Identification and the Influence Function of Olley and Pakes' (1996) Production Function Estimator

Jinyong Hahn* — UCLA
Zhipeng Liao† — UCLA
Geert Ridder‡ — USC

First Version: March 2020; This Version: March 2021

Abstract

In this paper, we reconsider the assumptions that ensure the identification of the production function in Olley and Pakes (1996). We show that an index restriction plays a crucial role in the identification, especially if the capital stock is measured by the perpetual inventory method. The index restriction is not sufficient for identification under sample selectivity. The index restriction makes it possible to derive the influence function and the asymptotic variance of Olley-Pakes estimator.

JEL Classification: C13, C14, C31

Keywords: Identification; Index Restriction; Influence Function; Perpetual Inventory Method; Production Function Estimation

*Department of Economics, UCLA, Los Angeles, CA 90095-1477 USA. Email: hahn@econ.ucla.edu

†Department of Economics, UCLA, Los Angeles, CA 90095-1477 USA. Email: zhipeng.liao@econ.ucla.edu

‡Department of Economics, University of Southern California, Los Angeles, CA 90089. Email: ridder@usc.edu.

1 Introduction

Production functions are a central component of economics. For that reason, their estimation has a long history in applied econometrics. To our knowledge, the most prominent estimator used in modern empirical analysis is due to Olley and Pakes (1996, OP hereafter).[1]

The econometric analysis of the OP estimator is a challenge, and a correct asymptotic variance is currently not available.[2] Pakes and Olley (1995) derive an expression for the variance matrix. However their derivation does not address the generated regressor problem correctly, because they ignore the variability of the conditional expectation given the generated regressor (see their (28a)). Their asymptotic variance formula is therefore incorrect. [3] The OP estimator is a two-step estimator. The first step is a partially linear regression, in which the output elasticity of the variable production factor labor, and a non-parametric index that captures the contribution of capital and factor neutral productivity to log output are estimated. The second step in which the productivity of capital is estimated, is a variant of a partial linear regression as described in Section 2.

The OP estimator has some similarity to the class of estimators considered by Hahn and Ridder (2013, HR hereafter), although there is an important difference. In both HR and OP's first step, a variable is estimated and is used as a generated regressor in the second step. The second step is in the case of OP a variant of a partial linear regression, and in the case of HR a non-parametric regression with the generated regressor as an independent variable. In HR, the last step involves a moment that is a known functional of the second step non-parametric regression. The second step in OP can be thought of as having two sub-steps: (i) the estimation of a non-parametric function by partial linear regression treating the coefficient on capital as

[1]Ackerberg et al. (2007) discuss the innovation that OP introduced in production function estimation, and Ackerberg, Caves and Frazer (2015) give a partial list of the many applications of the estimator.

[2]The challenge in characterizing the influence function is due to the semiparametric estimation in the second step of OP. The difficulty disappears if the second step is completely parametric, which is *not* the specification in OP. Cattaneo, Jansson and Ma (2018) adopt such a parametric specification in their second step. The influence function can then be derived as a straightforward application of Newey (1994). Their primary contribution is therefore the analysis and characterization of the higher order bias for a fully parametric specification.

[3]Another attempt at a correct asymptotic variance of the OP estimator is in the working paper by Mammen, Rothe, and Schienle (2014). Inspection of their proof reveals that the derivation is not complete. In particular, the derivative of the key conditional expectation with respect to the capital coefficient is mentioned on p. 32, but an expression for this derivative that appears in the influence function is not provided. This derivative is not obvious (see our Proposition 1) and is the reason that the derivation of the OP influence function is not just an application of the three-step semiparametric framework in Hahn and Ridder (2013) and Mammen, Rothe, and Schienle (2016). The published version omits the proposed asymptotic variance of the OP estimator.

known and with the generated regressor as independent variable, and (ii) the estimation of the capital coefficient as the solution to the first-order condition of a non-linear least squares problem assuming the function estimated in (i) is known. Because the first-order condition in (ii) depends on the function in (i) and in addition, the capital coefficient also appears in the non-parametric function in (i) OP does not directly fit into the HR framework. The step (ii) is more complicated, than the final step in HR, and requires special attention.[4]

In practice, the standard error of the OP estimator can be calculated without an explicit expression for the asymptotic variance if some regularity conditions are satisfied, and if the nonparametric regressions in the OP procedure are estimated using the method of sieves. Hahn, Liao and Ridder (2018, HLR hereafter) show that under these assumptions the standard error of the OP estimator can be calculated as if the finite dimensional sieve approximation is in fact exact, i.e., as the standard error of a parametric estimator.

Despite the convenience, HLR's result is useless when nonparametric estimation is done by local methods, such as kernel estimation. It is therefore useful to have an explicit characterization of the asymptotic variance. Moreover, HLR is predicated under the assumption that the parameters are (locally) identified by the moments that OP use. One of the contributions of this paper is that we verify the local identification and find that the output elasticity of capital is only identified if an index/conditional independence assumption[5] holds that is implicit in OP. The index restriction also makes it possible to derive the asymptotic variance. We show that the index restriction is not necessary for identification if the capital stock is measured directly and not by the perpetual inventory method (PIM). If plants can close down, then the index restriction is not sufficient for the identification of the production function and the survival probability.

The rest of the paper is organized as follows. In Section 2, we discuss the identification of the production function and the implicit index restriction. Section 3 shows that identification depends on how the capital stock is measured. We also consider identification of the production function and the survival probability, if plants can close down. In Section 4, we derive the influence function of the OP estimator. Section 5 concludes. The Appendix offers proofs of the main results in the paper. Additional theoretical results are in the Supplemental Appendix to this paper.

[4]See more discussion on this in Section 4.

[5]This assumption seems to be known to IO economists, but not to econometricians.

2 Identification of the Production Function and the Index Restriction

In this section, we review and discuss the production function estimator developed by OP. We argue that given their other assumptions, one particular additional assumption is necessary for the identification of the productivity of capital. This assumption has not received much attention from econometricians. The assumption was called the first order Markov assumption in Ackerberg, Caves and Frazer (2015, p.2416), although econometricians would call it a conditional independence or index restriction.[6] We will discuss its necessity for identification in this section, and its implication for the influence function and hence the asymptotic variance of OP's estimator in Section 4. For simplicity, we will begin with the case that plants survive forever and next consider identification if plants can close down and do so selectively.

