} = \{\frac{\partial^2\Delta\mu(Z_{it};\gamma)}{\partial\gamma_j\partial\gamma_l}\}_{t=2}^{T}{}_{,i=1}^{n}$, we have

$$\frac{\partial L_N(\gamma)}{\partial\gamma'} = \frac{2}{N}(\Delta Y + \Delta\mu(\gamma))'(I-P_{\Delta\Phi})\frac{\partial\Delta\mu(\gamma)}{\partial\gamma'}, \text{ and } \frac{\partial^2 L_N(\gamma)}{\partial\gamma\partial\gamma'} = A_1(\gamma) + A_2(\gamma) + A_3(\gamma).$$

Denoting the $(jl)-th$ element of a matrix A by A_{jl}, $A_{1jl}(\gamma) = \frac{2}{N}(\frac{\partial\Delta\mu(\gamma)}{\partial\gamma_j})'(I-P_{\Delta\Phi})\frac{\partial\Delta\mu(\gamma)}{\partial\gamma_l}$, $A_{2jl}(\gamma) = \frac{2}{N}(\Delta g + \Delta\epsilon)'(I-P_{\Delta\Phi})\frac{\partial^2\Delta\mu(\gamma)}{\partial\gamma_j\partial\gamma_l}$, and $A_{3jl}(\gamma) = \frac{2}{N}(\Delta\mu(\gamma) - \Delta\mu(\gamma_0))'(I-P_{\Delta\Phi})\frac{\partial^2\Delta\mu(\gamma)}{\partial\gamma_j\partial\gamma_l}$. Recalling $\Sigma_{jl}(\gamma) =$

$\frac{2}{T-1}\sum_{t=2}^{T}E[\partial_j\Delta_{it}(\gamma)-\partial_j h(X_{it},X_{it-1};\gamma)][\partial_l\Delta_{it}(\gamma)-\partial_l h(X_{it},X_{it-1};\gamma)]$, defining $A_{3jl0}(\gamma)=\frac{2}{T-1}\sum_{t=2}^{T}E[\Delta_1\mu_{it}(\gamma)-\Delta_1\mu_{it}(\gamma_0)]\partial_{jl}\Delta_{it}(\gamma)$, we show below that

(1) $\sup_{\gamma\in\Gamma}|A_{1jl}(\gamma)-\Sigma_{jl}(\gamma)|=o_p(1)$.

(2) $\sup_{\gamma\in\Gamma}|A_{2jl}(\gamma)|=o_p(1)$.

(3) $\sup_{\gamma\in\Gamma}|A_{3jl}(\gamma)-A_{3jl0}(\gamma)|=o_p(1)$.

(4) $\sqrt{n}\left.\frac{\partial L_N(\gamma)}{\partial\gamma}\right|_{\gamma_0}\xrightarrow{d}N(0,\Omega_T)$.

With assumption B3, $\Sigma_{jl}(\gamma)+A_{3jl0}(\gamma)$ is continuous at $\gamma_0\in\Gamma$. Because $\hat{\gamma}\xrightarrow{p}\gamma_0$ as shown in part (a), $\gamma^*\xrightarrow{p}\gamma_0$. The $(j,l)-th$ element in $\frac{\partial^2 L_N(\gamma)}{\partial\gamma\partial\gamma'}$ as $A_{1jl}(\gamma)+A_{2jl}(\gamma)+A_{3jl}(\gamma)$, by results (1)-(3), converges in probability to $\Sigma_{jl}(\gamma)+A_{3jl0}(\gamma)$ uniformly in γ in an open neiborhood of γ_0 by A2(1). Then by Theorem 4.1.5 in Amemiya (1985), and the fact that $A_{3jl}(\gamma_0)=0$, $A_{1jl}(\gamma)+A_{2jl}(\gamma)+A_{3jl}(\gamma)|_{\gamma^*}\xrightarrow{p}\Sigma_{jl}(\gamma_0)$. So $\left.\frac{\partial^2 L_N(\gamma)}{\partial\gamma\partial\gamma'}\right|_{\gamma^*}\xrightarrow{p}\Sigma(\gamma_0)$. It is a nonstochastic invertible matrix as assumed by B4, so $1(\cdot)\xrightarrow{p}1$. Together with result (4), we obtain the claim in (b). Below we show results (1)-(4).

(1) $A_{1jl}(\gamma)=\frac{2}{N}\sum_{i=1}^{n}\sum_{t=2}^{T}\partial_j\Delta\mu_{it}(\gamma)\partial_l\Delta\mu_{it}(\gamma)-\frac{2}{N}(\frac{\partial\Delta\mu(\gamma)}{\partial\gamma_j})'P_{\Delta\Phi}\frac{\partial\Delta\mu(\gamma)}{\partial\gamma_l}=A_{11jl}(\gamma)-A_{12jl}(\gamma)$.

We claim that $\sup_{\gamma\in\Gamma}|A_{11jl}(\gamma)-A_{11jl0}(\gamma)|=o_p(1)$, where $A_{11jl0}(\gamma)=\frac{2}{T-1}\sum_{t=2}^{T}E\partial_j\Delta\mu_{it}(\gamma)\partial_l\Delta\mu_{it}(\gamma)$. $A_{11jl}(\gamma)=\frac{2}{T-1}\sum_{t=2}^{T}\frac{1}{n}\partial_j\Delta\mu_{it}(\gamma)\partial_l\Delta\mu_{it}(\gamma)=\frac{2}{T-1}\sum_{t=2}^{T}\frac{1}{n}G(Z_{it},Z_{it-1};\gamma)$. Because $ab-cd=(a-c)(b-d)+c(b-d)+(a-c)d$, with assumption B1, for $\gamma\neq\gamma'$,

$$\begin{aligned}
& |G(Z_{it},Z_{it-1};\gamma)-G(Z_{it},Z_{it-1};\gamma')| \\
\leq\ & |\partial_j\mu_{it}(\gamma)-\partial_j\mu_{it-1}(\gamma)-(\partial_j\mu_{it}(\gamma')-\partial_j\mu_{it-1}(\gamma'))||\partial_l\mu_{it}(\gamma)-\partial_l\mu_{it-1}(\gamma)-(\partial_l\mu_{it}(\gamma')-\partial_l\mu_{it-1}(\gamma'))| \\
& +|\partial_j\mu_{it}(\gamma')-\partial_j\mu_{it-1}(\gamma')||\partial_l\mu_{it}(\gamma)-\partial_l\mu_{it-1}(\gamma)-(\partial_l\mu_{it}(\gamma')-\partial_l\mu_{it-1}(\gamma'))| \\
& +|\partial_j\mu_{it}(\gamma)-\partial_j\mu_{it-1}(\gamma)-(\partial_j\mu_{it}(\gamma')-\partial_j\mu_{it-1}(\gamma'))||\partial_l\mu_{it}(\gamma')-\partial_l\mu_{it-1}(\gamma')| \\
\leq\ & ||\gamma-\gamma'||[B_{\partial_j\mu}(Z_{it})+B_{\partial_j\mu}(Z_{it-1})][|\partial_l\mu_{it}(\gamma)|+|\partial_l\mu_{it-1}(\gamma)|+|\partial_l\mu_{it}(\gamma')|+|\partial_l\mu_{it-1}(\gamma'))| \\
& +||\gamma-\gamma'||[B_{\partial_l\mu}(Z_{it})+B_{\partial_l\mu}(Z_{it-1})][|\partial_j\mu_{it}(\gamma')|+|\partial_j\mu_{it-1}(\gamma')||] \\
& +||\gamma-\gamma'||[B_{\partial_j\mu}(Z_{it})+B_{\partial_j\mu}(Z_{it-1})][|\partial_l\mu_{it}(\gamma')|+|\partial_l\mu_{it-1}(\gamma')||].
\end{aligned}$$

Given that $EB^2_{\partial_j\mu}(Z_{it})<C$ in B1, $E\sup_{\gamma\in\Gamma}|\partial_j\mu_{it}(\gamma)|^2<C$ in B5, we have $|G(Z_{it},Z_{it-1};\gamma)-G(Z_{it},Z_{it-1};\gamma')|\leq ||\gamma-\gamma'||B_4(Z_{it},Z_{it-1})$ for some $B_4(Z_{it},Z_{it-1})$ such that $EB_4(Z_{it},Z_{it-1})<C$. Given A1(1), and the Cramer's condition for $\partial_j\mu_{it}(\gamma)\partial_l\mu_{it'}(\gamma)$ in B6, we have all three conditions (i)-(iii) in Lemma 5 and conclude that $\sup_{\gamma\in\Gamma}|\frac{1}{n}G(Z_{it},Z_{it-1};\gamma)-EE\partial_j\Delta\mu_{it}(\gamma)\partial_l\Delta\mu_{it}(\gamma)|=O_p((\frac{Lnn}{n})^{\frac{1}{2}})$, which leads to the claim above.

Define $E_{\mathcal{G}}(E(\frac{\partial \Delta\mu(\gamma)}{\partial \gamma_j})|X) = \{\partial_j h(X_{it}, X_{it-1}; \gamma)\}_{t=2}^{T},_{i=1}^{n}$, then

$$\begin{aligned} A_{12jl}(\gamma) = & \tfrac{2}{N}[\tfrac{\partial\Delta\mu(\gamma)}{\partial\gamma_j} - E_{\mathcal{G}}(E(\tfrac{\partial\Delta\mu(\gamma)}{\partial\gamma_j})|X)]' P_{\Delta\Phi}[\tfrac{\partial\Delta\mu(\gamma)}{\partial\gamma_j} - E_{\mathcal{G}}(E(\tfrac{\partial\Delta\mu(\gamma)}{\partial\gamma_j})|X)] \\ & + \tfrac{4}{N}[\tfrac{\partial\Delta\mu(\gamma)}{\partial\gamma_j} - E_{\mathcal{G}}(E(\tfrac{\partial\Delta\mu(\gamma)}{\partial\gamma_j})|X)]' P_{\Delta\Phi} E_{\mathcal{G}}(E(\tfrac{\partial\Delta\mu(\gamma)}{\partial\gamma_j})|X) \\ & + \tfrac{2}{N}[E_{\mathcal{G}}(E(\tfrac{\partial\Delta\mu(\gamma)}{\partial\gamma_j})|X)]' P_{\Delta\Phi}[E_{\mathcal{G}}(E(\tfrac{\partial\Delta\mu(\gamma)}{\partial\gamma_j})|X)] \\ = & A_{121jl}(\gamma) + A_{122jl}(\gamma) + A_{123jl}(\gamma). \end{aligned}$$

By the Cauchy-Schwartz inequality, $A_{122jl}(\gamma) \le [A_{121jl}(\gamma)]^{\frac{1}{2}}[A_{123jl}(\gamma)]^{\frac{1}{2}}$, so we only focus on $A_{121jl}(\gamma)$ and $A_{123jl}(\gamma)$. We follow Lemma 4's argument to show that uniformly for all $\gamma \in \Gamma$, $A_{121jl}(\gamma) \le O_p(1)[\frac{1}{N}(\frac{\partial\Delta\mu(\gamma)}{\partial\gamma_j} - E_{\mathcal{G}}(E(\frac{\partial\Delta\mu(\gamma)}{\partial\gamma_j})|X))'\Delta\Phi][\frac{1}{N}(\Delta\Phi)'(\frac{\partial\Delta\mu(\gamma)}{\partial\gamma_l} - E_{\mathcal{G}}(E(\frac{\partial\Delta\mu(\gamma)}{\partial\gamma_;})|X))] = o_p(1)$, with assumptions B1 and B2 that $\partial_j\Delta_1\mu_{it}(\gamma)$ and $\partial_j\Delta_1\mu_{it}(\gamma)\partial_j\Delta_1\mu_{it'}(\gamma)$ satisfy the Cramer's condition. We claim that $\sup_{\gamma\in\Gamma}|A_{123jl}(\gamma) - A_{123jl0}(\gamma)| = o_p(1)$ with $A_{123jl0} = \frac{2}{T-1}\sum_{t=2}^{T} E[\partial_j h(X_{it}, X_{it-1}; \gamma)\partial_j h(X_{it}, X_{it-1}; \gamma)]$. First, we define $\hat{B}_{\partial_j h} = [(\Delta\Phi)'\Delta\Phi]^{-1}(\Delta\Phi)' E_{\mathcal{G}}(E(\frac{\partial\Delta\mu(\gamma)}{\partial\gamma_j})|X)$, then by Lemma 2, we have $\sup_{\gamma\in\Gamma}||\hat{B}_{\partial_j h} - B_{\partial_j h}|| = O_p(d\kappa^{-\delta_1} + d_1\kappa^{-2\delta_2})$. Following Lemma 3, we obtain $\sup_{\gamma\in\Gamma}\frac{2}{N}[E_{\mathcal{G}}(E(\frac{\partial\Delta\mu(\gamma)}{\partial\gamma_j})|X)]'(I_N - P_{\Delta\Phi})[E_{\mathcal{G}}(E(\frac{\partial\Delta\mu(\gamma)}{\partial\gamma_j})|X)] = O_p(d\kappa^{-\delta_1} + d_1\kappa^{-2\delta_2}) = o_p(1)$. Next, we write

$$\begin{aligned} \tfrac{2}{N}[E_{\mathcal{G}}(E(\tfrac{\partial\Delta\mu(\gamma)}{\partial\gamma_j})|X)]'[E_{\mathcal{G}}(E(\tfrac{\partial\Delta\mu(\gamma)}{\partial\gamma_j})|X)] = & \tfrac{2}{T-1}\sum_{t=2}^{T}\tfrac{1}{n}\sum_{i=1}^{n}\partial_j h(X_{it}, X_{it-1}; \gamma)\partial_l h(X_{it}, X_{it-1}; \gamma) \\ = & \tfrac{2}{T-1}\sum_{t=2}^{T}\tfrac{1}{n}\sum_{i=1}^{n} G(X_{it}, X_{it-1}; \gamma). \end{aligned}$$

With B1 and B5, we obtain that for $\gamma \ne \gamma'$, $|G(X_{it}, X_{it-1}; \gamma) - G(X_{it}, X_{it-1}; \gamma)| \le ||\gamma - \gamma'||B_5(X_{it}, X_{it-1})$ with some function $B_5(\cdot)$ such that $EB_5(X_{it}, X_{it-1}) < C$. With $B6$ that $\partial_j h(X_{it}, X_{it-1}; \gamma)\partial_l h(X_{it}, X_{it-1}; \gamma)$ satisfies the Cramer's condition, $\sup_{\gamma\in\Gamma}|\frac{1}{n}\sum_{i=1}^{n}|G(X_{it}, X_{it-1}; \gamma) - E\partial_j h(X_{it}, X_{it-1}; \gamma)\partial_l h(X_{it}, X_{it-1}; \gamma)| = o_p(1)$ by applying Lemma 5. The two steps above imply the claim that $\sup_{\gamma\in\Gamma}|A_{123jl}(\gamma) - A_{123jl0}(\gamma)| = o_p(1)$. So in all, $\sup_{\gamma\in\Gamma}|A_{12jl}(\gamma) - A_{123jl0}(\gamma)| = o_p(1)$.

$$\begin{aligned} & A_{11jl0}(\gamma) - A_{123jl0}(\gamma) \\ = & \tfrac{2}{T-1}\sum_{t=2}^{T}[E\partial_j\Delta\mu_{it}(\gamma)\partial_l\Delta\mu_{it}(\gamma) - \partial_j h(X_{it}, X_{it-1}; \gamma)\partial_l h(X_{it}, X_{it-1}; \gamma)] \\ = & \tfrac{2}{T-1}\sum_{t=2}^{T} E[\partial_j\Delta\mu_{it}(\gamma) - \partial_j h(X_{it}, X_{it-1}; \gamma)][\partial_l\Delta\mu_{it}(\gamma) - \partial_l h(X_{it}, X_{it-1}; \gamma)] \end{aligned}$$

since $E[\partial_j\Delta\mu_{it}(\gamma) - \partial_j h(X_{it}, X_{it-1}; \gamma)][\partial_l h(X_{it}, X_{it-1}; \gamma)] = 0$ due to the facts that $\partial_j h(X_{it}, X_{it-1}; \gamma) \in \mathcal{G}$, and $\partial_j\Delta\mu_{it}(\gamma) - \partial_j h(X_{it}, X_{it-1}; \gamma) \in \mathcal{G}^{\perp}$. Combing results above, we obtain the claim in (1).