2.1 Model and Estimator

We will begin with the description of OP's model. We simplify their model by omitting the age of the plant.[7] The production function takes the form[8]

$$y_{t,i} = \beta_0 + \beta_{k,0} k_{t,i} + \beta_{l,0} l_{t,i} + \omega_{t,i} + \eta_{t,i}, \tag{OP 6}$$

where $y_{t,i}$ is the log of output from plant i at time t, $k_{t,i}$ the log of its capital stock, $l_{t,i}$ the log of its labor input, $\omega_{t,i}$ its productivity, and $\eta_{t,i}$ is either measurement error or a shock to productivity which is not forecastable. Both $\omega_{t,i}$ and $\eta_{t,i}$ are unobserved, and they differ from each other in that $\omega_{t,i}$ is a state variable in the firm's decision problem, while $\eta_{t,i}$ is not. To keep the notation simple, we will omit the i subscript below when obvious.

It is assumed that

$$k_{t+1} = (1-\delta)\, k_t + i_t. \tag{OP 1}$$

This is the perpetual inventory method (PIM) of capital stock measurement as discussed on p.1295 of OP. It requires only an initial estimate of the capital stock and investment data. It assumes that the depreciation rate is the same across plants and over time. We discuss its implications for identification of $\beta_{k,0}$ in Section 3. A second assumption is that

$$i_t = i_t\left(\omega_t, k_t\right). \tag{OP 5}$$

[6] OP themselves did not name the assumption.

[7] We will present only the most salient aspects of their model and estimation strategy. See OP for details.

[8] (OP 6) is equation (6) in OP with the variable age of the plant omitted.

with $i_t(\omega_t, k_t)$ monotonically increasing in ω_t for all k_t (OP, p. 1274). The investment choice follows from the Bellman equation

$$V_t(\omega_t, k_t) = \max\left\{\Phi,\ \sup_{i_t \geq 0}(\pi_t(\omega_t, k_t) - c(i_t) + \beta\mathbb{E}(V_{t+1}(\omega_{t+1}, k_{t+1}|J_t))\right\} \tag{OP 3}$$

where Φ denotes the liquidation value, $\pi_t(\omega_t, k_t)$ is the profit function as a function of the state variables and $c(i_t)$ is the cost of investment, the information at time t, J_t contains at the minimum the state variables ω_t, k_t, and as do OP, we take $J_t = \{\omega_t, k_t\}$. In (OP 3), we can set the liquidation value $\Phi = -\infty$ to ensure that the plant is not liquidated. We shall discuss the model with possible liquidation in Section 3.

By the monotonicity assumption, we can invert (OP 5) and write

$$\omega_t = h_t(i_t, k_t), \tag{OP 7}$$

which allows us to rewrite

$$y_t = \beta_{l,0}l_t + \phi_t(i_t, k_t) + \eta_t, \tag{OP 8}$$

where

$$\phi_t(i_t, k_t) \equiv \beta_0 + \beta_{k,0}k_t + \omega_t = \beta_0 + \beta_{k,0}k_t + h_t(i_t, k_t). \tag{OP 9}$$

The assumption that a firm never liquidates a plant[9] implies by the first expression on p. 1276 of OP , that $g\left(\underline{\omega}_{t+1}(k_{t+1}), \omega_t\right) = \mathbb{E}[\omega_{t+1}|\,\omega_t] + \beta_0 \equiv g(\omega_t)$ (substitute $\underline{\omega}_t(k_t)) = -\infty$). Therefore their equations (11) and (12) can be rewritten[10]

$$\mathbb{E}[y_{t+1} - \beta_{l,0}l_{t+1}|\,k_{t+1}] = \beta_{k,0}k_{t+1} + g(\omega_t), \tag{OP 11}$$

$$y_{t+1} - \beta_{l,0}l_{t+1} = \beta_{k,0}k_{t+1} + g(\phi_t(i_t, k_t) - \beta_{k,0}k_t) + \xi_{t+1} + \eta_{t+1}, \tag{OP 12}$$

where

$$\xi_{t+1} \equiv \omega_{t+1} - \mathbb{E}[\omega_{t+1}|\,\omega_t]. \tag{1}$$

OP's estimator is based on the following multi-step identification strategy:

[9]Because $\underline{\omega}_t(k_t)$ in their equation (4) is understood to be equal to $-\infty$, the P_t in their equation (10) is equal to 1.

[10]In view of the definition of $\phi_t(i_t, k_t)$ in (OP 9), (OP 12) should be written as

$$y_{t+1} - \beta_{l,0}l_{t+1} = \beta_{k,0}k_{t+1} + \tilde{g}(\phi_t(i_t, k_t) - \beta_{k,0}k_t) + \xi_{t+1} + \eta_{t+1}$$

where $\tilde{g}(v) = g(v - \beta_0)$. Since $g(\cdot)$ is nonparametrically specified and β_0 is not of interest, we write $g(\cdot)$ for $\tilde{g}(\cdot)$ for notational simplicity in the rest of the paper.

1. In the first step, $\beta_{l,0}$ and ϕ_t in (OP 8) are identified by standard methods for partially linear models, where β_l and ϕ_t are identified as the solution[11] to

$$\min_{\beta_l,\phi_t} \mathbb{E}\left[(y_t - \beta_l l_t - \phi_t(i_t, k_t))^2\right]. \tag{2}$$

2. The $\beta_{k,0}$ and g in (OP 12) are identified as the solution to

$$\min_{\beta_k,g} \mathbb{E}\left[(y_{t+1} - \beta_{l,0} l_{t+1} - \beta_k k_{t+1} - g(\phi_t(i_t,k_t) - \beta_k k_t))^2\right],$$

where we substitute $\beta_{l,0}$ and $\phi_t(i_t, k_t)$ that were identified in the first step. [12]

2.2 Index restriction

Equation (OP 11) above is a simplified version of equation (11) in OP, where the simplification is due to the fact that we omit the age variable and have no sample selectivity. Except for these simplifications, it is a direct quote from OP. We argue that (i) it should be derived rigorously under the same (but simplified) assumptions as in OP; and (ii) that derivation will uncover an implicit assumption that needs to be made explicit in order to understand the source of identification.

Equation (OP 11) equates a conditional expectation given k_{t+1} to a function of k_{t+1} and $\omega_t = h_t(i_t, k_t)$. Note that the right-hand side (RHS) is not a function of k_{t+1} only, but a function of k_{t+1} and i_t, or equivalently because of the PIM, of k_{t+1} and k_t. Superficially, this would mean that the arguments in the left-hand side (LHS) and the RHS of (OP 11) are not the same, which cannot be mathematically correct. For this purpose, we start with the derivation of the LHS, under the OP's assumptions.

On p.1275, OP state that (OP 11) is "the expectation of $y_{t+1} - \beta_l l_{t+1}$ conditional on information at t". The information at t includes the state variables ω_t, k_t. Therefore, the LHS of (OP 11) must be $\mathbb{E}[y_{t+1} - \beta_{l,0} l_{t+1} | J_t]$. Now consider the RHS. By the monotonicity of investment demand the information at t is equivalent to i_t, k_t. If the capital stock is measured by the PIM, then

$$\mathbb{E}[k_{t+1}|\omega_t, k_t] = \mathbb{E}[k_{t+1}|i_t, k_t] = k_{t+1}. \tag{3}$$

[11] This minimization itself can be understood to consist of two substeps: For given β_l the function is minimized at $\phi_t(i_t, k_t) = \mathbb{E}[y_t | i_t, k_t] - \beta_l \mathbb{E}[l_t | i_t, k_t]$. Substitution and minimization over β_l identifies that parameter. The second step below also has a two-step interpretation.

[12] Because β_k appears both in the linear part and in the nonparametric function, this is not a standard partially linear regression.

By (OP 6)

$$y_{t+1} - \beta_{l,0} l_{t+1} = \beta_0 + \beta_{k,0} k_{t+1} + \omega_{t+1} + \eta_{t+1},$$

so that, if we, as did OP, assume $\mathbb{E}\left[\eta_{t+1} | J_t\right] = 0$,

$$\begin{aligned} \mathbb{E}\left[y_{t+1} - \beta_{l,0} l_{t+1} | J_t\right] &= \beta_0 + \beta_{k,0} \mathbb{E}\left[k_{t+1} | J_t\right] + \mathbb{E}\left[\omega_{t+1} | J_t\right] + \mathbb{E}\left[\eta_{t+1} | J_t\right] \\ &= \beta_0 + \beta_{k,0} k_{t+1} + \mathbb{E}\left[\omega_{t+1} | \omega_t, k_t\right]. \end{aligned} \tag{4}$$

This suggests that (OP 11) should be read as

$$\mathbb{E}\left[y_{t+1} - \beta_{l,0} l_{t+1} | i_t, k_t\right] = \beta_0 + \beta_{k,0} k_{t+1} + \mathbb{E}\left[\omega_{t+1} | i_t, k_t\right] \tag{5}$$

Comparing with (OP 11) we conclude that OP make an additional assumption

$$\beta_0 + \mathbb{E}\left[\omega_{t+1} | i_t, k_t\right] = \beta_0 + \mathbb{E}\left[\omega_{t+1} | \omega_t, k_t\right] = g(\omega_t). \tag{6}$$

This is either an index restriction with ω_t an index for i_t and k_t, or a conditional mean independence assumption.

OP make the conditional independence assumption implicitly in their equation (2). They state that the distribution of ω_{t+1} conditional on the information at t has a distribution function that belongs to the family $F_\omega = \{F(\cdot | \omega), \omega \in \Omega\}$. This is consistent with ω_t being an index or with ω_{t+1} being conditionally independent of k_t given ω_t. The assumption in OP's equation (2) is also made in Ackerberg, Caves and Frazer (2015, p.2416) who call it the first order Markov assumption.

The index restriction plays a crucial role in the identification of β_k. Under a mild full rank condition, $\beta_{l,0}$ and $\phi_t(i_t, k_t)$ are identified by the partial linear regression in the first step of the OP procedure. So we can assume that $\beta_{l,0}$ and $\phi_t(i_t, k_t)$ are known, and examine identification of $\beta_{k,0}$ by (OP 11) in the second step. Suppose that the index/conditional independence restriction (6) is violated. In that case

$$\beta_0 + \mathbb{E}\left[\omega_{t+1} | \omega_t, k_t\right] = g(\omega_t, k_t). \tag{7}$$

There are economic reasons why the evolution of productivity can depend on the capital stock, an example being learning-by-doing.

By (OP 1), (5) and (7), for all β_k

$$\begin{aligned} \mathbb{E}\left[y_{t+1} - \beta_{l,0} l_{t+1} | \omega_t, k_t\right] &= \beta_k k_{t+1} + g(\omega_t, k_t) + (\beta_{k,0} - \beta_k) k_{t+1} \\ &= \beta_k k_{t+1} + g(\omega_t, k_t) + (\beta_{k,0} - \beta_k)((1-\delta) k_t + i_t) \\ &= \beta_k k_{t+1} + g(\omega_t, k_t) + (\beta_{k,0} - \beta_k)((1-\delta) k_t + i_t(\omega_t, k_t)) \\ &= \beta_k k_{t+1} + \bar{g}(\omega_t, k_t) \end{aligned} \tag{8}$$

for $\bar{g}(\omega_t, k_t) \equiv g(\omega_t, k_t) + (\beta_{k,0} - \beta_k)((1-\delta)k_t + i_t(\omega_t, k_t))$. Because both g and $\bar{g}$ are nonparametric, we conclude that $(\beta_{k,0}, g)$ and $(\beta_k, \bar{g})$ are observationally equivalent, so that $\beta_{k,0}$ and g are not identified.

3 Discussion

3.1 Perpetual Inventory Method

The non-identification of $\beta_{k,0}$, if the index restriction is not satisfied, is a consequence of (3) which in turn is implied by the measurement of the capital stock by the PIM as in (OP 1). We argue that it is possible to identify $\beta_{k,0}$ without the index restriction if the capital stock satisfies

$$k_{t+1} = (1-\delta)k_t + i_t + u_t \tag{9}$$

with u_t a shock to the value of the capital stock, e.g., because of technological progress that makes part of the capital stock obsolete. For the purpose of identification, we further assume that (i) $u_t \in J_t$, but u_t is not correlated over time so that it is not a state variable in (OP 3), and (ii) $\mathbb{E}[\omega_{t+1}|\,\omega_t, k_t, u_t] = \mathbb{E}[\omega_{t+1}|\,\omega_t, k_t]$.