$$\begin{aligned} (2)\ A_{2jl}(\gamma) = & \tfrac{2}{N}(\Delta g)'(I - P_{\Delta\Phi})\tfrac{\partial^2\Delta\mu(\gamma)}{\partial\gamma_j\partial\gamma_l} + \tfrac{2}{N}(\Delta\epsilon)'\tfrac{\partial^2\Delta\mu(\gamma)}{\partial\gamma_j\partial\gamma_l} - \tfrac{2}{N}(\Delta\epsilon)' P_{\Delta\Phi}\tfrac{\partial^2\Delta\mu(\gamma)}{\partial\gamma_j\partial\gamma_l} \\ = & A_{21jl}(\gamma) + A_{22jl}(\gamma) - A_{23jl}(\gamma). \end{aligned}$$

(i) $\frac{1}{2}A_{21jl}(\gamma) \le [\frac{1}{N}(\Delta g)'(I - P_{\Delta\Phi})\Delta g]^{1/2}[\frac{1}{N}(\frac{\partial^2\Delta\mu(\gamma)}{\partial\gamma_j\partial\gamma_l})'(I - P_{\Delta\Phi})(\frac{\partial^2\Delta\mu(\gamma)}{\partial\gamma_j\partial\gamma_l})]^{1/2} = [L_{31N}]^{1/2}[A_{211jl}(\gamma)]^{1/2}$. We know from part (a) (4) that $L_{31N} = O_p((d\kappa^{-\delta_1} + d_1\kappa^{-2\delta_2})^2) = o_p(1)$. We write

$A_{211jl}(\gamma) = \frac{1}{N}\sum_{i=1}^{n}\sum_{t=2}^{T}(\partial_{jl}\Delta\mu_{it}(\gamma))^2 - \frac{1}{N}(\frac{\partial^2\Delta\mu(\gamma)}{\partial\gamma_j\partial\gamma_l})'P_{\Delta\Phi}\frac{\partial^2\Delta\mu(\gamma)}{\partial\gamma_j\partial\gamma_l} = A_{2111jl}(\gamma) - A_{2112jl}(\gamma)$. The structure is similar to what we observe in (1) for $A_{1jl}(\gamma)$. So with $\partial_{jl}\mu_{it}(\gamma)$ satisfying the Lipchitz's condition and $EB^2_{\partial_{jl}\mu}(Z_{it}) < C$ in B1, B2, $E\sup_{\gamma\in\Gamma}|\partial_{jl}\mu_{it}(\gamma)|^2 < C$ in B5, and $(\partial_{jl}\mu_{it}(\gamma))^2$, $(\partial_{jl}h(X_{it}, X_{it-1};\gamma))^2$,$\partial_{jl}\Delta_1\mu_{it}(\gamma)$, and $\partial_{jl}\Delta_1\mu_{it}(\gamma)\partial_{jl}\Delta_1\mu_{it'}(\gamma)$ satisfy the Cramer's condition in B6, we obtain $\sup_{\gamma\in\Gamma}|A_{211jl}(\gamma) - A_{2110jl}(\gamma)| = o_p(1)$, and $\sup_{\gamma\in\Gamma}|A_{2112jl}(\gamma) - A_{21120jl}(\gamma)| = o_p(1)$, where $A_{2110jl}(\gamma) = \frac{2}{T-1}\sum_{t=2}^{T}E\partial_{jl}\Delta_1\mu_{it}(\gamma) < C$ and $A_{21120jl}(\gamma) = \frac{2}{T-1}\sum_{t=2}^{T}E[\partial_{jl}h(X_{it}, X_{it-1};\gamma)]^2 < C$ by B5. So $\sup_{\gamma\in\Gamma}A_{211jl}(\gamma) = O_p(1)$ and $\sup_{\gamma\in\Gamma}A_{21jl}(\gamma) = o_p(1)$.

(ii) $\frac{1}{2}A_{23jl}(\gamma) \leq [\frac{1}{N}(\Delta\epsilon)'P_{\Delta\Phi}\Delta\epsilon]^{1/2}[\frac{1}{N}(\frac{\partial^2\Delta\mu(\gamma)}{\partial\gamma_j\partial\gamma_l})'P_{\Delta\Phi}(\frac{\partial^2\Delta\mu(\gamma)}{\partial\gamma_j\partial\gamma_l})]^{1/2} = [A_{231jl}]^{1/2}[A_{2112jl}(\gamma)]^{1/2}$. We know from (i) above that $\sup_{\gamma\in\Gamma}|A_{2112jl}(\gamma)| = O_p(1)$. By Lemma 4, we have $A_{231jl} = o_p(1)$. So we obtain $\sup_{\gamma\in\Gamma}A_{23jl}(\gamma) = o_p(1)$.

(iii) $A_{22jl}(\gamma) = \frac{2}{T-1}\sum_{t=2}^{T}\frac{1}{n}\sum_{i=1}^{n}\Delta\epsilon_{it}\partial_{jl}\Delta\mu_{it}(\gamma) = \frac{2}{T-1}\sum_{t=2}^{T}\frac{1}{n}\sum_{i=1}^{n}G(Q_{it}, Q_{it-1};\gamma)$. By B1, $|G(Q_{it}, Q_{it-1};\gamma) - G(Q_{it}, Q_{it-1};\gamma')| \leq ||\gamma - \gamma'||B_6(Q_{it}, Q_{it-1})$, where $EB_6(Q_{it}, Q_{it-1}) < C$ by $EB^2_{\partial_{jl}\mu}(Z_{it}) < C$ in B1, $Ev^2_{it} = \sigma^2_v < C$ in A1(3), and $E\tilde{u}^2_{it} < C$ in A1(4). Furthermore, $EG(Q_{it}, Q_{it-1};\gamma) = 0$ by A1(5). Together with A1(1), and $\Delta\epsilon_{it}\partial_{jl}\Delta\mu_{it}(\gamma)$ satisfies the Cramer's condition in B6, we apply Lemma 5 to obtain $\sup_{\gamma\in\Gamma}|A_{22jl}(\gamma) = o_p(1)$.

Combining results (i)-(iii) above, we obtain $\sup_{\gamma\in\Gamma}|A_2(\gamma)| = o_p(1)$.

$$\begin{aligned}
(3)\ A_{3jl}(\gamma) = \quad & \tfrac{2}{N}(\Delta\mu(\gamma) - \Delta\mu(\gamma_0))'\tfrac{\partial^2\Delta\mu(\gamma)}{\partial\gamma_j\partial\gamma_l} \\
& -\tfrac{2}{N}(\Delta\mu(\gamma) - \Delta\mu(\gamma_0) - (E_{\mathcal{G}}(E(\Delta\mu(\gamma)|X)) - E_{\mathcal{G}}(E(\Delta\mu(\gamma_0)|X))))'P_{\Delta\Phi}\tfrac{\partial^2\Delta\mu(\gamma)}{\partial\gamma_j\partial\gamma_l} \\
& -\tfrac{2}{N}(E_{\mathcal{G}}(E(\Delta\mu(\gamma)|X)) - E_{\mathcal{G}}(E(\Delta\mu(\gamma_0)|X)))'P_{\Delta\Phi}\tfrac{\partial^2\Delta\mu(\gamma)}{\partial\gamma_j\partial\gamma_l} \\
= \quad & A_{31jl}(\gamma) - A_{32jl}(\gamma) - A_{33jl}(\gamma).
\end{aligned}$$

$$\begin{aligned}
\tfrac{1}{2}A_{32jl}(\gamma) \leq \quad & [\tfrac{1}{N}(\Delta\mu(\gamma) - \Delta\mu(\gamma_0) - E_{\mathcal{G}}(E(\Delta\mu(\gamma) - \Delta\mu(\gamma_0)|X)))'P_{\Delta\Phi} \\
& (\Delta\mu(\gamma) - \Delta\mu(\gamma_0) - E_{\mathcal{G}}(E(\Delta\mu(\gamma) - \Delta\mu(\gamma_0)|X)))]^{1/2}[\tfrac{1}{N}(\tfrac{\partial^2\Delta\mu(\gamma)}{\partial\gamma_j\partial\gamma_l})'P_{\Delta\Phi}\tfrac{\partial^2\Delta\mu(\gamma)}{\partial\gamma_j\partial\gamma_l}]^{1/2} \\
= \quad & o_p(1)[A_{2112jl}(\gamma)]^{1/2} \text{ uniformly over } \gamma\in\Gamma \text{ by Lemma 4(2).}
\end{aligned}$$

We show before that $\sup_{\gamma\in\Gamma}A_{2112jl}(\gamma) = O_p(1)$ in (2)(i), so $\sup_{\gamma\in\Gamma}A_{32jl}(\gamma) = o_p(1)$. We write

$$\begin{aligned}
\tfrac{1}{2}A_{33jl}(\gamma) \leq \quad & \tfrac{1}{N}E_{\mathcal{G}}(E(\Delta\mu(\gamma) - \Delta\mu(\gamma_0)|X))'(P_{\Delta\Phi} - I_N)\tfrac{\partial^2\Delta\mu(\gamma)}{\partial\gamma_j\partial\gamma_l} \\
& +\tfrac{1}{N}E_{\mathcal{G}}(E(\Delta\mu(\gamma) - \Delta\mu(\gamma_0)|X))'\tfrac{\partial^2\Delta\mu(\gamma)}{\partial\gamma_j\partial\gamma_l} \\
= \quad & A_{331jl}(\gamma) + A_{332jl}(\gamma).
\end{aligned}$$

By Cauchy-Schwartz inequality, the fact that $\sup_{\gamma\in\Gamma}L_{13N}(\gamma) = o_p(1)$ in (a) (2), and $\sup_{\gamma\in\Gamma}A_{211jl}(\gamma) = O_p(1)$ in (b)(2)(i) above, we have $\sup_{\gamma\in\Gamma}|A_{331jl}(\gamma)| = o_p(1)$. So in all, $A_{33jl}(\gamma) = 2A_{332jl}(\gamma) + o_p(1)$

uniformly over $\gamma \in \Gamma$. Combining above, we have $A_{3jl}(\gamma) = A_{31jl}(\gamma) - 2A_{332jl}(\gamma) + o_p(1)$ uniformly over $\gamma \in \Gamma$. Recalling the definition of $\Delta_1\mu_{it}(\gamma)$,

$$\begin{aligned} A_{31jl}(\gamma) - 2A_{332jl}(\gamma) = & \ \tfrac{2}{N}(\Delta\mu(\gamma) - \Delta\mu(\gamma_0) - E_{\mathcal{G}}(E(\Delta\mu(\gamma) - \Delta\mu(\gamma_0)|X)))' \tfrac{\partial^2 \Delta\mu(\gamma)}{\partial\gamma_j \partial\gamma_l} \\ = & \ \tfrac{2}{T-1}\sum_{t=2}^{T}\tfrac{1}{n}\sum_{i=1}^{n}(\Delta_1\mu_{it}(\gamma) - \Delta_1\mu_{it}(\gamma_0))\partial_{jl}\Delta\mu_{it}(\gamma) \\ = & \ \tfrac{2}{T-1}\sum_{t=2}^{T}\tfrac{1}{n}\sum_{i=1}^{n}Q(Z_{it}, Z_{it-1};\gamma). \end{aligned}$$

With the Lipschitz conditions on $\mu(Z_{it};\gamma)$ in A2(3) and $\partial_{jl}\mu_{it}(\gamma)$ in B1, we can show that $|Q(Z_{it}, Z_{it-1};\gamma) - Q(Z_{it}, Z_{it-1};\gamma_0)| < ||\gamma - \gamma_0||B_7(Z_{it}, Z_{it-1})$, where $EB_7(Z_{it}, Z_{it-1}) < C$ by $E(B_\mu^2(Z_{it})) < C$ in A2(3), $E\sup_{\gamma\in\Gamma}|\mu(Z_{it};\gamma)|^2 < C$ in A2(5), $E(B^2_{\partial_{jl}\mu}(Z_{it})) < C$ in B1, and $E\sup_{\gamma\in\Gamma}|\partial_{jl}\mu_{it}(\gamma)|^2 < C$ in B5. Furthermore, with A1(1), $\Delta_1\mu_{it}(\gamma)\partial_{jl}\mu_{it}(\gamma)$ and $\Delta_1\mu_{it}(\gamma_0)\partial_{jl}\mu_{it}(\gamma)$ satisfying the Cramer's condition in B6, we have the three conditions in Lemma 5 and conclude that $\sup_{\gamma\in\Gamma}|\frac{1}{n}\sum_{i=1}^{n}Q(Z_{it}, Z_{it-1};\gamma) - EQ(Z_{it}, Z_{it-1};\gamma)| = o_p(1)$. Because $\frac{2}{T-1}\sum_{t=2}^{T}EQ(Z_{it}, Z_{it-1};\gamma) = A_{3jl0}(\gamma)$, we obtain $\sup_{\gamma\in\Gamma}|A_{3jl}(\gamma) - A_{3jl0}(\gamma)| = o_p(1)$.

(4) $\sqrt{n}\frac{\partial L_N(\gamma)}{\partial\gamma}|_{\gamma_0} = \frac{2\sqrt{n}}{N}(\Delta g + \Delta\epsilon)'(I_N - P_{\Delta\Phi})\frac{\partial\Delta\mu(\gamma)}{\partial\gamma'}|_{\gamma_0} = B(\gamma_0)$. We consider the $j-th$ element as $B_j(\gamma_0) = \frac{2\sqrt{n}}{N}(\Delta g)'(I_N - P_{\Delta\Phi})\frac{\partial\Delta\mu(\gamma)}{\partial\gamma_j}|_{\gamma_0} + \frac{2\sqrt{n}}{N}(\Delta\epsilon)'(I_N - P_{\Delta\Phi})\frac{\partial\Delta\mu(\gamma)}{\partial\gamma_j}|_{\gamma_0} = B_{1j}(\gamma_0) + B_{2j}(\gamma_0)$.

(i) $B_{1j}(\gamma_0) \le 2\sqrt{n}[\frac{1}{N}(\Delta g)'(I_N - P_{\Delta\Phi})\Delta g]^{1/2}[\frac{1}{N}(\frac{\partial\Delta\mu(\gamma)}{\partial\gamma_j})'(I_N - P_{\Delta\Phi})\frac{\partial\Delta\mu(\gamma)}{\partial\gamma_j}|_{\gamma_0}]^{1/2} = 2\sqrt{n}[L_{31N}]^{1/2}[\frac{1}{2}A_{1jj}(\gamma_0)]^{1/2}$. We know from part (a) (4) and A3(3) that $\sqrt{n}[L_{31N}]^{1/2} = O_+p(\sqrt{n}(d\kappa^{-\delta} + d_1\kappa^{-2\delta_2})) = o_p(1)$. In part (b) (1) above, we know that $A_{1jj}(\gamma_0) - [A_{11jj0}(\gamma_0) - A_{123jj0}(\gamma_0)] = o_p(1)$, where $A_{11jj0}(\gamma_0) - A_{123jj0}(\gamma_0) = \frac{2}{T-1}\sum_{t=2}^{T}E(\partial_j\Delta\mu_{it}(\gamma) - \partial_j h(X_{it}, X_{it-1};\gamma_0))^2 < C$, so $A_{1jj}(\gamma_0) = O_p(1)$. In all, we have $B_{1j}(\gamma_0) = o_p(1)$.

$$\begin{aligned} \text{(ii) } B_{2j}(\gamma_0) = & \ \tfrac{2\sqrt{n}}{N}(\Delta\epsilon)'[\tfrac{\partial\Delta\mu(\gamma)}{\partial\gamma_j}|_{\gamma_0} - E_{\mathcal{G}}(E(\tfrac{\partial\Delta\mu(\gamma)}{\partial\gamma_j}|X))] - \tfrac{2\sqrt{n}}{N}(\Delta\epsilon)'[P_{\Delta\Phi}\tfrac{\partial\Delta\mu(\gamma)}{\partial\gamma_j}|_{\gamma_0} - E_{\mathcal{G}}(E(\tfrac{\partial\Delta\mu(\gamma)}{\partial\gamma_j}|X))] \\ = & \ B_{21j}(\gamma_0) - B_{22j}(\gamma_0). \end{aligned}$$

$$\begin{aligned} B_{22j}(\gamma_0) = & \ \tfrac{2\sqrt{n}}{N}(\Delta\epsilon)'P_{\Delta\Phi}[\tfrac{\partial\Delta\mu(\gamma)}{\partial\gamma_j}|_{\gamma_0} - E_{\mathcal{G}}(E(\tfrac{\partial\Delta\mu(\gamma)}{\partial\gamma_j}|X))] \\ & + \tfrac{2\sqrt{n}}{N}(\Delta\epsilon)'[P_{\Delta\Phi}E_{\mathcal{G}}(E(\tfrac{\partial\Delta\mu(\gamma)}{\partial\gamma_j}|X)) - E_{\mathcal{G}}(E(\tfrac{\partial\Delta\mu(\gamma)}{\partial\gamma_j}|X))] \\ = & \ B_{221j}(\gamma_0) + B_{222j}(\gamma_0). \end{aligned}$$

Defining $\partial_j\Delta_1\mu(\gamma_0) = \frac{\partial\Delta\mu(\gamma)}{\partial\gamma_j}|_{\gamma_0} - E_{\mathcal{G}}(E(\frac{\partial\Delta\mu(\gamma)}{\partial\gamma_j}|X))$, and $Z = \{Z_{it}\}_{t=2,}^{T}{}_{i=1}^{n}$. By assumption A1(4), we use similar arguments in Lemma 4 to have

$$\begin{aligned} E(B^2_{221j}(\gamma_0)|Z) = & \ \tfrac{4}{(T-1)^2 n}E[(\Delta\epsilon)'P_{\Delta\Phi}\partial_j\Delta_1\mu(\gamma_0)\partial_j\Delta_1\mu(\gamma_0)'P_{\Delta\Phi}\Delta\epsilon|Z] \\ = & \ \tfrac{4}{(T-1)N}tr(P_{\Delta\Phi}\partial_j\Delta_1\mu(\gamma_0)(\partial_j\Delta_1\mu(\gamma_0))'P_{\Delta\Phi}[E(\Delta\epsilon(\Delta\epsilon)'|Z) - C'(T-1)I_N]) \\ & + \tfrac{4}{(T-1)N}tr((\partial_j\Delta_1\mu(\gamma_0))'P_{\Delta\Phi}\partial_j\Delta_1\mu(\gamma_0)C'(T-1)) \\ \le & \ \tfrac{4C}{N}(\partial_j\Delta_1\mu(\gamma_0))'P_{\Delta\Phi}\partial_j\Delta_1\mu(\gamma_0) \\ = & \ o_p(1) \text{ by results on } A_{121jj}(\gamma) \text{ in (1). So we have } B_{221j}(\gamma_0) = o_p(1). \end{aligned}$$

$$
\begin{aligned}
B_{222j}(\gamma_0) = & \ \frac{2\sqrt{n}}{N}(\Delta\epsilon)'(P_{\Delta\Phi}-I_N)E_{\mathcal{G}}(E(\frac{\partial\Delta\mu(\gamma_0)}{\partial\gamma_j}|X)) \\
\leq & \ 2\sqrt{n}[\frac{1}{N}(\Delta\epsilon)'(P_{\Delta\Phi}-I_N)\Delta\epsilon]^{1/2}[\frac{1}{N}(E_{\mathcal{G}}(E(\frac{\partial\Delta\mu(\gamma_0)}{\partial\gamma_j}|X)))'(P_{\Delta\Phi}-I_N)E_{\mathcal{G}}(E(\frac{\partial\Delta\mu(\gamma_0)}{\partial\gamma_j}|X))]^{1/2} \\
= & \ 2\sqrt{n}[B_{2221j}(\gamma_0)]^{1/2}[B_{2222j}(\gamma_0)]^{1/2}.
\end{aligned}
$$

In Part (1)'s argument, we obtain $\sup_{\gamma\in\Gamma} B_{2222j}(\gamma) = O_p((d\kappa^{-\delta_1}+d_1\kappa^{-2\delta_2})^2)$, so by A3(3), we have $B_{2222j}(\gamma_0) = o_p(\frac{1}{n})$.

$B_{2221j}(\gamma_0) = \frac{1}{N}(\Delta\epsilon)'P_{\Delta\Phi}\Delta\epsilon - \frac{1}{N}(\Delta\epsilon)'\Delta\epsilon = -\frac{1}{N}(\Delta\epsilon)'\Delta\epsilon + o_p(1)$ by the result in (2) that $A_{231jl} = o_p(1)$. Furthermore, $E|\frac{1}{N}(\Delta\epsilon)'P_{\Delta\Phi}\Delta\epsilon| = \frac{1}{T-1}\sum_{t=2}^{T}[2\sigma_v^2 + E(u_{it}-\mu(Z_{it};\gamma_0))^2 + E(u_{it-1}-\mu(Z_{it-1};\gamma_0))^2 - 2E(u_{it}-\mu(Z_{it};\gamma_0))(u_{it-1}-\mu(Z_{it-1};\gamma_0))] < C$, so $B_{2221j}(\gamma_0) = O_p(1)$. We have $B_{222j}(\gamma_0) = o_p(1)$, $B_{22j}(\gamma_0) = o_p(1)$.

Now consider $B_{21}(\gamma_0) = \{B_{21j}(\gamma_0)\}_{j=1}^p$, where $B_{21j}(\gamma_0) = \frac{2\sqrt{n}}{N}\sum_{i=1}^{n}\sum_{t=2}^{T}\Delta\epsilon_{it}\partial_j\Delta\mu_{it}(\gamma_0)$. We let $q = (q_1,\cdots,q_p)'$ be a $p\times 1$ vector of arbitrary nonzero constants. Then $q'B_{21}(\gamma_0) = \frac{2\sqrt{n}}{N}\sum_{j=1}^p\sum_{i=1}^{n}\sum_{t=2}^{T}q_j\Delta\epsilon_{it}\partial_j\Delta\mu_{it}(\gamma_0) = \frac{1}{\sqrt{n}}\sum_{i=1}^{n}\frac{2}{T-1}\sum_{t=2}^{T}q_j\Delta\epsilon_{it}\partial_j\Delta\mu_{it}(\gamma_0)$. By assumptions A1(3) and A1(5), $E(q'B_{21}(\gamma_0)) = 0$. $V(q'B_{21}(\gamma_0)) = \sum_{j=1}^p\sum_{l=1}^p q_jq_l\Omega_{Tjl}$, where Ω_T is a $p\times p$ matrix with its $(jl)-th$ element being

$$
\begin{aligned}
\Omega_{Tjl} = & \ \frac{4}{(T-1)^2}\sum_{t=2}^{T}\sum_{\tau=2}^{T}E(\Delta\epsilon_{it}\Delta\epsilon_{i\tau}\partial_j\Delta\mu_{it}(\gamma_0)\partial_l\Delta\mu_{i\tau}(\gamma_0)) \\
= & \ \frac{4}{(T-1)^2}[\sum_{t=2}^{T}(2\sigma_v^2 + E((\tilde{u}_{it}-\tilde{u}_{it-1})^2|Z_{it},Z_{it-1}))\partial_j\Delta\mu_{it}(\gamma_0)\partial_l\Delta\mu_{it}(\gamma_0) \\
& + \sum_{\substack{t=2 \\ t\neq\tau}}^{T}\sum_{\tau=2}^{T}E(\Delta\epsilon_{it}\Delta\epsilon_{i\tau}|Z_{it},Z_{it-1})\partial_j\Delta\mu_{it}(\gamma_0)\partial_l\Delta\mu_{i\tau}(\gamma_0)] < C.
\end{aligned}
$$

By the Lindeberg-Levy Central Limit Theorem (Theorem 23.3 in Davidson (1994)), we have $q'B_{21}(\gamma_0) \xrightarrow{d} N(0, q'\Omega_T q)$. By the Cramer's device, we have $B_{21}(\gamma_0) \xrightarrow{d} N(0,\Omega_T)$. Together with the result that $B_{22j}(\gamma_0) = o_p(1)$, we obtain $B_2(\gamma_0) \xrightarrow{d} N(0,\Omega_T)$, where $B_2(\gamma_0) = \{B_{2j}(\gamma_0)\}_{j=1}^p$.

Combining results in (i) and (ii), for $B_1(\gamma_0) = \{B_{1j}(\gamma_0)\}_{j=1}^p$, we have the claim that $\sqrt{n}\frac{\partial L_N(\gamma)}{\partial\gamma}|_{\gamma_0} = B(\gamma_0) = B_1(\gamma_0) + B_2(\gamma_0) \xrightarrow{d} N(0,\Omega_T)$. □

Theorem 2.

Proof. (a). We define $\theta = [(\theta^m)', (\theta^H)']'$, where $\theta^m = [(\theta^1)',\cdots,(\theta^d)']'$ with $\theta^j = B_{m_j}$ in assumption A3(2), and $\theta^H = [(\theta^{12})',\cdots,(\theta^{1d})',(\theta^{23})',\cdots,(\theta^{2d})',\cdots,(\theta^{d-1d})']'$, with $\theta^{jl} = B_{H_{jl}}$ in assumption A3(2). Recalling that $\Delta Y = \Delta g - \Delta\mu(\gamma_0) + \Delta\epsilon$, with q_N in Lemma 2, we write

$$
\begin{aligned}
1_N(\hat{\theta}(\hat{\gamma})-\theta) = & \ 1_N[\frac{1}{N}\Delta\Phi'\Delta\Phi]\frac{1}{N}\Delta\Phi'\Delta\epsilon + 1_N[\frac{1}{N}\Delta\Phi'\Delta\Phi]\frac{1}{N}\Delta\Phi'(\Delta g - \Delta\Phi'\theta) \\
& + 1_N[\frac{1}{N}\Delta\Phi'\Delta\Phi]\frac{1}{N}\Delta\Phi'(\Delta\mu(\hat{\gamma})-\Delta\mu(\gamma_0)) \\
= & \ T_{1n} + T_{2n} + T_{3n}.
\end{aligned}
$$

T_{1n}: with $1_N \neq 0$, $\hat{\mathcal{Q}}$ is PD, so $\hat{\mathcal{Q}}^{-1}\hat{\mathcal{Q}}^{-1} = \hat{\mathcal{Q}}^{-1/2}\hat{\mathcal{Q}}^{-1}\hat{\mathcal{Q}}^{-1/2}$. So with similar arguments in Lemma 2,

$$\begin{aligned} ||T_{1n}|| = & \ 1_N||[\tfrac{1}{N}\Delta\Phi'\Delta\Phi]^{-1}\tfrac{1}{N}\Delta\Phi'\Delta\epsilon|| = 1_N[(\tfrac{1}{N}\Delta\epsilon'\Delta\Phi)\hat{\mathcal{Q}}^{-1}\hat{\mathcal{Q}}^{-1}\tfrac{1}{N}\Delta\Phi'\Delta\epsilon]^{1/2} \\ = & \ 1_N[(\tfrac{1}{N}\Delta\epsilon'\Delta\Phi)\hat{\mathcal{Q}}^{-1/2}\hat{\mathcal{Q}}^{-1}\hat{\mathcal{Q}}^{-1/2}\tfrac{1}{N}\Delta\Phi'\Delta\epsilon]^{1/2} = O_p(1)1_N[(\tfrac{1}{N}\Delta\epsilon'\Delta\Phi)\hat{\mathcal{Q}}^{-1}\tfrac{1}{N}\Delta\Phi'\Delta\epsilon]^{1/2} \\ = & \ O_p(1)1_N[(\tfrac{1}{N}\Delta\epsilon' P_{\Delta\Phi}\Delta\epsilon]^{1/2} = O_p(1)1_N O_p((\tfrac{S(\kappa)}{n})^{1/2}) = O_p((\tfrac{S(\kappa)}{n})^{1/2}), \end{aligned}$$

where the we have used Lemma 4 (i) and Lemma 2 (1). So $||T_{1n}|| = O_p((\frac{S(\kappa)}{n})^{1/2})$.

T_{2n} : by triangle inequality and assumption A3(2), we have $\sup_{X_{it}\in[-1,1]^d, X_{it-1}\in[-1,1]^d} |\Delta g(X_{it}) - \Delta\Phi(X_{it})'\theta| \leq \sum_{j=1}^d \sup_{X_{it}^j\in[-1,1]} |m_j(X_{it}^j) - \phi^\kappa(X_{it}^j)'\theta^j| + \sum\sum_{1\leq j<l\leq d} \sup_{X_{it}^j\in[-1,1], X_{it}^l\in[-1,1]} |H_{jl}(X_{it}^j, X_{it}^l) - \phi^{\kappa^2}(X_{it}^j, X_{it}^l)'\theta^{jl}| + \sum_{j=1}^d \sup_{X_{it-1}^j\in[-1,1]} |m_j(X_{it-1}^j) - \phi^\kappa(X_{it-1}^j)'\theta^j| + \sum\sum_{1\leq j<l\leq d} \sup_{X_{it-1}^j\in[-1,1], X_{it-1}^l\in[-1,1]} |H_{jl}(X_{it-1}^j, X_{it-1}^l) - \phi^{\kappa^2}(X_{it-1}^j, X_{it-1}^l)'\theta^{jl}| = O(d\kappa^{-\delta_1} + d_1\kappa^{-2\delta_2})$.

$$\begin{aligned} ||T_{2n}|| = & \ 1_N[(\tfrac{1}{N}(\Delta g - \Delta\Phi'\theta)'\Delta\Phi)\hat{\mathcal{Q}}^{-1/2}\hat{\mathcal{Q}}^{-1}\hat{\mathcal{Q}}^{-1/2}\tfrac{1}{N}\Delta\Phi'(\Delta g - \Delta\Phi'\theta)]^{1/2} \\ = & \ O_p(1)1_N[(\tfrac{1}{N}(\Delta g - \Delta\Phi'\theta)'\Delta\Phi\hat{\mathcal{Q}}^{-1}\tfrac{1}{N}\Delta\Phi'(\Delta g - \Delta\Phi'\theta)]^{1/2} \\ = & \ O_p(1)1_N[(\tfrac{1}{N}(\Delta g - \Delta\Phi'\theta)'P_{\Delta\Phi}(\Delta g - \Delta\Phi'\theta)]^{1/2} \\ \leq & \ O_p(1)1_N[(\tfrac{1}{N}(\Delta g - \Delta\Phi'\theta)'(\Delta g - \Delta\Phi'\theta)]^{1/2} \text{ since } P_{\Delta\Phi} \text{ is idempotent,} \\ = & \ O_p(1)1_N[\tfrac{1}{N}\sum_{i=1}^n\sum_{t=2}^T(\Delta g(X_{it}) - \Delta\Phi(X_{it})'\theta)^2]^{1/2} \\ = & \ O(d\kappa^{-\delta_1} + d_1\kappa^{-2\delta_2}). \end{aligned}$$

T_{3n} : by $c-r$ inequality and A2(3), $(\Delta\mu(Z_{it};\hat{\gamma}) - \Delta\mu(Z_{it};\gamma_0))^2 \leq C[(\mu(Z_{it};\hat{\gamma}) - \mu(Z_{it};\gamma_0))^2 + (\mu(Z_{it-1};\hat{\gamma}) - \mu(Z_{it-1};\gamma_0))^2] \leq C[B_\mu^2(Z_{it})||\hat{\gamma} - \gamma_0||^2 + B_\mu^2(Z_{it-1})||\hat{\gamma} - \gamma_0||^2]$. With similar arguments, we have

$$\begin{aligned} ||T_{3n}|| = & \ O_p(1)1_N[(\tfrac{1}{N}(\Delta\mu(\hat{\gamma}) - \Delta\mu(\gamma_0))'(\Delta\mu(\hat{\gamma}) - \Delta\mu(\gamma_0))]^{1/2} \\ = & \ O_p(1)1_N[\tfrac{1}{N}\sum_{i=1}^n\sum_{t=2}^T(\Delta\mu(Z_{it};\hat{\gamma}) - \Delta\mu(Z_{it};\gamma_0))^2]^{1/2} \\ \leq & \ O_p(1)1_N\tfrac{1}{N}\sum_{i=1}^n\sum_{t=2}^T(B_\mu^2(Z_{it}) + B_\mu^2(Z_{it-1}))||\hat{\gamma} - \gamma_0|| = O_p(n^{-1/2}), \end{aligned}$$

by assumption A2(3) and Theorem 1.