Under these assumptions and with the updated J_t, (4) becomes

$$\begin{aligned}
\mathbb{E}[y_{t+1} - \beta_{l,0}l_{t+1}|\,\omega_t, k_t, u_t] &= \beta_0 + \beta_{k,0}k_{t+1} + \mathbb{E}[\omega_{t+1}|\,\omega_t, k_t, u_t] \\
&= \beta_0 + \beta_{k,0}k_{t+1} + \mathbb{E}[\omega_{t+1}|\,\omega_t, k_t] \\
&= \beta_{k,0}k_{t+1} + g(\omega_t, k_t),
\end{aligned}$$

since $\mathbb{E}[k_{t+1}|\omega_t, k_t, u_t] = k_{t+1}$. Because $k_{t+1} = (1-\delta)k_t + i_t + u_t \neq \mathbb{E}[k_{t+1}|\,i_t, k_t] = \mathbb{E}[k_{t+1}|\,\omega_t, k_t]$, we can estimate $\beta_{k,0}$ by regressing $y_{t+1} - \beta_{l,0}l_{t+1} - \mathbb{E}[y_{t+1} - \beta_{l,0}l_{t+1}|\,i_t, k_t]$ on $k_{t+1} - \mathbb{E}[k_{t+1}|\,i_t, k_t]$. Therefore if the capital stock is measured by a method that does not involve an exact relation between k_{t+1} and i_t, k_t, then we can relax the index restriction, or even test the restriction by comparing estimates of $\beta_{k,0}$ with and without the index restriction.

If the capital stock data are constructed using the PIM then $u_t \equiv 0$ and $\beta_{k,0}$ is not identified. The accounting identity $k_{t+1} = k_t + i_t - d_t$ with d_t the depreciation in period t implies that in (9) $d_t = \delta k_t - u_t$. Therefore the depreciation depends on other variables than the current capital stock. For instance a machine is scrapped because a technologically more advanced one has become available. To identify $\beta_{k,0}$ without the index restriction, plant level data on k_t or d_t are required, as available in the Compustat® database. With the subsample from the Compustat data used by İmrohoroğlu and Tüzel (2014) it is easily checked that the depreciation rate d_t/k_t differs between firms and over time.[13]

3.2 Sample Selection

The preceding analysis of the PIM raises concerns regarding the identification of $\beta_{k,0}$, if firms can close down plants. In fact, it can be shown that $\beta_{k,0}$ is not identified with sample selectivity if (OP 1) is satisfied, which contradicts OP's claim.

In their equation (4), OP specify a threshold model for plant survival (see OP, p.1273): $\chi_t = 1$ iff $\omega_t \geq \underline{\omega}_t(k_t)$ with $\underline{\omega}_t(k_t)$ the value that makes the firm indifferent between scrapping and continuing the plant. Therefore their equation (11) that accounts for the scrapping of plants is (if we additionally condition on the information at t as stated by OP)

$$\mathbb{E}\left[y_{t+1} - \beta_{l,0} l_{t+1} \middle| k_{t+1}, \omega_t, k_t, , \chi_{t+1} = 1\right] = \beta_{k,0} k_{t+1} + g\left(\underline{\omega}_{t+1}(k_{t+1}), \omega_t\right). \qquad \text{(OP 11')}$$

Note that we impose the index restriction. OP (p.1276) define $g\left(\underline{\omega}_{t+1}(k_{t+1}), \omega_t\right)$ by

$$g\left(\underline{\omega}_{t+1}(k_{t+1}), \omega_t\right) = \beta_0 + \int_{\underline{\omega}_{t+1}(k_{t+1})} \omega_{t+1} \frac{F\left(d\omega_{t+1} \middle| \omega_t\right)}{\int_{\underline{\omega}_{t+1}(k_{t+1})} F\left(d\omega_{t+1} \middle| \omega_t\right)} = \mathbb{E}\left[\omega_{t+1} \middle| \omega_{t+1} \geq \underline{\omega}_{t+1}(k_{t+1}), \omega_t\right].$$

The problem is that $\underline{\omega}_{t+1}(k_{t+1})$ is a function of k_{t+1}, which raises the question of under-identification.

Note that $g\left(\underline{\omega}_{t+1}(k_{t+1}), \omega_t\right)$ is a strictly increasing function of $\underline{\omega}_{t+1}(k_{t+1})$ for any given ω_t. Also $\underline{\omega}_{t+1}(k_{t+1})$ is a strictly decreasing function of k_{t+1}, so that $\bar{g}(\omega_t, k_{t+1}) \equiv g\left(\underline{\omega}_{t+1}(k_{t+1}), \omega_t\right)$ is strictly decreasing in k_{t+1}.

As in the previous section, there are observationally equivalent parameters. By (OP 11') and the PIM

$$\begin{aligned} \mathbb{E}\left[y_{t+1} - \beta_l l_{t+1} \middle| \omega_t, k_t, \chi_t = 1\right] &= \bar{\beta}_k k_{t+1} + \bar{g}(\omega_t, k_{t+1}) + \left(\beta_k - \bar{\beta}_k\right) k_{t+1} \\ &= \bar{\beta}_k k_{t+1} + \breve{g}(\omega_t, k_{t+1}) \end{aligned}$$

for $\breve{g}(\omega_t, k_{t+1}) \equiv \bar{g}(\omega_t, k_{t+1}) + \left(\beta_k - \bar{\beta}_k\right) k_{t+1}$, which is strictly decreasing in k_{t+1} if $\bar{\beta}_k > \beta_k$. Because both $\bar{g}$ and $\breve{g}$ are nonparametric, we conclude that β_k is not identified.

[13] Among others, Brynjolfsson and Hitt (2003) and İmrohoroğlu and Tüzel (2014) have used the Compustat capital stock and depreciation data with the adjustments suggested by Hall (1990). İmrohoroğlu and Tüzel (2014) use the OP estimator. Piketty and Zucman (2014) criticize the use of the PIM for the measurement of the capital stock. Hulten (1990) discusses practical aspects of the measurement of capital. We thank Monica Morlacco for discussions on this topic.