So in all, $1_N||\hat{\theta}(\hat{\gamma}) - \theta|| \leq ||T_{1n}|| + ||T_{2n}|| + ||T_{3n}|| = O_p((\frac{S(\kappa)}{n})^{1/2} + d\kappa^{-\delta_1} + d_1\kappa^{-2\delta_2} + n^{-1/2}) = O_p((\frac{S(\kappa)}{n})^{1/2} + d\kappa^{-\delta_1} + d_1\kappa^{-2\delta_2})$.

(b) Given results in (a), we easily have $1_N||\hat{\theta}^j(\hat{\gamma}) - \theta^j|| = O_p((\frac{S(\kappa)}{n})^{1/2} + d\kappa^{-\delta_1} + d_1\kappa^{-2\delta_2})$, and $1_N||\hat{\theta}^{jl}(\hat{\gamma}) - \theta^{jl}|| = O_p((\frac{S(\kappa)}{n})^{1/2} + d\kappa^{-\delta_1} + d_1\kappa^{-2\delta_2})$. So

$$\begin{aligned} & 1_N \sup_{x^j\in[-1.1]} |\hat{m}_j(x_j) - m_j(x^j)| = 1_N \sup_{x^j\in[-1.1]} |\phi^\kappa(x^j)'\hat{\theta}^j(\hat{\gamma}) - m_j(x^j)| \\ \leq & \ 1_N \sup_{x^j\in[-1.1]} |\phi^\kappa(x^j)'(\hat{\theta}^j(\hat{\gamma}) - \theta^j)| + 1_N sup_{x^j\in[-1,1]}|\phi^\kappa(x^j)'\theta^j - m_j(x^j)| \text{ by A3(2) and (4)} \\ \leq & \ \xi_{S(\kappa)}1_N||\hat{\theta}^j(\hat{\gamma}) - \theta^j|| + O(\kappa^{-\delta_1}) = O_p(\xi_{S(\kappa)}((\tfrac{S(\kappa)}{n})^{1/2} + d\kappa^{-\delta_1} + d_1\kappa^{-2\delta_2})). \end{aligned}$$

(c) With assumptions A3(2) and (4), we use similar arguments to obtain

$$1_N \sup_{x^j\in[-1.1], x^l\in[-1.1]} |\hat{H}_{jl}(x_j, x^l) - H_{jl}(x^j, x^l)| = O_p(\xi_{S(\kappa)}((\tfrac{S(\kappa)}{n})^{1/2} + d\kappa^{-\delta_1} + d_1\kappa^{-2\delta_2})).$$

□

Theorem 3.

Proof. (a) We start by considering the infeasible estimator $\tilde{m}_j^o(x^j)$, which is defined as in (7) using $Q_{it,-j}$ instead of $\hat{Q}_{it,-j}$, where $Q_{it,-j} = \Delta Y_{it} + m_j(X_{it-1}^j) - \sum_{l=1,l\neq j}^d \Delta m_l(X_{it}^l) - \sum\sum_{1\leq j<l\leq d} \Delta H_{jl}(X_{it}^j, x_{it}^l) + \Delta\mu(Z_{it};\gamma_0)$. So $\tilde{m}_j^o(x^j) = e_{1,2}' S_{N,j}^{-1}(x^j) \frac{1}{Nh}\sum_{i=1}^n\sum_{t=2}^T K_{it,j} \begin{bmatrix} 1 \\ X_{it}^j - x^j \end{bmatrix} Q_{it,-j}$, where $e_{1,2}$ is a 2×1 vector, with the first element being 1 and second being 0, $K_{it,j} = K(\frac{X_{it}^j - x^j}{h})$, $S_{N,j}(x_j) = \begin{bmatrix} S_{N,j0}(x^j) & S_{N,j1}(x^j) \\ S_{N,j1}(x^j) & S_{N,j2}(x^j) \end{bmatrix}$, and $S_{N,jl}(x^j) = \frac{1}{Nh}\sum_{i=1}^n\sum_{t=2}^T K_{it,j}(X_{it}^j - x^j)^l$ for $l = 0,1,2$. So with $m_j^{(i)}(x^j)$ being the $i-th$ derivative of $m_j(x^j)$, we have

$$\begin{aligned}
\tilde{m}_j^o(x^j) - m_j(x^j) = & \; e_{1,2}' S_{N,j}^{-1}(x^j)[\frac{1}{Nh}\sum_{i=1}^n\sum_{t=2}^T K_{it,j} \begin{bmatrix} 1 \\ X_{it}^j - x^j \end{bmatrix} Q_{it,-j} - S_{N,j}(x^j) \begin{bmatrix} m_j(x^j) \\ m_j^{(1)}(x^j) \end{bmatrix} \\
= & \; e_{1,2}' S_{N,j}^{-1}(x^j) \frac{1}{Nh}\sum_{i=1}^n\sum_{t=2}^T K_{it,j} \begin{bmatrix} 1 \\ X_{it}^j - x^j \end{bmatrix} [\frac{1}{2}(X_{it}^j - x^j)^2 m_j^{(2)}(X_{it}^{j*}) + \Delta\epsilon_{it}].
\end{aligned}$$

(1) We show $\sqrt{nh}(\tilde{m}_j^o(x^j) - m_j(x^j) - (\frac{h^2}{2}m_j^{(2)}(x^j)\mu_{k,2} + o_p(h^2))) \xrightarrow{d} N(0, \sigma_{m_j}^2(x^j))$.

Consider $S_{N,jl}(x^j)$. With assumptions C1-C3, we apply Lemma 6 to have $\sup_{x^j\in\mathcal{G}} |S_{N,jl}(x^j) - ES_{N,jl}(x^j)| \leq \frac{1}{T-1}\sum_{t=2}^T \sup_{x^j\in\mathcal{G}} |\frac{1}{nh}\sum_{i=1}^n K_{it,j}(X_{it}^j - x^j)^l - E\frac{1}{nh}\sum_{i=1}^n K_{it,j}(X_{it}^j - x^j)^l| = O_p((\frac{lnn}{nh})^{1/2}h^l)$. With assumption C2, $\frac{1}{h^l}ES_{N,jl}(x^j) = \frac{1}{T-1}\sum_{t=2}^T \frac{1}{h}EK_{it,j}(\frac{X_{it}^j - x^j}{h})^l = \frac{1}{T-1}\sum_{t=2}^T \int K(\psi)\psi^l[f_{X_t^j}(x^j) + h\psi f_{X_t^j}^{(1)}(x^j) + \frac{h^2\psi^2}{2} f_{X_t^j}^{(2)}(x^{j^*})]d\psi$, where x^{j*} is between x^j and X_{it}^j. So we have, uniformly for all $x^j\in\mathcal{G}$, $S_{Nj0}(x^j) = \frac{1}{T-1}\sum_{t=2}^T f_{X_t^j}(x^j) + O_p(h^2 + (\frac{lnn}{nh})^{1/2})$, $S_{Nj1}(x^j) = \frac{1}{T-1}\sum_{t=2}^T f_{X_t^j}^{(1)}(x^j)\mu_{k,2}h^2 + O_p(h(\frac{lnn}{nh})^{1/2})$, and $S_{Nj2}(x^j) = \frac{1}{T-1}\sum_{t=2}^T f_{X_t^j}(x^j)\mu_{k,2}h^2 + O_p(h^2(\frac{lnn}{nh})^{1/2})$. Let $G_H = \begin{bmatrix} 1 & 0 \\ 0 & h^2 \end{bmatrix}$, $\tilde{S}_{N,j}(x_j) = G_H S_{N,j}(x_j) = \begin{bmatrix} S_{N,j0}(x^j) & S_{N,j1}(x^j) \\ \frac{1}{h^2}S_{N,j1}(x^j) & \frac{1}{h^2}S_{N,j2}(x^j) \end{bmatrix}$, and $e_{1,2}'\tilde{S}_{N,j}^{-1}(x_j) = [A_{11}, A_{12}]$. Specifically, $A_{11} = \frac{\frac{1}{h^2}S_{N,j2}(x^j)}{S_{N,j0}(x^j)\frac{1}{h^2}S_{N,j2}(x^j) - \frac{1}{h^2}S_{N,j1}^2(x^j)} = \frac{1}{\frac{1}{T-1}\sum_{t=2}^T f_{X_t^j(x^j)}} + o_p(1)$, and $A_{12} = -\frac{S_{N,j1}(x^j)}{S_{N,j0}(x^j)\frac{1}{h^2}S_{N,j2}(x^j) - \frac{1}{h^2}S_{N,j1}^2(x^j)} = O_p(h^2)$. So we have

$$\begin{aligned}
\tilde{m}_j^o(x^j) - m_j(x^j) = & \; e_{1,2}' S_{N,j}^{-1}(x^j) G_H^{-1} G_H \frac{1}{Nh}\sum_{i=1}^n\sum_{t=2}^T K_{it,j} \begin{bmatrix} 1 \\ X_{it}^j - x^j \end{bmatrix} [\frac{1}{2}(X_{it}^j - x^j)^2 m_j^{(2)}(X_{it}^{j*}) + \Delta\epsilon_{it}] \\
= & \; e_{1,2}' \tilde{S}_{N,j}^{-1}(x^j) \frac{1}{Nh}\sum_{i=1}^n\sum_{t=2}^T K_{it,j} \begin{bmatrix} 1 \\ \frac{X_{it}^j - x^j}{h^2} \end{bmatrix} [\frac{1}{2}(X_{it}^j - x^j)^2 m_j^{(2)}(X_{it}^{j*}) + \Delta\epsilon_{it}] \\
= & \; [A_{11},\, A_{12}] \frac{1}{Nh}\sum_{i=1}^n\sum_{t=2}^T K_{it,j} \begin{bmatrix} 1 \\ \frac{X_{it}^j - x^j}{h^2} \end{bmatrix} [\frac{1}{2}(X_{it}^j - x^j)^2 m_j^{(2)}(X_{it}^{j*}) + \Delta\epsilon_{it}] \\
= & \; [\frac{1}{T-1}\sum_{t=2}^T f_{X_t^j}(x^j)]^{-1} \frac{1}{Nh}\sum_{i=1}^n\sum_{t=2}^T K_{it,j}[\frac{1}{2}(X_{it}^j - x^j)^2 m_j^{(2)}(X_{it}^{j*}) + \Delta\epsilon_{it}](1 + o_p(1)) \\
& + O_p(h)\frac{1}{Nh}\sum_{i=1}^n\sum_{t=2}^T K_{it,j}(\frac{X_{it}^j - x^j}{h})[\frac{1}{2}(X_{it}^j - x^j)^2 m_j^{(2)}(X_{it}^{j*}) + \Delta\epsilon_{it}] \\
= & \; I_{1n}(1 + o_p(1)) + O_p(h)I_{2n}.
\end{aligned}$$

Define $I_{anl} = \frac{1}{Nh}\sum_{i=1}^n\sum_{t=2}^T K_{it,j}(\frac{X_{it}^j - x^j}{h})^l \frac{1}{2}(X_{it}^j - x^j)^2 m_j^{(2)}(X_{it}^{j*})$, and $I_{bnl} = \frac{1}{Nh}\sum_{i=1}^n\sum_{t=2}^T K_{it,j}(\frac{X_{it}^j - x^j}{h})^l \Delta\epsilon_{it}$ for $l = 0,1$. We claim that

(i) $I_{anl} = \frac{h^2}{2(T-1)}\sum_{t=2}^{T} f_{X_t^j}(x^j) m_j^{(2)}(x^j) \int K(\psi)\psi^{l+2} d\psi + o_p(h^2) + O_p(h^2(\frac{lnn}{nh})^{1/2})$ uniformly over $x^j \in G$, a compact subset of $[-1,1]$.

(ii) $\sup_{x^j \in G} |I_{bnl}| = O_p((\frac{lnn}{nh})^{1/2})$.

(iii) $\sqrt{nh} I_{bn0} \xrightarrow{d} N(0, \sigma^2_{m_j}(x^j)[\frac{1}{T-1}\sum_{t=2}^{T} f_{X_t^j}(x^j)]^2)$.

With (i) and (ii), $\sqrt{nh}O_p(h)I_{2n} = \sqrt{nh}O_p(h)[I_{an1} + I_{bn1}] = O_p(\sqrt{nh}h(\frac{lnn}{nh})^{1/2}) + o_p(\sqrt{nh}h^3) = o_p(1)$ with C4(1). $I_{1n} = [\frac{1}{T-1}\sum_{t=2}^{T} f_{X_t^j}(x^j)]^{-1}[I_{an0}+I_{bn0}](1+o_p(1)) = \frac{h^2}{2} m_j^{(2)}(x^j)\mu_{k,2} + o_p(h^2) + [\frac{1}{T-1}\sum_{t=2}^{T} f_{X_t^j}(x^j)]^{-1} I_{bn0}(1+o_p(1))$. Together with (iii), we obtain $\sqrt{nh}(\tilde{m}_j^o(x^j) - m_j(x^j) - (\frac{h^2}{2} m_j^{(2)}(x^j)\mu_{k,2} + o_p(h^2))) \xrightarrow{d} N(0, \sigma^2_{m_j}(x^j))$. Below, we show the claims (i)-(iii).

(i) $|I_{anl} - EI_{anl}| \leq \frac{h^2}{2(T-1)}\sum_{t=2}^{T} |\frac{1}{nh}\sum_{i=1}^{n}[K_{it,j}(\frac{X_{itj^j} - x^j}{h})^{l+2} m_j^{(2)}(X_{it}^{j*}) - EK_{it,j}(\frac{X_{itj^j} - x^j}{h})^{l+2} m_j^{(2)}(X_{it}^{j*})]| = O_p((\frac{lnn}{nh})^{1/2}h^2)$ uniformly over $x^j \in G$, where we apply Lemma 6 together with assumption C2(1). $EI_{anl} = \frac{h^2}{2(T-1)}\sum_{t=2}^{T} f_{X_t^j}(x^j) \int K(\psi)\psi^{l+2} d\psi m_j^{(2)}(x^j)(1 + o_p(1))$ uniformly over $x^j \in G$. These give the claim in (i).

(ii) By Assumption A1(3) and (5), $E(\Delta\epsilon_{it}|X_{it}) = 0$, thus, $E(I_{bnl}) = 0$. By C2(2), $E|\Delta\epsilon_{it}|^{2+\delta} < C$. With A1(1) and C3(1), $f_{X_t^j|\Delta\epsilon_{it}}(x^j) < C$, and we can apply Lemma 6 to have $\sup_{x^j \in G}|I_{bnl}| \leq \frac{1}{T-1}\sum_{t=2}^{T} \sup_{x^j \in G} |\frac{1}{n}\sum_{i=1}^{n} K_{it,j}(\frac{X_{it}^j - x^j}{h})^l \Delta\epsilon_{it}| = O_p((\frac{lnn}{nh})^{1/2})$, and we obtain the claim in (ii).