This issue can be seen slightly differently. First we note that by the third and fifth lines of their equation (10), we have for the survival probability

$$P_t \equiv \mathfrak{p}_t \left(\underline{\omega}_{t+1}(k_{t+1}), \omega_t\right) = \mathfrak{p}_t(i_t, k_t).$$

If we read for $\mathfrak{p}_t$ the conditional survival function of ω_{t+1} given ω_t, then we can invert the relation to obtain $\underline{\omega}_{t+1}(k_{t+1})$ as a function of P_t and ω_t. OP (p. 1276) therefore obtain for the truncated conditional mean

$$g\left(\underline{\omega}_{t+1}(k_{t+1}), \omega_t\right) = g\left(P_t, \omega_t\right).$$

As in our discussion of the index restriction above, we rewrite (OP 11) as conditional on the state variables and survival

$$\mathbb{E}\left[y_{t+1} - \beta_{l,0} l_{t+1} | \, \omega_t, k_t, \chi_t = 1\right] = \beta_{k,0} k_{t+1} + g\left(P_t, \omega_t\right),$$

where the equality follows from the PIM.

Because P_t is strictly increasing in k_{t+1} given ω_t, we can invert the relationship and write k_{t+1} as a function of (P_t, ω_t). Therefore, the partially linear regression of $y_{t+1} - \beta_l l_{t+1}$ on k_{t+1} (using (P_t, ω_t) as an argument of the nonparametric component) fails to identify β_k.

The problem disappears if k_{t+1} cannot be written a a function of (P_t, ω_t). For example, $\underline{\omega}_{t+1}(k_{t+1}) = \max\left(k_{t+1}, C\right)$ may eliminate the under-identification problem. However, it is not clear if that choice is consistent with the optimal scrapping rule in (OP 3). Also note that the PIM was used to find an observationally equivalent model. Whether the model with attrition is identified if the capital stock is not measured using the PIM is beyond the scope of this paper.

4 The Influence Function of the Estimator

In this section, we discuss how the asymptotic distribution of the OP estimator can be characterized using recent results on inference in semi-parametric models with generated regressors. We argue that the index restriction not only plays a crucial role in the identification, but it also makes it possible to characterize the influence function.

As discussed in the previous section, the OP estimator is based on a two-step identification strategy. Our derivation of the asymptotic distribution is based on an alternative characterization of the minimization in the second step. It is convenient to start with the case that $\beta_{l,0}$ and $\phi_t(\cdot)$ are known. We characterize the second step as consisting of two sub-steps:

1. For given β_k, we minimize the objective function

$$\mathbb{E}\left[\left(y_{t+1} - \beta_{l,0} l_{t+1} - \beta_k k_{t+1} - g\left(\phi_t - \beta_k k_t\right)\right)^2\right] \tag{10}$$

with respect to g, where ~~we let~~ $\phi_t \equiv \phi_t(i_t, k_t)$. The solution that depends on β_k is equal to (note that we omit the conditioning variables in ϕ_t)

$$\mathbb{E}\left[y_{t+1} - \beta_{l,0} l_{t+1} \middle| \phi_t - \beta_k k_t \right] - \beta_k \mathbb{E}\left[k_{t+1} \middle| \phi_t - \beta_k k_t \right] \tag{11}$$

2. Upon substitution of (11) in the objective function (10), we obtain a concentrated objective function, that we minimize with respect to β_k.

To keep the notation simple, we write $Y_1 \equiv y_{t+1} - \beta_{l,0} l_{t+1}$, $Y_2 \equiv k_{t+1}$, and $\gamma_j \left(\phi_t - \beta_k k_t \right) \equiv \mathbb{E}\left[Y_j \middle| \phi_t - \beta_k k_t \right]$ for $j = 1, 2$. With this notation, we can write

$$g\left(\phi_t - \beta_k k_t \right) = \gamma_1 \left(\phi_t - \beta_k k_t \right) - \beta_k \gamma_2 \left(\phi_t - \beta_k k_t \right). \tag{12}$$

The minimization problem in the second sub-step is

$$\min_{\beta_k} \mathbb{E}\left[\frac{1}{2} \left(Y_1 - \gamma_1 \left(\phi_t - \beta_k k_t \right) - \beta_k \left(Y_2 - \gamma_2 \left(\phi_t - \beta_k k_t \right) \right) \right)^2 \right]. \tag{13}$$

For the first-order condition we need the derivative of the concentrated objective function with respect to β_k. There are two complications. First, we note that even if $\beta_{l,0}$ and ϕ_t are known, the conditional expectations $\mathbb{E}\left[Y_j \middle| \phi_t - \beta_k k_t \right]$ depend on β_k. This means that the derivative of the function under the expectation

$$M\left(Y_1, Y_2, \phi_t, k_t; \beta_k \right) \equiv \frac{1}{2} \left(Y_1 - \gamma_1 \left(\phi_t - \beta_k k_t \right) - \beta_k \left(Y_2 - \gamma_2 \left(\phi_t - \beta_k k_t \right) \right) \right)^2$$

has to take account of this dependence. Second, the ϕ_t is in fact estimated, so that its sampling variation affects the conditioning variable in γ_1 and γ_2.

A nice feature is that the estimation of γ_1 and γ_2 has no contribution to the influence function, i.e. we can consider their estimates as the population parameters. This follows from Newey (1994, p. 1357-58).[14] Newey shows that if an infinitely dimensional parameter as g, and therefore (γ_1, γ_2), is the solution to a minimization problem as (10), then its estimation does not have a contribution to the influence function of the estimator of β_k. By the same argument there is no contribution to the influence function of $\hat{\beta}_l$ from the estimation of ϕ_t that is the solution to the minimization problem (2).

Even with this simplification, the derivative of $M\left(Y_1, Y_2, \phi_t, k_t; \beta_k \right)$ with respect to β_k is not a trivial object. Consider $\mathbb{E}\left[Y_j \middle| \nu_t \left(\beta_k \right) \right]$ with

$$\nu_t \left(\beta_k \right) \equiv \phi_t - \beta_k k_t.$$

[14]We are referring to the argument that leads to Proposition 2 on p.1358. Unfortunately, there is a typo there; the proposition actually refers to equation (3.11) instead of equation (3.10).

The conditional expectation $\mathbb{E}\left[Y_j | \nu_t(\beta_k)\right]$ depends on β_k in two ways. First, the dependence is directly through its argument $\nu_t(\beta_k)$. Second, β_k affects the entire shape of the conditional distribution of Y_j given $\nu_t(\beta_k)$ and therefore its conditional mean. So a better notation for the conditional mean is

$$\mathbb{E}\left[Y_j | \nu_t(\beta_k)\right] = \gamma_j\left(\nu_t(\beta_k); \beta_k\right)$$

This notation emphasizes the two roles of β_k in this conditional expectation. The total derivative is the sum of the partial derivatives with respect to both appearances of β_k. Of these the derivative $\partial \gamma_j(\nu; \beta_k) / \partial \beta_k$ is not obvious. Hahn and Ridder (2013) characterized the *expectation of* such derivatives, but not the derivatives themselves.