(iii) $\sqrt{nh} I_{bn0} = \sum_{i=1}^{n} \frac{1}{\sqrt{nh}(T-1)}\sum_{t=2}^{T} K_{it,j}\Delta\epsilon_{it} = \sum_{i=1}^{n} W_{ni}$. $S_n^2 = \sum_{i=1}^{n} V(W_{ni}) = \frac{1}{(T-1)^2}\sum_{t=2}^{T}\frac{1}{nh}\sum_{i=1}^{n} EK_{it,j}^2\Delta\epsilon_{it}^2 + \frac{1}{(T-1)^2}\sum_{\substack{t=2 \\ t\neq\tau}}^{T}\sum_{\tau=2}^{T}\frac{1}{nh}\sum_{i=1}^{n} EK_{it,j}K_{i\tau,j}\Delta\epsilon_{it}\Delta\epsilon_{i\tau} = S_{n1}^2 + S_{n2}^2$. $S_{n1}^2 = \frac{1}{(T-1)^2}\sum_{t=2}^{T}\int K(\psi)^2\sigma^2_{\Delta\epsilon_{it}}(x^j + h\psi) f_{X_t^j}(x^j + h\psi)d\psi \to \frac{1}{(T-1)^2}\sum_{t=2}^{T}\sigma^2_{\Delta\epsilon_{it}}(x^j) f_{X_t^j}(x^j)\int K(\psi)^2 d\psi$ by assumptions C2(3) and C3(3). With A1(3) and A1(4), we have $E(|\Delta\epsilon_{it}|^2||X_{it}, X_{i\tau}) \leq CE(v_{it}^2 + v_{it-1}^2 + \tilde{u}_{it}^2 + \tilde{u}_{it-1}^2|X_{it}, X_{t\tau}) < C$ because $E(\tilde{u}_{it}^2|X_{it}, X_{i\tau}) \leq C[E(u_{it}^2|X_{it}, X_{i\tau}) + E(\mu^2(Z_{it};\gamma_0)|X_{it}, X_{i\tau})] < C$. By the Cauchy-Schwartz inequality, $E(|\Delta\epsilon_{it}\Delta\epsilon_{i\tau}||X_{it}, X_{i\tau}) \leq [E(|\Delta\epsilon_{it}|^2|X_{it}, X_{i\tau})]^{1/2}[E(|\Delta\epsilon_{i\tau}|^2|X_{it}, X_{i\tau})]^{1/2}$. So $S_{n2}^2 = \frac{1}{(T-1)^2}\sum_{\substack{t=2 \\ t\neq\tau}}^{T}\sum_{\tau=2}^{T}\frac{1}{h}E[K_{it,j}K_{i\tau,j}E(\Delta\epsilon_{it}\Delta\epsilon_{i\tau}|X_{it}, X_{i\tau})] \leq \frac{C}{(T-1)^2}\sum_{\substack{t=2 \\ t\neq\tau}}^{T}\sum_{\tau=2}^{T}\frac{1}{h}E[K_{it,j}K_{i\tau,j}] = O(h)$ with assumption C3(3). So $S_n^2 = \frac{1}{(T-1)^2}\sum_{t=2}^{T}\sigma^2_{\Delta\epsilon_{it}}(x^j) f_{X_t^j}(x^j)\int K(\psi)^2 d\psi + o(1) = \sigma^2_{m_j}(x^j)[\frac{1}{T-1}\sum_{t=2}^{T} f_{X_t^j}(x^j)]^2 + o(1)$. Furthermore, we have the Liapunov's condition that $\sum_{i=1}^{n} E|\frac{W_{ni}}{S_n}|^{2+\delta} \to 0$ as $n \to \infty$ for some $\delta > 0$, because $S_n^{2+\delta} \to C$, and $\sum_{i=1}^{n} E|W_{ni}|^{2+\delta} = (nh)^{-\delta/2}\frac{1}{h}E|\frac{1}{T-1}\sum_{t=2}^{T} K_{it,j}\Delta\epsilon_{it}|^{2+\delta} < C(nh)^{-\delta/2}\frac{1}{(T-1)^{2+\delta}}\sum_{t=2}^{T}\int K(\psi)^{2+\delta} f_{X_t^j}(x^j + h\psi)d\psi \to 0$ by assumptions C1, C2(2) and C3(3). Thus, by the

Lindeberg's Central Limit Theorem, we obtain the claim that $\sqrt{nh}I_{bn0} \xrightarrow{d} N(0, \sigma^2_{m_j}(x^j)[\frac{1}{T-1}\sum_{t=2}^{T} f_{X_t^j}(x^j)]^2)$.

(2) We show that $\sqrt{nh}(\tilde{m}_j(x^j) - \tilde{m}_j^o(x^j)) = o_p(1)$. Note that (1) and (2) implies the claim in (a) that $\sqrt{nh}(\tilde{m}_j(x^j) - m_j(x^j) - (\frac{h^2}{2}m_j^{(2)}(x^j)\mu_{k,2} + o_p(h^2))) \xrightarrow{d} N(0, \sigma^2_{m_j}(x^j))$.

Given that $\hat{Q}_{it,-j} - Q_{it,-j} = \hat{m}_j(X^j_{it-1}) - m_j(X^j_{it-1}) - \sum_{l=1,l\neq j}^{d}[\Delta\hat{m}_l(X^l_{it}) - \Delta m_l(X^l_{it})] - \sum\sum_{1\leq j<l\leq d}[\Delta\hat{H}_{jl}(X^j_{it}, x^l_{it}) - \Delta H_{jl}(X^j_{it}, x^l_{it})] + \Delta\mu(Z_{it}; \hat{\gamma}) - \Delta\mu(Z_{it}; \gamma_0) = \hat{I}_{1n} + \hat{I}_{2n} + \hat{I}_{3n} + \hat{I}_{4n}$.

$$
\begin{aligned}
\tilde{m}_j(x^j) - \tilde{m}_j^o(x^j) = \;& e'_{1,2}\tilde{S}^{-1}_{N,j}(x^j)\frac{1}{Nh}\sum_{i=1}^{n}\sum_{t=2}^{T}K_{it,j}\begin{bmatrix} 1 \\ \frac{X^j_{it}-x^j}{h^2}\end{bmatrix}[\hat{Q}_{it,-j} - Q_{it,-j}] \\
= \;& [A_{11},\, A_{12}]\frac{1}{Nh}\sum_{i=1}^{n}\sum_{t=2}^{T}K_{it,j}\begin{bmatrix} 1 \\ \frac{X^j_{it}-x^j}{h^2}\end{bmatrix}[\hat{I}_{1n} + \hat{I}_{2n} + \hat{I}_{3n} + \hat{I}_{4n}] \\
= \;& [\frac{1}{T-1}\sum_{t=2}^{T} f_{X_t^j}(x^j)]^{-1}\frac{1}{Nh}\sum_{i=1}^{n}\sum_{t=2}^{T}K_{it,j}[\hat{I}_{1n} + \hat{I}_{2n} + \hat{I}_{3n} + \hat{I}_{4n}](1 + o_p(1)) \\
& + O_p(h)\frac{1}{Nh}\sum_{i=1}^{n}\sum_{t=2}^{T}K_{it,j}(\frac{X^j_{it}-x^j}{h})[\hat{I}_{1n} + \hat{I}_{2n} + \hat{I}_{3n} + \hat{I}_{4n}].
\end{aligned}
$$

(i) $\sqrt{nh}\frac{1}{nh}\sum_{i=1}^{n}\sum_{t=2}^{T}K_{it,j}\hat{I}_{1n} = o_p(1)$, and $\sqrt{nh}h\frac{1}{nh}\sum_{i=1}^{n}\sum_{t=2}^{T}K_{it,j}(\frac{X_{it,j}-x^j}{h})\hat{I}_{1n} = o_p(1)$.

(ii) $\sqrt{nh}\frac{1}{nh}\sum_{i=1}^{n}\sum_{t=2}^{T}K_{it,j}(\hat{I}_{2n} + \hat{I}_{3n}) = o_p(1)$, and $\sqrt{nh}h\frac{1}{nh}\sum_{i=1}^{n}\sum_{t=2}^{T}K_{it,j}(\frac{X_{it,j}-x^j}{h})(\hat{I}_{2n} + \hat{I}_{3n}) = o_p(1)$.

(iii) $\frac{1}{nh}\sum_{i=1}^{n}\sum_{t=2}^{T}K_{it,j}(\frac{X_{it,j}-x^j}{h})^l\hat{I}_{4n} = O_p(n^{-1/2})$.

Claims (i)-(iii) above lead to the result in (2).

(i) $\sqrt{nh}\frac{1}{nh}\sum_{i=1}^{n}\sum_{t=2}^{T}K_{it,j}(\frac{X_{it,j}-x^j}{h})^l\hat{I}_{1n} = \sqrt{nh}\frac{1}{nh}\sum_{i=1}^{n}\sum_{t=2}^{T}K_{it,j}(\frac{X_{it,j}-x^j}{h})^l(\hat{m}_j(X_{it-1}) - m_j(X^j_{it-1}))$, where $\hat{m}_j(x^j) - m_j(x^j) = \phi^\kappa(x^j)'(\hat{\theta}^j(\hat{\gamma}) - \theta^j) + \phi^\kappa(x^j)'\theta^j - m_j(x^j)$.

$\hat{\theta}^j(\hat{\gamma}) - \theta^j = e'_{j\kappa}(\hat{\theta}(\hat{\gamma}) - \theta) = e'_{j\kappa}\hat{\mathcal{Q}}^{-1}\frac{1}{N}\Delta\Phi'(\Delta g - \Delta\Phi'\theta + \Delta\epsilon + \Delta\mu(\hat{\gamma}) - \Delta\mu(\gamma_0))$, where $e_{j\kappa}$ is a $\kappa \times (\kappa d + \kappa^2 d(d-1)/2)$ matrix of zeros, with its $(j-1)\kappa + 1$ to $(j+1)\kappa - 1$ columns being replaced by the identity matrix I_κ. By Lemma 1, $||\mathcal{Q} - I_{S(\kappa)}||^2 = O_p(\frac{1}{N}\xi^2_{S(\kappa)}S(\kappa))$. With assumption 3(5) and the argument in Lemma 1, we have the smallest eigenvalue of $\mathcal{Q}$ is one. By the sub-multiplicative property of the spectral norm $||\cdot||_{sp}$, we have $||\hat{\mathcal{Q}}^{-1} - I^{-1}_{S(\kappa)}||_{sp} = ||\hat{\mathcal{Q}}^{-1}(I_{S(\kappa)} - \hat{\mathcal{Q}})I^{-1}_{S(\kappa)}||_{sp} \leq ||\hat{\mathcal{Q}}^{-1}||_{sp}||I_{S(\kappa)} - \hat{\mathcal{Q}}||_{sp}||I^{-1}_{S(\kappa)}||_{sp}$. $||I_{S(\kappa)} - \hat{\mathcal{Q}}||_{sp} \leq ||I_{S(\kappa)} - \hat{\mathcal{Q}}|| = O_p(\frac{1}{\sqrt{N}}\xi_{S(\kappa)}S^{1/2}(\kappa))$ by Lemma 1. $||I^{-1}_{S(\kappa)}||_{sp} = [\lambda_{min}(I'_{S(\kappa)}I_{S(\kappa)})]^{-1/2} = [\lambda_{min}(I^2_{S(\kappa)})]^{-1/2} = [\lambda_{min}(I_{S(\kappa)})]^{-1} = 1$. $||\hat{\mathcal{Q}}^{-1}||_{sp} = [\lambda_{min}(\hat{\mathcal{Q}})]^{-1} - [\lambda_{min}(I_{S(\kappa)})]^{-1} + [\lambda_{min}(I_{S(\kappa)})]^{-1} = \frac{\lambda_{min}(I_{S(\kappa)}) - \lambda_{min}(\hat{\mathcal{Q}})}{\lambda_{min}(\hat{\mathcal{Q}})\lambda_{min}(I_{S(\kappa)})} + [\lambda_{min}(I_{S(\kappa)})]^{-1} = [\lambda_{min}(I_{S(\kappa)})]^{-1} + o_p(1) = 1 + o_p(1)$ by Lemma 2 (1). So $||\hat{\mathcal{Q}}^{-1} - I^{-1}_{S(\kappa)}||_{sp} = O_p(\frac{1}{\sqrt{N}}\xi_{S(\kappa)}S^{1/2}(\kappa))$. So $\hat{m}_j(x^j) - m_j(x^j) = \phi^\kappa(x^j)'e'_{j\kappa}[I_{S(\kappa)} + R_{\theta n}]\frac{1}{N}\Delta\Phi'(\Delta g - \Delta\Phi'\theta + \Delta\epsilon + \Delta\mu(\hat{\gamma}) - \Delta\mu(\gamma_0)) + \phi^\kappa(x^j)'\theta^j - m_j(x^j)$, where $||R_{\theta n}||_{sp} = O_p(\frac{1}{\sqrt{N}}\xi_{S(\kappa)}S^{1/2}(\kappa))$. So $\frac{1}{nh}\sum_{i=1}^{n}\sum_{t=2}^{T}K_{it,j}(\frac{X_{it,j}-x^j}{h})^l\hat{I}_{1n} =$

$\frac{1}{nh}\sum_{i=1}^{n}\sum_{t=2}^{T}K_{it,j}(\frac{X_{it,j}-x^j}{h})^l[\phi^\kappa(X_{it-1}^j)'e_{j\kappa}'[I_{S(\kappa)}+R_{\theta n}]\frac{1}{N}\Delta\Phi'(\Delta g-\Delta\Phi'\theta+\Delta\epsilon+\Delta\mu(\hat{\gamma})-\Delta\mu(\gamma_0))+\phi^\kappa(X_{it-1}^j)'\theta^j-m_j(X_{it-1}^j)]=(D_{11l}+D_{12l}+D_{13l})+(D_{11lR}+D_{12lR}+D_{13lR})+D_{14l}$.

(A) $\sqrt{nh}D_{12l}=o_p(1)$, where for $d_{12l}(x^j)=\frac{1}{nh}\sum_{i=1}^{n}K_{it,j}(\frac{X_{it,j}-x^j}{h})^l\phi^\kappa(X_{it-1}^j)=\{d_{12l,k}(x^j)\}_{k=1}^{\kappa}$, $d_{12l,k}(x^j)=\frac{1}{nh}\sum_{i=1}^{n}K_{it,j}(\frac{X_{it,j}-x^j}{h})^l\phi_k(X_{it-1}^j)$, we have for $l=0,1$,

$$\begin{aligned}\sqrt{nh}D_{12l} = & \sqrt{nh}\frac{1}{nh}\sum_{i=1}^{n}\sum_{t=2}^{T}K_{it,j}(\frac{X_{it,j}-x^j}{h})^l\phi^\kappa(X_{it-1}^j)'e_{j\kappa}'\frac{1}{N}\Delta\Phi'\Delta\epsilon\\ = & \frac{1}{(T-1)^2}\sum_{t=2}^{T}\sum_{\tau=2}^{T}\sqrt{nh}(d_{12l}(x^j))'\frac{1}{n}\sum_{m=1}^{n}\Delta\phi^\kappa(X_{m\tau})\Delta\epsilon_{m\tau}.\end{aligned}$$

By C1, C3(2), (3) and C4(1), we apply Lemma 6 to conclude that $\sup_{x^j\in G}|d_{12l,k}(x^j)-Ed_{12l,k}(x^j)|=O_p((\frac{Lnn}{nh})^{1/2})$, where $Ed_{12l,k}(x^j)=\int K(\psi)\psi^l\phi_k(X_{it-1}^j)[f_{X_t^j,X_{t-1}^j}(x^j,X_{it-1}^j)+h\psi\frac{\partial}{\partial X_t^j}f_{X_t^j,X_{t-1}^j}(x^j,X_{it-1}^j)+(1/2)h^2\psi^2\frac{\partial^2}{\partial^2X_t^j}f_{X_t^j,X_{t-1}^j}(x^{j*},X_{it-1}^j)]d\psi dX_{it-1}^j$, so $Ed_{120,k}(x^j)=\int\phi_k(X_{it-1}^j)f_{X_t^j,X_{t-1}^j}(x^j,X_{it-1}^j)dX_{it-1}^j+O(h^2)$, and $Ed_{121,k}(x^j)=h\int\phi_k(X_{it-1}^j)\frac{\partial}{\partial X_t^j}f_{X_t^j,X_{t-1}^j}(x^j,X_{it-1}^j)dX_{it-1}^j\mu_{k,2}+O(h^2)$. So $d_{12l}(x^j)=d_{12l0}(x^j)+R_{121l}+R_{122l}$ for $l=0,1$, where $d_{1200}(x^j)=\int\phi^\kappa(X_{it-1}^j)f_{X_t^j,X_{t-1}^j}(x^j,X_{it-1}^j)dX_{it-1}^j$, $d_{1210}(x^j)=0$,

$\sup_{x^j\in G}||R_{121l}||=O_p(\sqrt{\kappa}(\frac{Lnn}{nh})^{1/2})$, $\sup_{x^j\in G}||R_{1220l}||=O_p(\sqrt{\kappa}h^2)$, and $\sup_{x^j\in G}||R_{1221l}||=O_p(\sqrt{\kappa}h)$. So we have $\sqrt{nh}D_{12l}=\frac{1}{T-1}\sum_{t=2}^{T}\sqrt{nh}[d_{12l0}(x^j)+R_{121l}+R_{122l}]'\frac{1}{n}\sum_{m=1}^{n}\sum_{\tau=2}^{T}\Delta\phi^\kappa(X_{m\tau})\Delta\epsilon_{m\tau}$. As observed in p.2435 of Lemma 7 of Horowitz and Mammen (2004), for each $x^j\in[-1,1]$, $d_{1200}(x^j)$'s components are the Fourier coefficients of a function (by C3(4), $f_{X_t^j,X_{t-1}^j}(X_{it}^j,X_{it-1}^j)\in C^2$) that is uniformly bounded over x^j, so $\sup_{|x^j|\leq1}||d_{1200}(x^j)||^2\leq C$.

Let $D_{12lA}=d_{12l0}(x^j)'\frac{1}{n}\sum_{m=1}^{n}\sum_{\tau=2}^{T}\Delta\phi^\kappa(X_{m\tau})\Delta\epsilon_{m\tau}$. Denoting $\Delta\phi^\kappa=\{(\Delta\phi^\kappa(X_{m\tau}))'\}_{m=1}^{n},\tau=2^T$, then $E(\sqrt{nh}D_{12lA})^2=\frac{nh}{N}d_{12l0}(x^j)'\frac{1}{N}E((\Delta\phi^\kappa)'\Delta\epsilon\Delta\epsilon'\Delta\phi^\kappa)d_{12l0}(x^j)$. From Lemma 1, $\mathcal{Q}=E[\Delta\Phi(X_{it})\Delta\Phi(X_{it})']=I_{S(\kappa)}$ for all t. $\frac{1}{N}E[(\Delta\phi^\kappa)'\Delta\phi^\kappa)=e_{j\kappa}'\frac{1}{N}E[(\Delta\Phi)'\Delta\Phi]e_{j\kappa}=e_{j\kappa}'e_{j\kappa}=I_\kappa$. From Lemma 4 (i), we have $E((\Delta\epsilon)(\Delta\epsilon)'|X)-C'(T-1)I_N$ is ND, so $E(\sqrt{nh}D_{12lA})^2=\frac{nh}{N}d_{12l0}(x^j)'\frac{1}{N}E((\Delta\phi^\kappa)'[E(\Delta\epsilon\Delta\epsilon'|X)-C'(T-1)I_N+C'(T-1)I_N]\Delta\phi^\kappa)d_{12l0}(x^j)\leq\frac{nh}{N}d_{12l0}(x^j)'I_\kappa d_{12l0}(x^j)C'(T-1)=hd_{12l0}(x^j)'d_{12l0}(x^j)C'=O(h)$. For $l=0$, $E(\sqrt{nh}D_{12lA})^2=0$. So $\sqrt{nh}D_{120A}=O_p(h^{1/2})$ and $\sqrt{nh}D_{121A}=0$.

Let $\sqrt{nh}D_{12lB}=\sqrt{nh}(R_{121l}+R_{122l})'\frac{1}{N}\sum_{m=1}^{n}\sum_{\tau=2}^{T}\Delta\phi^\kappa(X_{m\tau})\Delta\epsilon_{m\tau}$. $E||\frac{1}{N}\sum_{m=1}^{n}\sum_{\tau=2}^{T}\Delta\phi^\kappa(X_{m\tau})\Delta\epsilon_{m\tau}||^2=\frac{1}{N^2}tr(E(\Delta\phi^\kappa(\Delta\phi^\kappa)'[E(\Delta\epsilon\Delta\epsilon'|X)-C'(T-1)I_N+C'(T-1)I_N]))\leq\frac{1}{N^2}tr(E(\Delta\phi^\kappa(\Delta\phi^\kappa)'C'(T-1)I_N))=\frac{C'\kappa}{n}$, where the inequality is due to the facts that $\Delta\phi^\kappa(\Delta\phi^\kappa)'$ is PSD, and $E((\Delta\epsilon)(\Delta\epsilon)'|X)-C'(T-1)I_N$ is ND, and the equality follows from the assumption that $\mathcal{Q}=E[\Delta\Phi(X_{it})\Delta\Phi(X_{it})']=I_{S(\kappa)}$. So

$\sqrt{nh}D_{12lB} \leq \sqrt{nh}(||R_{121l}||+||R_{122l}||)||\frac{1}{N}\sum_{m=1}^{n}\sum_{\tau=2}^{T}\Delta\phi^{\kappa}(X_{m\tau})\Delta\epsilon_{m\tau}||$. So $\sqrt{nh}D_{120B} = O_p(\sqrt{nh}(\sqrt{\kappa}(\frac{Lnn}{nh})^{1/2}+\sqrt{\kappa}h^2)(\frac{\kappa}{n})^{1/2}) = O_p((\frac{\kappa^2}{n})^{1/2}(Lnn)^{1/2}+\kappa h^{5/2}) = o_p(1)$ by assumption C4(1). $\sqrt{nh}D_{12lB} = O_p(\sqrt{nh}(\sqrt{\kappa}(\frac{Lnn}{nh})^{1/2}+\sqrt{\kappa}h)(\frac{\kappa}{n})^{1/2}) = o_p(1) + O_p(h^{-1}\kappa h^{5/2}) = o_p(1)$ by C4(1) as well. So we have $\sqrt{nh}D_{12lB} = o_p(1)$ for $l = 0, 1$. In all, $\sqrt{nh}D_{12l} = \frac{1}{T-1}\sum_{t=2}^{T}\sqrt{nh}(D_{12lA} + D_{12lB})$. So $\sqrt{nh}D_{12l} = o_p(1)$ for $l = 0, 1$.

(B) $\sqrt{nh}h^l D_{11l} = o_p(1)$ for $l = 0, 1$, where $D_{11l} = \frac{1}{nh}\sum_{i=1}^{n}\sum_{t=2}^{T}K_{it,j}(\frac{X_{it,j}-x^j}{h})^l\phi^{\kappa}(X_{it-1}^j)'e_{j\kappa}'\frac{1}{N}\Delta\Phi'(\Delta g - \Delta\Phi'\theta) = \frac{1}{T-1}\sum_{t=2}^{T}[d_{12l0}(x^j)+R_{121l}+R_{122l}]'d_{11}$, and $d_{11} = e_{j\kappa}'\frac{1}{N}\Delta\Phi'(\Delta g-\Delta\Phi'\theta) = \frac{1}{N}(\Delta\phi^{\kappa})'(\Delta g-\Delta\Phi'\theta)$. Let $W = \Delta g - \Delta\Phi'\theta$, then by Theorem 2 (a) term T_{2n}, we have $\frac{1}{N}||W||^2 = O_p((d\kappa^{-\delta_1}+d_1\kappa^{-2\delta_2})^2)$. By the exercise on p.342 of Horn and Johnson (2013), $||d_{11}||^2 = \frac{||W||^2}{N}\frac{W'}{||W||}[\frac{1}{N}\Delta\phi^{\kappa}(\Delta\phi^{\kappa})']\frac{W}{||W||} \leq \frac{||W||^2}{N}\lambda_{max}(\frac{1}{N}\Delta\phi^{\kappa}(\Delta\phi^{\kappa})') = \frac{||W||^2}{N}\lambda_{max}(\frac{1}{N}(\Delta\phi^{\kappa})'\Delta\phi^{\kappa})$. For $\hat{\mathcal{Q}}_{\kappa} = \frac{1}{N}(\Delta\phi^{\kappa})'\Delta\phi^{\kappa})$, we apply Weyl's theorem as in Lemma 2 (1) to obtain that $\lambda_{min}(I_{\kappa} - \hat{\mathcal{Q}}_{\kappa}) \leq \lambda_{max}(I_{\kappa}) - \lambda_{max}(\hat{\mathcal{Q}}_{\kappa}) \leq \lambda_{min}(I_{\kappa} - \hat{\mathcal{Q}}_{\kappa})$. By Lemma 1, $|\lambda_{min}(I_{\kappa} - \hat{\mathcal{Q}}_{\kappa})| \leq ||I_{\kappa} - \hat{\mathcal{Q}}_{\kappa}|| = O_p(\frac{1}{\sqrt{n}}\xi_{S(\kappa)}S^{\frac{1}{2}}(\kappa))$, $|\lambda_{max}(I_{\kappa} - \hat{\mathcal{Q}}_{\kappa})| \leq ||I_{\kappa} - \hat{\mathcal{Q}}_{\kappa}|| = O_p(\frac{1}{\sqrt{n}}\xi_{S(\kappa)}S^{\frac{1}{2}}(\kappa))$, and $\lambda_{max}(I_{\kappa}) - \lambda_{max}(\hat{\mathcal{Q}}_{\kappa}) = O_p(\frac{1}{\sqrt{n}}\xi_{S(\kappa)}S^{\frac{1}{2}}(\kappa))$. So $||d_{11}||^2 = \frac{||W||^2}{N}\lambda_{max}(\hat{\mathcal{Q}}_{\kappa} - \lambda_{max}(I_{\kappa}) + \lambda_{max}(I_{\kappa})) = O_p((d\kappa^{-\delta_1} + d_1\kappa^{-2\delta_2})^2)$. Then $\sqrt{nh}D_{11l} \leq \frac{1}{T-1}\sum_{t=2}^{T}\sqrt{nh}[||d_{12l0}(x^j)|| + ||R_{121l}|| + ||R_{122l}||]||d_{11}||$. So $\sqrt{nh}D_{110} = [O_p(1) + O_p(\sqrt{\kappa}(\frac{Lnn}{nh})^{1/2}) + O_p(\sqrt{\kappa}h^2)]O_p((d\kappa^{-\delta_1} + d_1\kappa^{-2\delta_2})\sqrt{nh}) = o_p(1)$ by C4(1). Similarly, $\sqrt{nh}D_{11l} \leq \frac{1}{T-1}\sum_{t=2}^{T}\sqrt{nh}[||d_{12l0}(x^j)|| + ||R_{121l}|| + ||R_{122l}||]||d_{11}||$. So $\sqrt{nh}hD_{111} = [O_p(\sqrt{\kappa}(\frac{Lnn}{nh})^{1/2}h) + O_p(\sqrt{\kappa}h^2)]O_p((d\kappa^{-\delta_1} + d_1\kappa^{-2\delta_2})\sqrt{nh}) = o_p(1)$.

(C) $\sqrt{nh}D_{13l} = o_p(1)$ for $l = 0, 1$, where $D_{13l} = \frac{1}{nh}\sum_{i=1}^{n}\sum_{t=2}^{T}K_{it,j}(\frac{X_{it,j}-x^j}{h})^l\phi^{\kappa}(X_{it-1}^j)'e_{j\kappa}'\frac{1}{N}\Delta\Phi'(\Delta\mu(\hat{\gamma}) - \Delta\mu(\gamma_0))$. Here, because $\frac{1}{N}||\Delta\mu(\hat{\gamma}) - \Delta\mu(\gamma_0)||^2 = O_p(n^{-1})$ as in T_{3n} of Theorem 2 (a), we follow part (B) to obtain the result in a similar fashion.

(D) $\sqrt{nh}D_{12lR} = o_p(1)$, where for a $\kappa \times \kappa$ matrix $R_{\theta n,1}$ such that $||R_{\theta n,1}||_{sp} = O_p(\frac{1}{\sqrt{N}}\xi_{S(\kappa)}S^{1/2}(\kappa))$, $D_{121R} = \frac{1}{Nh}\sum_{i=1}^{n}\sum_{t=2}^{T}K_{it,j}(\frac{X_{it}^j-x^j}{h})^j\phi^{\kappa}(X_{it-1}^j)'R_{\theta n,1}e_{j\kappa}'\frac{1}{N}\Delta\Phi'\Delta\epsilon$. Note this is a result of (2)(i) on $e_{j\kappa}'R_{\theta n}$. So $\sqrt{nh}D_{12lR} \leq \frac{1}{T-1}\sum_{t=2}^{T}\sqrt{nh}[||d_{12l0}(x^j)|| + ||R_{121l}|| + ||R_{122l}||]||R_{\theta n,1}||_{sp}O_p((\frac{\kappa}{n})^{1/2})$, because as shown in (A), $||\frac{1}{N}\sum_{m=1}^{n}\sum_{\tau=2}^{T}\Delta\phi^{\kappa}(X_{m\tau})\Delta\epsilon_{m\tau}|| = O_p((\frac{\kappa}{n})^{1/2})$. The claim $\sqrt{nh}D_{12lR} = o_p(1)$ follows similarly.

(E) $D_{11lR} = \frac{1}{Nh}\sum_{i=1}^{n}\sum_{t=2}^{T}K_{it,j}(\frac{X_{it}^j-x^j}{h})^j\phi^{\kappa}(X_{it-1}^j)'R_{\theta n,1}e_{j\kappa}'\frac{1}{N}\Delta\Phi'(\Delta g-\Delta\Phi'\theta) \leq \frac{1}{T-1}\sum_{t=2}^{T}[||d_{12l0}(x^j)||+||R_{121l}||+||R_{122l}||]||R_{\theta n,1}||_{sp}||d_{11}|| = o_p(D_{11l})$, because $||R_{\theta n,1}||_{sp} = o_p(1)$.

(F) Similarly, $D_{11lR} = \frac{1}{Nh}\sum_{i=1}^{n}\sum_{t=2}^{T} K_{it,j}(\frac{X_{it}^j - x^j}{h})^j \phi^\kappa(X_{it-1}^j)' R_{\theta n,1} e'_{j\kappa} \frac{1}{N}\Delta\Phi'(\Delta\mu(\hat{\gamma}) - \Delta\mu(\gamma_0)) = o_p(D_{13l})$.

(G) $|D_{14l}| = \frac{1}{Nh}\sum_{i=1}^{n}\sum_{t=2}^{T} K_{it,j}(\frac{X_{it}^j - x^j}{h})^j (\phi^\kappa(X_{it-1}^j)'\theta^j - m_j(X_{it-1}^j)) \leq \sup_{x^j \in [-1,1]} |\phi^\kappa(x^j)'\theta^j - m_j(x^j)| \frac{1}{Nh}\sum_{i=1}^{n}\sum_{t=2}^{T} K_{it,j}(\frac{X_{it}^j - x^j}{h})^j$ $O_p(\kappa^{-\delta_1})$, given A3(2) and the fact that $E|\frac{1}{Nh}\sum_{i=1}^{n}\sum_{t=2}^{T} K_{it,j}(\frac{X_{it}^j - x^j}{h})^j| \to \frac{1}{T-1}\sum_{t=2}^{T} f_{X_t^j}(x^j) \int K(\psi)|\psi|^l d\psi = O(1)$. So with assumption A3(3), we have $\sqrt{nh} D_{14l} = O_p(\sqrt{nh}\kappa^{-\delta_1}) = o_p(1)$.