To find the derivative we use a result in Newey (1994, p.1358), who shows how to calculate the derivative if the parameter enters in an index as is the case in OP. In the previous section, it was argued that an index restriction is crucial for identification. We now exploit the index restriction for the first order condition for (13). As noted above, Hahn and Ridder (2013) do not derive the derivatives of the conditional expectations. On the other hand, Newey (1994, p. 1358, l. 19)[15] derives an expression for such a derivative under the index restriction. Although the index restriction does not necessarily hold for γ_1 and γ_2, it does hold for g by the discussion in the previous section. This means that we can apply Newey's result to characterize the derivative of g, and obtain the first order condition below in Proposition 1.[16] This is the third important implication of the index restriction.

Proofs of the following results are collected in the appendix.

Proposition 1 *The first-order condition for the minimization problem (13) is given by*

$$0 = \mathbb{E}\left[\left(\xi_{t+1} + \eta_{t+1}\right)\left(\left(k_{t+1} - \mathbb{E}\left(k_{t+1} | \nu_t\right)\right) - \frac{\partial g(\nu_t)}{\partial \nu_t}\left(k_t - \mathbb{E}\left[k_t | \nu_t\right]\right)\right)\right], \tag{14}$$

where $\nu_t \equiv \nu_t(\beta_{k,0})$, η *and* ξ *are defined in (OP 6) and (1) respectively.*

Newey's (1994) result that is based on an index restriction, can be utilized to verify the (local) identification of β_k. Proposition 2 gives the second derivative Υ of (13) with respect to β_k. The second derivative $\Upsilon > 0$ in general, so that β_k is locally identified.

Proposition 2 *The second order derivative of (13) with respect to* β_k *is*

$$\Upsilon = \mathbb{E}\left[\left(\left(k_{t+1} - \mathbb{E}\left(k_{t+1} | \nu_t\right)\right) - \frac{\partial g(\nu_t)}{\partial \nu_t}\left(k_t - \mathbb{E}\left[k_t | \nu_t\right]\right)\right)^2\right]. \tag{15}$$

[15]See the working paper version of Hahn and Ridder (2013) for more detailed analysis.

[16]Specifically, the term $-\frac{\partial g(\nu_t)}{\partial \nu_t}\left(k_t - \mathbb{E}\left[k_t | \nu_t\right]\right)$ on the RHS of (14) is the result of applying Newey's argument.

The next proposition gives the influence function of $\hat{\beta}_k$.

Proposition 3 *The influence function of $\hat{\beta}_k$ is the sum of the main term that is the normalized sum of the moment functions in (24) in the Appendix:*

$$\Upsilon^{-1}\left(\frac{1}{\sqrt{n}}\sum_{i=1}^{n}\begin{pmatrix} y_{i,t+1}-\beta_{l,0}l_{i,t+1}-\beta_{k,0}k_{i,t+1} \\ -\mathbb{E}\left[\left.y_{i,t+1}-\beta_{l,0}l_{i,t+1}-\beta_{k,0}k_{i,t+1}\right|\nu_{i,t}\right]\end{pmatrix}\cdot\begin{pmatrix} k_{i,t+1}-\mathbb{E}\left[\left.k_{i,t+1}\right|\nu_{i,t}\right] \\ -\frac{\partial g(\nu_{i,t})}{\partial\nu_{i,t}}\left(k_{i,t}-\mathbb{E}\left[\left.k_{i,t}\right|\nu_{i,t}\right]\right)\end{pmatrix}\right) \tag{16}$$

and the adjustment

$$-\Upsilon^{-1}\frac{1}{\sqrt{n}}\sum_{i=1}^{n}\left(\Lambda_{1i}+\Lambda_{2i}\right)\left(\left(y_{i,t}-\beta_{l,0}l_{i,t}\right)-\mathbb{E}\left[\left.y_{i,t}-\beta_{l,0}l_{i,t}\right|i_{i,t},k_{i,t}\right]\right), \tag{17}$$

where

$$\Lambda_{1i}\equiv\frac{\partial g\left(\nu_{i,t}\right)}{\partial\nu_{i,t}}\left(\left(k_{i,t+1}-\mathbb{E}\left[\left.k_{i,t+1}\right|\nu_{i,t}\right]\right)-\frac{\partial g\left(\nu_{i,t}\right)}{\partial\nu_{i,t}}\left(k_{i,t}-\mathbb{E}\left[\left.k_{i,t}\right|\nu_{i,t}\right]\right)\right),$$

$$\Lambda_{2i}\equiv\frac{\Gamma}{\mathbb{E}\left[\left(l_{i,t}-\mathbb{E}\left[\left.l_{i,t}\right|i_{i,t},k_{i,t}\right]\right)^2\right]}\left(l_{i,t}-\mathbb{E}\left[\left.l_{i,t}\right|i_{i,t},k_{i,t}\right]\right),$$

and

$$\Gamma\equiv\mathbb{E}\left[\begin{array}{c}\left(\left(l_{t+1}-\mathbb{E}\left[\left.l_{t+1}\right|\nu_t\right]\right)-\frac{\partial g(\nu_{i,t})}{\partial\nu_{i,t}}\mathbb{E}\left[\left.l_t\right|i_t,k_t\right]\right) \\ \times\left(\left(k_{t+1}-\mathbb{E}\left[\left.k_{t+1}\right|\nu_t\right]\right)-\frac{\partial g(\nu_{i,t})}{\partial\nu_{i,t}}\left(k_t-\mathbb{E}\left[\left.k_t\right|\nu_t\right]\right)\right)\end{array}\right].$$

Remark 1 *Our decomposition of the influence function is helpful in case $\hat{\beta}_l$ is not estimated by the partially linear regression method, which is useful because Ackerberg, Caves, and Frazer (2015) raised concerns about identification of $\beta_{l,0}$ by such strategy. If so, it may be desired to estimate it by some other method. If the alternative method is such that the influence function is $\varepsilon_{1,i}$, a straightforward modification of our proof indicates that (17) should be replaced by*

$$-\Upsilon^{-1}\frac{1}{\sqrt{n}}\sum_{i=1}^{n}\Lambda_{1i}\left(\left(y_{i,t}-\beta_{l,0}l_{i,t}\right)-\mathbb{E}\left[\left.y_{i,t}-\beta_{l,0}l_{i,t}\right|i_{i,t},k_{i,t}\right]\right)-\Upsilon^{-1}\Gamma\frac{1}{\sqrt{n}}\sum_{i=1}^{n}\varepsilon_{1,i}.$$

Remark 2 *We derived the influence function directly by Newey's (1994) results. A derivation of the asymptotic distribution by stochastic expansion is included in the Supplemental Appendix of this paper, which is available from the authors upon request.*

5 Summary

In this paper, we examined the identifying assumptions Olley and Pakes (1996). We argued that an index restriction plays a crucial role in identification, especially if the capital stock is measured by the perpetual inventory method. We argued that the index restriction is not sufficient for identification under sample selectivity. Finally, we exploited the index restriction to derive the influence function of the OP estimator.