So combining (A) -(G), we have the claim in (i).

(ii) Define $\Delta\tilde{\Phi}_{-j}^{d\kappa}(X_{it}) = [\Delta\phi^\kappa(x^1)', \cdots, \Delta\phi^\kappa(x^{j-1})', 0'_\kappa, \Delta\phi^\kappa(x^{j+1}),', \cdots, \Delta\phi^\kappa(x^d)']'$ and $\Delta\tilde{\Phi}_{-j}(X_{it})' = [\Delta\tilde{\Phi}_{-j}^{d\kappa}(X_{it})', \Delta\tilde{\Phi}^{d_1\kappa^2}(X_{it})']$. Then

$$
\begin{aligned}
& \frac{1}{Nh}\sum_{i=1}^{n}\sum_{t=2}^{T} K_{it,j}(\frac{X_{it}^j - x^j}{h})^l (I_{2n} + I_{3n}) \\
= & -\frac{1}{Nh}\sum_{i=1}^{n}\sum_{t=2}^{T} K_{it,j}(\frac{X_{it}^j - x^j}{h})^l [\Delta\tilde{\Phi}_{-j}(X_{it})'(\hat{\theta}(\hat{\gamma}) - \theta) + \Delta\tilde{\Phi}_{-j}(X_{it})'\theta - \Delta g_{-j}(X_{it})] \\
= & -[\frac{1}{Nh}\sum_{i=1}^{n}\sum_{t=2}^{T} K_{it,j}(\frac{X_{it}^j - x^j}{h})^l \Delta\tilde{\Phi}_{-j}(X_{it})'(I_{s(\kappa)} + R_{\theta n})\frac{1}{N}\Delta\Phi'[\Delta g - \Delta'\theta + \Delta\epsilon + \Delta\mu(\hat{\gamma}) - \Delta\mu(\gamma_0)] \\
& = -[D_{21l} + D_{22l} + D_{23l} + (D_{21lR} + D_{22lR} + D_{23lR}) + D_{24l}].
\end{aligned}
$$

(A) Claim: $\sqrt{nh}h^l D_{22l} = o_p(1)$, for $\sqrt{nh} D_{22l} = \frac{1}{T-1}\sum_{t=2}^{T}\sqrt{nh}\frac{1}{nh}\sum_{i=1}^{n} K_{it,j}(\frac{X_{it}^j - x^j}{h})^l \Delta\tilde{\Phi}_{-j}(X_{it})' \frac{1}{N}\Delta\Phi(X_{m\tau})\Delta\epsilon_{m\tau}$. Let $d_{22l}(x^j) = \frac{1}{nh}\sum_{i=1}^{n} K_{it,j}(\frac{X_{it}^j - x^j}{h})^l \Delta\tilde{\Phi}_{-j}(X_{it}) = \{d_{22l,k}\}_{k=1}^{s(\kappa)}$, where $d_{22l,k} = \frac{1}{nh}\sum_{i=1}^{n} K_{it,j}(\frac{X_{it}^j - x^j}{h})^l \Delta\tilde{\Phi}_{-j,k}(X_{it})$ with $\Delta\tilde{\Phi}_{-j,k}(X_{it})$ being the $k-th$ element in $\Delta\tilde{\Phi}_{-j}(X_{it})$. With assumptions C1, C3(2), (3) and C4(1), we can apply Lemma 6 to obtain that $\sup_{x^j \in [-1,1]} |d_{22l,k} - Ed_{22l,k}| = O_p(\frac{Lnn}{nh})^{1/2})$. $\forall m \in \{1, \cdots, d\}$, $m \neq j$ and $k' = 1, \cdots, \kappa$, $k = (m-1)\kappa + k'$ for some m and k'. So for these k, we have $Ed_{220,k}(x^j) = \int \phi_{k'}(X_{it}^m) f_{X_t^j, X_t^m}(x^j, X_{it}^m) dX_{it}^m - \int \phi_{k'}(X_{it-1}^m) f_{X_t^j, X_{t-1}^m}(x^j, X_{it-1}^m) dX_{it-1}^m + O(h^2) = d_{2200,kA} + R_{2220,kA}$, $Ed_{221,k}(x^j) = h\int \phi_{k'}(X_{it}^m)\partial_{X_t^j} f_{X_t^j, X_t^m}(x^j, X_{it}^m) dX_{it}^m - \int \phi_{k'}(X_{it-1}^m)\partial_{X_t^j} f_{X_t^j, X_{t-1}^m}(x^j, X_{it-1}^m) dX_{it-1}^m \mu_{k,2} + O(h^2) = R_{2221,kA}$. □

Lemma 1. *For* $\hat{\mathcal{Q}} = \frac{1}{N}\Delta\Phi'\Delta\Phi$ *and with assumption A1(1), A3(4) and A3(5), we have* $||\hat{\mathcal{Q}} - I_{S(\kappa)}|| = o_p(1)$.

Proof. Define $\mathcal{Q} = E[\Delta\Phi(X_{it})\Delta\Phi(X_{it})']$. We follow Newey (1997) to normalize $\mathcal{Q} = I_{S(\kappa)}$ for all t because $\Delta\Phi(X_{it})'\hat{\theta}$ is invariant to nonsingular transformation of $\Delta\Phi(X_{it})$ and by assumption A3(5), $\mathcal{Q}_t$ has smallest eigenvalue bounded away from zero.

Denoting the $j-th$ element in $\Delta\Phi(X_{it})$ by $\Delta\Phi_j(X_{it})$, $I_{(jk)} = 1$ if $j = k$ and $I_{(jk)} = 0$ otherwise,

$$
\begin{aligned}
& E||\hat{\mathcal{Q}} - I_{S(\kappa)}||^2 = E(tr(\hat{\mathcal{Q}} - I_{S(\kappa)})'(\hat{\mathcal{Q}} - I_{S(\kappa)})) = Etr[\tfrac{\Delta\Phi'\Delta\Phi}{N}\tfrac{\Delta\Phi'\Delta\Phi}{N} - 2\tfrac{\Delta\Phi'\Delta\Phi}{N} + I_{S(\kappa)}] \\
= \; & E\sum_{k=1}^{S(\kappa)}\sum_{j=1}^{S(\kappa)}[(\tfrac{1}{N}\sum_{i=1}^{n}\sum_{t=2}^{T}\Delta\Phi_k(X_{it})\Delta\Phi_j(X_{it}))^2 - \tfrac{2}{N}\sum_{i=1}^{n}\sum_{t=2}^{T}\Delta\Phi_k(X_{it})\Delta\Phi_j(X_{it})I_{(jk)} + I_{(jk)}] \\
= \; & \sum_{k=1}^{S(\kappa)}\sum_{j=1}^{S(\kappa)}[\tfrac{1}{N^2}\sum_{i=1}^{n}\sum_{t=2}^{T}E(\Delta\Phi_k(X_{it})\Delta\Phi_j(X_{it}))^2 \\
& + \tfrac{1}{N^2}\sum_{i=1}^{n}\sum_{t=2}^{T}\sum_{\substack{i'=1 \\ (i,t)\neq(i',t')}}^{n}\sum_{t'=1}^{T-1}E(\Delta\Phi_k(X_{it})\Delta\Phi_j(X_{it}))E(\Delta\Phi_k(X_{i't'})\Delta\Phi_j(X_{i't'})) - I_{(jk)}]
\end{aligned}
$$

The second term is not zero only when $j = k$ and by the normalization above, it is $\frac{N-1}{N}$, so

$$
\begin{aligned}
& E||\hat{\mathcal{Q}} - I_{S(\kappa)}||^2 \\
\leq \; & \sum_{k=1}^{S(\kappa)}\sum_{j=1}^{S(\kappa)}\tfrac{1}{N^2}\sum_{i=1}^{n}\sum_{t=2}^{T}E(\Delta\Phi_k(X_{it})\Delta\Phi_j(X_{it}))^2 \\
= \; & \sum_{k=1}^{S(\kappa)}\sum_{j=1}^{S(\kappa)}\tfrac{1}{N(T-1)}\sum_{t=2}^{T}E(\Delta\Phi_k(X_{it})\Delta\Phi_j(X_{it}))^2 \text{ by A1(1)} \\
= \; & \tfrac{1}{N(T-1)}\sum_{t=2}^{T}E[\sum_{k=1}^{S(\kappa)}(\Delta\Phi_k(X_{it}))^2\sum_{j=1}^{S(\kappa)}\Delta\Phi_j(X_{it})] \\
\leq \; & \xi_{S(\kappa)}^2\tfrac{1}{N(T-1)}\sum_{t=2}^{T}tr(I_{S(\kappa)}) \text{ by A3(4) and the normalization} \\
= \; & \tfrac{1}{N}\xi_{S(\kappa)}^2 S(\kappa) \to 0 \text{ by A3(4) again.}
\end{aligned}
$$

So we have $||\hat{\mathcal{Q}} - I_{S(\kappa)}|| = O_p(\frac{1}{\sqrt{N}}\xi_{S(\kappa)}S^{\frac{1}{2}}(\kappa)) = o(1)$.

□

Lemma 2. *Let $\hat{B}_\xi = (\Delta\Phi'\Delta\Phi)^{-1}\Delta\Phi'\xi$, for $\xi = \{m_j(X_{it}^j)\}_{t=1}^{T}{}_{,i=1}^{n}$, or $\xi = \{H_{jl}(X_{it}^j, X_{it}^l)\}_{t=1}^{T}{}_{,i=1}^{n}$, $B_\xi = B_{m_j}$, or $B_\xi = B_{H_{jl}}$, $\hat{B}_h = (\Delta\Phi'\Delta\Phi)^{-1}\Delta\Phi' h$, and $h = \{h(X_{it}, X_{it-1};\gamma)\}_{t=1}^{T}{}_{,i=1}^{n}$, we have $||\hat{B}_\xi - B_\xi|| = O_p(d\kappa^{-\delta_1} + d_1\kappa^{-2\delta_2})$, and $\sup_{\gamma\in\Gamma}||\hat{B}_h - B_h|| = O_p(d\kappa^{-\delta_1} + d_1\kappa^{-2\delta_2})$.*

Proof. (1) Define $1_N = I(\lambda_{min}(\hat{\mathcal{Q}}) > \frac{1}{2})$, we claim that $1_N \overset{p}{\to} 1$.

With the inequality (Weyl's Theorem) that $\lambda_{min}(I_{S(\kappa)} - \mathcal{Q}) \leq \lambda_{min}(I_{S(\kappa)}) - \lambda_{min}(\mathcal{Q}) \leq \lambda_{max}(I_{S(\kappa)} - \mathcal{Q})$, where $\lambda_{min}(A)$ and λ_{max} refer to the smallest and largest eigenvalues of matrix A, respectively. With $||A||_{sp}$ being the spectrum norm, we apply Theorem 5.6.9 in Horn and Johnson (2013) to have $|\lambda_{min}(I_{S(\kappa)} - \mathcal{Q})| \leq ||I_{S(\kappa)}) - \lambda_{min}(\mathcal{Q}||_{sp}$, and $|\lambda_{max}(I_{S(\kappa)} - \mathcal{Q})| \leq ||I_{S(\kappa)}) - \lambda_{min}(\mathcal{Q}||_{sp}$. Because $||A||_{sp} \leq ||A||$, and Lemma 1 that $||I_{S(\kappa)}) - \lambda_{min}(\mathcal{Q}|| = o_p(1)$, we have $\lambda_{min}(\mathcal{Q}) = \lambda_{min}(I_{S(\kappa)}) + o_p(1) = 1 + o_p(1)$. So $1_N \overset{p}{\to} 1$.

(2) Consider the matrix product of $A_nB_nC_n$ which is a scalar, C_nA_n exists and is PSD (positive semidefinite), and B_n is a $n \times n$ matrix such that $B_n - CI_n$ is ND (negative definite) for some $C > 0$. Then we claim that $A_nB_nC_n \leq Ctr(A_nC_n)$.

To show the claim, note $A_nB_nC_n = tr(A_nB_nC_n) = tr(C_nA_nB_n) = tr(C_nA_n(B_n - CI_n)) + Ctr(C_nA_n)$.

Given that C_nA_n is PSD, $B_n - CI_n$ is ND, we have $tr(C_nA_n(B_n - CI_n)) \leq 0$, which gives the claimed result.

(3) With $1_N = I(\lambda_{min}(\hat{\mathcal{Q}}) > \frac{1}{2}) = I((\lambda_{min}(\hat{\mathcal{Q}}))^{-1} = \lambda_{max}(\mathcal{Q}^{-1}) < 2) = I(\mathcal{Q}^{-1} - 2I_{S(\kappa)} \text{ is ND}) \xrightarrow{p} 1$, so $1_N||\hat{B}_\xi - B_\xi|| = 1_N||(\Delta\Phi'\Delta\Phi)^{-1}\Delta\Phi'(\xi - \Delta\Phi B_\xi)|| = 1_N\{(\xi - \Delta\Phi B_\xi)'\Delta\Phi(\Delta\Phi'\Delta\Phi)^{-1}(\frac{\Delta\Phi'\Delta\Phi}{N})^{-1}\Delta\Phi'(\xi - \Delta\Phi B_\xi)\frac{1}{N}\}^{1/2} = O_p(1)1_N\{(\xi - \Delta\Phi B_\xi)'\Delta\Phi(\Delta\Phi'\Delta\Phi)^{-1}\Delta\Phi'(\xi - \Delta\Phi B_\xi)\frac{1}{N}\}^{1/2}$. Here we use (2), with $A_n = (\xi - \Delta\Phi B_\xi)'\Delta\Phi(\Delta\Phi'\Delta\Phi)^{-1}$, $B_n = (\frac{\Delta\Phi'\Delta\Phi}{N})^{-1}$, and $C_n = \Delta\Phi'(\xi - \Delta\Phi B_\xi)\frac{1}{N}$. C_nA_n is PSD as it can be written as a product of two matrices, each being PSD. With 1_n, we have $B_n - 2I_{S(\kappa)} = \mathcal{Q}^{-1} - 2I_{S(\kappa)}$ being ND. Furthermore, because $P_{\Delta\Phi}$ is an idempotent projection matrix, we have $1_N||\hat{B}_\xi - B_\xi|| = O_p(1)1_N\{(\xi - \Delta\Phi B_\xi)'P_{\Delta\Phi}(\xi - \Delta\Phi B_\xi)\frac{1}{N}\}^{1/2} \leq O_p(1)\{(\xi - \Delta\Phi B_\xi)'(\xi - \Delta\Phi B_\xi)\frac{1}{N}\}^{1/2} = O_p(d\kappa^{-\delta_1} + d_1\kappa^{-2\delta_2})$ with assumption A3 (2). So we have $||\hat{B}_\xi - B_\xi|| = O_p(d\kappa^{-\delta_1} + d_1\kappa^{-2\delta_2})$. The claim that $\sup_{\gamma\in\Gamma}||\hat{B}_h - B_h|| = O_p(d\kappa^{-\delta_1} + d_1\kappa^{-2\delta_2})$ follows similarly with the assumption A3(1). □

Lemma 3. *With h and ξ as defined in Lemma 2, we assume the assumptions in Lemma 1 and Lemma 2, together with A3(1) and A3(2). Then (i)* $\sup_{\gamma\in\Gamma}\frac{1}{N}h'(I_N - P_{\Delta\Phi})h = O_p((d\kappa^{-\delta_1} + d_1\kappa^{-2\delta_2})^2)$. *(ii)* $\frac{1}{N}\xi'(I_N - P_{\Delta\Phi})\xi = O_p((d\kappa^{-\delta_1} + d_1\kappa^{-2\delta_2})^2)$.

Proof. (i) Because $I_N - P_{\Delta\Phi}$ is idempotent and symmetric, we have $\frac{1}{N}h'(I_N - P_{\Delta\Phi})h = \frac{1}{N}||h - P_{\Delta\Phi}h||^2 = \frac{1}{N}||h - \Delta\Phi\hat{B}_h||^2 = \frac{1}{N}||h - \Delta\Phi B_h + \Delta\Phi B_h - \Delta\Phi\hat{B}_h||^2 \leq \frac{1}{N}||h - \Delta\Phi B_h||^2 + \frac{1}{N}||\Delta\Phi B_h - \Delta\Phi\hat{B}_h||^2$. By assumption A3(1), $\sup_{\gamma\in\Gamma}\frac{1}{N}||h - \Delta\Phi B_h||^2 = O_p((d\kappa^{-\delta_1} + d_1\kappa^{-2\delta_2})^2)$. $\frac{1}{N}||\Delta\Phi B_h - \Delta\Phi\hat{B}_h||^2 = (B_h - \hat{B}_h)'(\frac{\Delta\Phi'\Delta\Phi}{N} - I_{S(\kappa)})(B_h - \hat{B}_h) + (B_h - \hat{B}_h)'(B_h - \hat{B}_h) \leq ||B_h - \hat{B}_h||^2||\frac{\Delta\Phi'\Delta\Phi}{N} - I_{S(\kappa)}|| + ||B_h - \hat{B}_h||^2$, since $||AB|| \leq ||A||.||B||$. So we have by Lemma 1 and Lemma 2 that $\sup_{\gamma\in\Gamma}\frac{1}{N}||\Delta\Phi B_h - \Delta\Phi\hat{B}_h||^2 = \sup_{\gamma\in\Gamma}||B_h - \hat{B}_h||^2(1 + o_p(1)) = O_p((d\kappa^{-\delta_1} + d_1\kappa^{-2\delta_2})^2)$. Thus, $\sup_{\gamma\in\Gamma}\frac{1}{N}h'(I_N - P_{\Delta\Phi})h = O_p((d\kappa^{-\delta_1} + d_1\kappa^{-2\delta_2})^2)$.

(ii) The claim here follows with similar arguments and assumption A3(2). □

Lemma 4. *With assumption A, we have*

(i) $\frac{1}{N}(\Delta\epsilon)'P_{\Delta\Phi}\Delta\epsilon = o_p(1)$.

(ii) $\sup_{\gamma\in\Gamma}\frac{1}{N}(\Delta\mu(\gamma) - E_{\mathcal{G}}(E(\Delta\mu(\gamma)|X)))'P_{\Delta\Phi}(\Delta\mu(\gamma) - E_{\mathcal{G}}(E(\Delta\mu(\gamma)|X))) = o_p(1)$.

Proof. (i) Recall $X = \{X_{it}\}_{t=1,}^{T}{}_{i=1}^{n}$, $E(\frac{1}{N}(\Delta\epsilon)'P_{\Delta\Phi}\Delta\epsilon|X) = \frac{1}{N}tr(P_{\Delta\Phi}E((\Delta\epsilon)(\Delta\epsilon)'|X))$. With A1, we know that $E((\Delta\epsilon)(\Delta\epsilon)'|X)$ is an block diagonal matrix. The $i - th$ diagonal block is an $(T-1)\times(T-1)$ matrix

A with its $(t,\tau)-th$ element being $E(\Delta\epsilon_{it}\Delta\epsilon_{i\tau}|X_i) \le E^{1/2}((\Delta\epsilon_{it})^2|X_i)E^{1/2}((\Delta\epsilon_{i\tau})^2|X_i)$. By $c-r$ inequality, $E((\Delta\epsilon_{it})^2|X_i) \le C(E((\epsilon_{it})^2|X_i) + E((\epsilon_{it-1})^2|X_i))$. Furthermore, $E((\epsilon_{it})^2|X_i) \le C[\sigma_v^2 + E(\tilde{u}_{it}^2|X_i)] \le C[\sigma_v^2 + E(u_{it}^2|X_i) + E(\mu^2(Z_{it};\gamma_0)|X_i)] < C$ by assumption A1(4). So each element in A is $a_{t\tau,i}$ and $|a_{t\tau,i}| < C$ $a.s.$. Then there is an $\epsilon > 0$ such that $C' = C + \epsilon$ and $C'(T-1)I_{T-1} - A$ is a positive definite (PD) matrix $a.s.$. To show it, consider an arbitrary $(T-1)\times 1$ nonzero constant vector λ, we have $\lambda_t\lambda_\tau \le \frac{1}{2}(x_t^2 + x_\tau^2)$, so $\lambda'(C'(T-1)I_{T-1} - A)\lambda = C'(T-1)\sum_{t=2}^{T} x_t^2 - \sum_{t=2}^{T}\sum_{\tau=2}^{T} a_{t\tau,i}\lambda_t\lambda_\tau \ge C'(T-1)\sum_{t=2}^{T} x_t^2 - \frac{C}{2}\sum_{t=2}^{T}\sum_{\tau=2}^{T}(\lambda_t^2 + \lambda_\tau^2) = \epsilon(T-1)\sum_{t=2}^{T} x_t^2 > 0$. So we have $C'(T-1)I_N - E((\Delta\epsilon)(\Delta\epsilon)'|X)$ is PD $a.s.$ by the arguments above. So $E(\frac{1}{N}(\Delta\epsilon)'P_{\Delta\Phi}\Delta\epsilon|X) = \frac{1}{N}tr[P_{\Delta\Phi}(E((\Delta\epsilon)(\Delta\epsilon)'|X) - C'(T-1)I_N)] + \frac{1}{N}tr(P_{\Delta\Phi}C'(T-1)I_N) \le \frac{1}{N}tr(P_{\Delta\Phi}C'(T-1)I_N) = \frac{C'(T-1)}{N}tr(P_{\Delta\Phi}) = \frac{C'(T-1)S(\kappa)}{N}$. So by A3(4), we have the claim that $\frac{1}{N}(\Delta\epsilon)'P_{\Delta\Phi}\Delta\epsilon = O_p(\frac{S(\kappa)}{N}) = o_p(1)$.

(ii) With arguments similar to Lemma 2 (3), we have $\sup_{\gamma\in\Gamma}\frac{1}{N}(\Delta\mu(\gamma) - E_{\mathcal{G}}(E(\Delta\mu(\gamma)|X)))'P_{\Delta\Phi}(\Delta\mu(\gamma) - E_{\mathcal{G}}(E(\Delta\mu(\gamma)|X))) \le O_p(1)\sup_{\gamma\in\Gamma}[\frac{1}{N}(\Delta\mu(\gamma) - E_{\mathcal{G}}(E(\Delta\mu(\gamma)|X)))'\Delta\Phi][\frac{1}{N}(\Delta\Phi)'(\Delta\mu(\gamma) - E_{\mathcal{G}}(E(\Delta\mu(\gamma)|X)))]$.

The claim in (ii) follows if we show $\sup_{\gamma\in\Gamma}||\frac{1}{N}(\Delta\Phi)'(\Delta\mu(\gamma) - E_{\mathcal{G}}(E(\Delta\mu(\gamma)|X)))|| = o_p(1)$. We write

$$
\begin{aligned}
& ||\tfrac{1}{N}(\Delta\Phi)'(\Delta\mu(\gamma) - E_{\mathcal{G}}(E(\Delta\mu(\gamma)|X)))||^2 \\
= & \tfrac{1}{N^2}\sum_{i=1}^{n}\sum_{t=2}^{T}\sum_{\tau=2}^{T}(\Delta\Phi(X_{it}))'(\Delta\Phi(X_{i\tau}))\Delta_1\mu_{it}(\gamma)\Delta_1\mu_{i\tau}(\gamma) \\
& + \tfrac{1}{N^2}\sum_{i=1}^{n}\sum_{\substack{j=1 \\ j\neq i}}^{n}\sum_{t=2}^{T}\sum_{\tau=2}^{T}(\Delta\Phi(X_{it}))'(\Delta\Phi(X_{j\tau}))\Delta_1\mu_{it}(\gamma)\Delta_1\mu_{j\tau}(\gamma) \\
= & T_1(\gamma) + T_2(\gamma).
\end{aligned}
$$

$$E|T_1(\gamma)| \le \sum_{t=2}^{T}\sum_{\tau=2}^{T}\frac{1}{N}E|(\Delta\Phi(X_{it}))'(\Delta\Phi(X_{i\tau}))\Delta_1\mu_{it}(\gamma)\Delta_1\mu_{i\tau}(\gamma)|.$$

By Cauchy Schwartz inequality, $(\Delta\Phi(X_{it}))'(\Delta\Phi(X_{i\tau})) \le ||\Delta\Phi(X_{it})||.||\Delta\Phi(X_{i\tau})|| \le 4(d+d_1)\xi_{S(\kappa)}^2$ by assumption A3(4). So $\sup_{\gamma\in\Gamma}E|T_1(\gamma)| \le 4(d+d_1)\xi_{S(\kappa)}^2\sum_{t=2}^{T}\sum_{\tau=2}^{T}\frac{1}{N}\sup_{\gamma\in\Gamma}E|\Delta_1\mu_{it}(\gamma)\Delta_1\mu_{i\tau}(\gamma)| = O(\frac{\xi_{S(\kappa)}^2}{n})$ by assumption A2(5).

$E|T_2(\gamma)| = \frac{1}{N^2}\sum_{i=1}^{n}\sum_{\substack{j=1 \\ j\neq i}}^{n}\sum_{t=2}^{T}\sum_{\tau=2}^{T}E[(\Delta\Phi(X_{it}))'\Delta_1\mu_{it}(\gamma)]E[(\Delta\Phi(X_{j\tau}))\Delta_1\mu_{j\tau}(\gamma)] = 0$, since $\Delta\Phi(X_{it}) \in \mathcal{G}$ and $\Delta_1\mu_{it}(\gamma) \in \mathcal{G}^{\perp}$. So in all, we have the claim that $\sup_{\gamma\in\Gamma}||\frac{1}{N}(\Delta\Phi)'(\Delta\mu(\gamma) - E_{\mathcal{G}}(E(\Delta\mu(\gamma)|X)))|| = o_p(1)$.

□

Lemma 5. *Let $S_n(\gamma) = \frac{1}{n}\sum_{i=1}^{n}G_n(Z_i,\gamma)$, where $EG_n(Z_i,\gamma) = 0$. We assume that (i) $|G_n(Z_i,\gamma) - G_n(Z_i,\gamma')| \le B_n(Z_i)||\gamma - \gamma'||$, where $EB_n(Z_i) < \infty$; (ii) Z_i is i.i.d. across i; (iii) $G_n(Z_i,\gamma)$ satisfies the Cramer's condi-*

tion. Then for $\gamma \in \Gamma$*, a compact set of* $\Re^p$, $\sup_{\gamma\in\Gamma} |S_n(\gamma) - ES_n(\gamma)| = O_p((\frac{lnn}{n})^{\frac{1}{2}})$.

References

Aigner, D., Lovell, C. A. K., Schmidt, P., 1977. Formulation and estimation of stochastic frontiers production function models. Journal of Econometrics 6, 21–37.

Amemiya, T., 1985. Advanced Econometrics. Harvard University Press, Cambridge, Massachusetts.

Baltagi, B. H., Li, D., 2002. Series estimation of partially linear panel data models with fixed effects. Annals of Economics and Finance 3, 103–116.

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

Coelli, T., 1995. Estimators and hypothesis tests for a stochastic frontier function. Journal of Productivity Analysis 6, 247–268.

Cornwell, C., Schmidt, P., Sickles, R. C., 1990. Production frontiers with cross-sectional and time-series variation in efficiency levels. Journal of Econometrics 46 (2), 185–200.

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

Fan, Y., Li, Q., Weersink, A., 1996. Semiparametric estimation of stochastic production frontier models. Journal of Business and Economic Statistics 14, 460–468.

Greene, W. H., 1993. The econometric approach to efficiency analysis. In: In: H. Fried, C. A. K. Lovell, and S. S. Schmidt, (Eds.), The Measurement of Productive Efficiency. Oxford University Press, Oxford.

Horn, R. A., Johnson, C. R., 2013. Matrix Analysis, second edi Edition. Cambridge University Press, New York.

Horowitz, J. L., Mammen, E., 2004. Nonparametric estimation of an additive model with a link function. Annals of Statistics 32 (6), 2412–2443.

Horrace, W. C., Parmeter, C. F., 2011. Semiparametric deconvolution with unknown error variance. Journal of Productivity Analysis 35 (2), 129–141.

Kumbhakar, S. C., Lovell, C. A. K., 2000. Stochastic frontier analysis. Cambridge university Press, Cambridge, UK.

Kumbhakar, S. C., Park, B. U., Simar, L., Tsionas, E., 2007. Nonparametric stochastic frontiers: a local maximum likelihood approach. Journal of Econometrics 137, 1–27.

Kumbhakar, S. C., Wang, H.-J., Horncastle, A. P., 2015. A Practitioner's Guide to Stochastic Frontier Analysis Using Stata. Cambridge University Press.

Lee, Y., Schmidt, P., 1993. A production frontier model with flexible temporal variation in technical efficiency. In: in L. H. Fried & S. Schmidt, eds, 'The Measurement of Productive Efficiency'. Oxford University Press, Oxford, United Kingdom.

Li, Q., 2000. Efficient estimation of additive partially linear models. International Economic Review 41, 1073–1092.

Li, Q., Racine, J., 2007. Nonparametric econometrics: theory and practice. Princeton University Press, Princeton, NJ.

Martins-Filho, C., Yao, F., 2015. Nonparametric stochastic frontier estimation via profile likelihood. Econometric Reviews 34 (4), 413–451.

Martins-Filho, C., Yao, F., Torero, M., 2018. Nonparametric estimation of conditional value-at-risk and expected shortfall based on extreme value theory. Econometric Theory 34 (1), 23–67.

Meeusen, W., van den Broeck, J., 1977. Efficiency estimation from cobb-douglas production functions with composed error. International Economic Review 18, 435–444.

Newey, W., jul 1997. Convergence rates and asymptotic normality for series estimators. Journal of Econometrics 79, 147–168.
URL http://linkinghub.elsevier.com/retrieve/pii/S0304407697000110

Ozabaci, D., Henderson, D. J., Su, L., 2014. Additive Nonparametric Regression in the Presence of Endogenous Regressors. Journal of Business and Economic Statistics 32 (4), 555–575.

Park, B. U., Simar, L., Zelenyuk, V., 2015. Categorical data in local maximum likelihood: theory and applications to productivity analysis. Journal of Productivity Analysis 43, 199–214.

Parmeter, C. F., Kumbhakar, S. C., 2014. Efficiency Analysis: A Primer on Recent Advances. Foundations and Trends® in Econometrics 7 (3-4), 191–385.

Parmeter, C. F., Wang, H.-J., Kumbhakar, S. C., 2014. Nonparametric estimation of the determinants of inefficiency. Department of Economics, University of Miami, Working Paper Series.

Qian, J., Wang, L., 2012. Estimating semiparametric panel data models by marginal integration. Journal of Econometrics 167 (2), 483–493.

Schmidt, P., Sickles, R. C., 1984. Production frontiers and panel data. Journal of Business & Economic Statistics 2 (2), 367–374.

Severini, T., Wong, W. H., 1992. Profile likelihood and conditionally parametric models. Annals of Statistics 20, 1768–1802.

Simar, L., Wilson, P., 2008. Statistical inference in nonparametric frontier models: recent developments and perspectives. In: Fried, H., Lovell, C. A. K., Schmidt, S. S. (Eds.), The Measurement of Productive Efficiency. Oxford University Press, Oxford, UK, Ch. Chapter 4.

Sperlich, S., Tjostheim, D., Yang, L., 2002. Nonparametric Estimation and Testing of Interaction in Additive Models. Econometric Theory 18, 197–251.

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

Sun, Y., Carroll, R. J., Li, D., 2009. Semiparametric estimation of fixed effects panel data varying coefficient models. Advances in Econometrics 25, 101–130.

Yao, F., Zhang, F., Kumbhakar, S. C., 2019. Semiparametric smooth coefficient stochastic frontier model with panel data. Journal of Business & Economic Statistics 37 (3), 556–572.