References

[1] Ackerberg, D.A.,L. Benkard, S. Berry, and A. Pakes, (2007): "Econometric Tools for Analyzing Market Outcomes," in Heckman, J. and Leamer, E., eds., Handbook of Econometrics, Volume 6A, Amsterdam: North Holland, 4171-4276.

[2] Ackerberg, D.A., K. Caves, and G. Frazer (2015): "Identification Properties of Recent Production Function Estimators," Econometrica, 83, 2411–2451.

[3] Ackerberg, D.A., Chen, X., and Hahn, J. (2012): "A Practical Asymptotic Variance Estimator for Two-Step Semiparametric Estimators," Review of Economics and Statistics, 94, 481-498.

[4] Brynjolfsson, E., L.M. Hitt (2003): "Computing Productivity: Firm-Level Evidence," Review of Economics and Statistics, 85, 793–808.

[5] Cattaneo, M., M. Jansson, and X. Ma (2019): "Two-Step Estimation and Inference with Possibly Many Included Covariates," Review of Economic Studies, 86, 1095–1122.

[6] Hahn, J., and Ridder, G. (2013): "The Asymptotic Variance of Semi-parametric Estimators with Generated Regressors," Econometrica, 81, 315–340.

[7] Hahn, J., Liao, Z., and Ridder, G. (2018): "Nonparametric Two-Step Sieve M Estimation and Inference," Econometric Theory, 34, 1281-1324.

[8] Hall, B. H. (1990): "The Manufacturing Sector Master File: 1959-1987," NBER, Working paper 3366.

[9] Hulten, C. R. (1990): "The Measurement of Capital," in E. R. Berndt and Jack E. T., eds., Fifty Years of Economic Measurement, Studies in Income and Wealth Volume 54, Chicago and London: The University of Chicago Press, 119-152.

[10] İmrohoroğlu, A., S. Tüzel (2014): "Firm Level Productivity, Risk, and Return," Management Science, 60, 2073-2090.

[11] Mammen, E., C. Rothe, and M. Schienle (2014): "Semiparametric Estimation with Generated Covariates," working paper SFB 649, 2014-043.

[12] Mammen, E., C. Rothe, and M. Schienle (2016): "Semiparametric Estimation with Generated Covariates," Econometric Theory, 32, 1140-1177.

[13] Newey, W. K. (1994): "The Asymptotic Variance of Semiparametric Estimators," Econometrica, 62(6), 1349-1382.

[14] Olley, S., and Pakes, A. (1996): "The Dynamics of Productivity in the Telecommunications Equipment Industry," Econometrica, 64, 1263-1297.

[15] Pakes, A. , and Olley, S.(1995): "A Limit Theorem for a Smooth Class of Semiparametric Estimators," Journal of Econometrics, 65, 295-332.

[16] Piketty, T., and G. Zucman (2014): "Capital is Back: Wealth-income Ratios in Rich Countries 1700–2010," Quarterly Journal of Economics, 129, 1255–1310.

Appendix

A Proof of Proposition 1

Recall that $\nu_t(\beta_k) \equiv \phi_t - \beta_k k_t$ and $\nu_t \equiv \nu_t(\beta_{k,0}) = \phi_t - \beta_{k,0} k_t$. If we interchange expectation and differentiation, then the first-order condition for (13) involves the derivative

$$\begin{aligned} &\left.\frac{\partial\left(Y_1 - \gamma_1\left(\phi_t - \beta_k k_t\right) - \beta_k\left(Y_2 - \gamma_2\left(\phi_t - \beta_k k_t\right)\right)\right)}{\partial \beta_k}\right|_{\beta_k=\beta_{k,0}} \\ &= -\left(Y_2 - \gamma_2\left(\phi_t - \beta_{k,0} k_t\right)\right) - \left.\frac{\partial\left(\gamma_1\left(\phi_t - \beta_k k_t\right) - \beta_{k,0}\gamma_2\left(\phi_t - \beta_k k_t\right)\right)}{\partial \beta_k}\right|_{\beta_k=\beta_{k,0}}. \end{aligned} \tag{18}$$

In the second line of (18), the derivative is with respect to β_k that appears in the common argument $\phi_t - \beta_k k_t$ of γ_1, γ_2. If we take the conditional expectation of $Y_1 - \beta_{k,0} Y_2 = \beta_0 + \omega_{t+1} + \eta_{t+1}$ with respect to $\nu_t(\beta_k)$, we obtain

$$\tau(\nu_t(\beta_k)) \equiv \mathbb{E}\left[\beta_0 + \omega_{t+1} + \eta_{t+1} \middle| \nu_t(\beta_k)\right] = \gamma_1\left(\nu_t(\beta_k)\right) - \beta_{k,0}\gamma_2\left(\nu_t(\beta_k)\right).$$

If evaluated at $\beta_k = \beta_{k,0}$, the above equations yield

$$\tau(\nu_t) = \mathbb{E}\left[\beta_0 + \omega_{t+1} + \eta_{t+1} \middle| \nu_t\right] = \mathbb{E}\left[\beta_0 + \omega_{t+1} \middle| \omega_t - \beta_0\right] = g(\nu_t)$$

because $\nu_t = \beta_0 + \omega_t$ and the index restriction holds. We now apply the derivative formula in Newey, (1994, Example 1 Continued, p.1358)

$$\begin{aligned} &\left.\frac{\partial\left(\gamma_1\left(\phi_t - \beta_k k_t\right) - \beta_{k,0}\gamma_2\left(\phi_t - \beta_k k_t\right)\right)}{\partial \beta_k}\right|_{\beta_k=\beta_{k,0}} \\ &= \left.\frac{\partial \tau(\nu_t)}{\partial \nu_t}\left(\frac{\partial \nu_t(\beta_k)}{\partial \beta_k} - \mathbb{E}\left[\left.\frac{\partial \nu_t(\beta_k)}{\partial \beta_k}\right| \nu_t\right]\right)\right|_{\beta_k=\beta_{k,0}} \\ &= -\frac{\partial g(\nu_t)}{\partial \nu_t}\left(k_t - \mathbb{E}\left[k_t \middle| \nu_t\right]\right). \end{aligned} \tag{19}$$

Combining (18) and (19), we obtain

$$\begin{aligned} &-\left.\frac{\partial\left(Y_1 - \gamma_1\left(\phi_t - \beta_k k_t\right) - \beta_k\left(Y_2 - \gamma_2\left(\phi_t - \beta_k k_t\right)\right)\right)}{\partial \beta_k}\right|_{\beta_k=\beta_{k,0}} \\ &= \left(Y_2 - \gamma_2\left(\nu_t\right)\right) - \frac{\partial g(\nu_t)}{\partial \nu_t}\left(k_t - \mathbb{E}\left[k_t \middle| \nu_t\right]\right), \end{aligned}$$

and hence, we may write

$$
\begin{aligned}
& - \left.\frac{\partial M\left(Y_1, Y_2, \phi_t, k_t; \beta_k\right)}{\partial \beta_k}\right|_{\beta_k=\beta_{k,0}} \\
& = \left(Y_1 - \gamma_1\left(\nu_t\right) - \beta_{k,0}\left(Y_2 - \gamma_2\left(\nu_t\right)\right)\right)\left(Y_2 - \gamma_2\left(\nu_t\right) - \frac{\partial g(\nu_t)}{\partial \nu_t}\left(k_t - \mathbb{E}\left[k_t \mid \nu_t\right]\right)\right) \\
& = \left(\xi_{t+1} + \eta_{t+1}\right)\left(k_{t+1} - \mathbb{E}[k_{t+1}|\nu_t] - \frac{\partial g(\nu_t)}{\partial \nu_t}\left(k_t - \mathbb{E}\left[k_t \mid \nu_t\right]\right)\right), \qquad (20)
\end{aligned}
$$

with the corresponding first-order condition for (13):

$$
0 = \mathbb{E}\left[\left(\xi_{t+1} + \eta_{t+1}\right)\left(\left(k_{t+1} - \mathbb{E}[k_{t+1}|\nu_t(\beta_k)]\right) - \frac{\partial g(\nu_t(\beta_k))}{\partial \nu_t(\beta_k)}\left(k_t - \mathbb{E}\left[k_t \mid \nu_t(\beta_k)\right]\right)\right)\right].
$$

that holds if $\beta_k = \beta_{k,0}$.

B Proof of Proposition 2

We calculate the second order derivative for (13) as well. The first derivative is by the second line of (20) the product of two factors. Minus the derivative of the first factor $Y_1 - \gamma_1\left(\nu_t\right) - \beta_k\left(Y_2 - \gamma_2\left(\nu_t\right)\right)$ is equal to the second factor, so that the second derivative is

$$
\begin{aligned}
& \mathbb{E}\left[\left(\left(Y_2 - \gamma_2\left(\nu_t\right)\right) - \frac{\partial g(\nu_t)}{\partial \nu_t}\left(k_t - \mathbb{E}\left[k_t \mid \nu_t\right]\right)\right)^2\right] \\
& - \mathbb{E}\left[\left(\xi_{t+1} + \eta_{t+1}\right)\left.\frac{\partial\left(\left(Y_2 - \mathbb{E}[Y_2|\nu_t(\beta_k)]\right) - \frac{\partial g(\nu_t(\beta_k))}{\partial \nu_t}\left(k_t - \mathbb{E}\left[k_t \mid \nu_t(\beta_k)\right]\right)\right)}{\partial \beta_k}\right|_{\beta_k=\beta_{k,0}}\right] \qquad (21)
\end{aligned}
$$

Note that the component in the second term on the right

$$
\begin{aligned}
& \left.\frac{\partial\left(\left(Y_2 - \mathbb{E}[Y_2|\nu_t(\beta_k)]\right) - \frac{\partial g(\nu_t(\beta_k))}{\partial \nu_t}\left(k_t - \mathbb{E}\left[k_t \mid \nu_t(\beta_k)\right]\right)\right)}{\partial \beta_k}\right|_{\beta_k=\beta_{k,0}} \\
& = -\frac{\partial \gamma_2\left(\nu_t\right)}{\partial \beta_k} - \frac{\partial\left(\frac{\partial g(\nu_t)}{\partial \nu_t}\right)}{\partial \beta_k}\left(k_t - \mathbb{E}\left[k_t \mid \nu_t\right]\right) + \frac{\partial g\left(\nu_t\right)}{\partial \nu_t}\frac{\partial \mathbb{E}\left[k_t \mid \nu_t\right]}{\partial \beta_k}
\end{aligned}
$$

is a function in (ϕ_t, k_t), which in turn is a function of (i_t, k_t). Because $\mathbb{E}\left[\xi_{t+1} + \eta_{t+1} \mid i_t, k_t\right] = 0$, we get the desired conclusion.

C Proof of Proposition 3

C.1 Main Term

In Section 4, we argued that the estimation of γ_1 and γ_2 does not require an adjustment of the influence function. This follows from Newey's (1994) Proposition 2 that states that if the parameter (here β_k) and the non-parametric function (here γ_1 and γ_2) are estimated by minimizing an objective function, then the estimation errors of γ_1 and γ_2 can be ignored in the influence function. In other words we can consider γ_1 and γ_2 as known in our analysis of the first order condition. The main term follows directly from the moment function in (24) multiplied by Υ^{-1}. We now turn to the adjustment.

C.2 Impact of the First Step Estimation on the Distribution of $\hat{\beta}_k$

We recall that

$$\phi_t(i_t, k_t) = \mathbb{E}\left[y_t | i_t, k_t\right] - \beta_l \mathbb{E}\left[l_t | i_t, k_t\right]$$

We apply Newey (1994, Proposition 4) to find the adjustment for the estimation of $\mathbb{E}[y_t|i_t, k_t]$, $\mathbb{E}[l_t|i_t, k_t]$ and β_l. We conclude the following:[17]

1. The adjustment for estimating $(\mathbb{E}\left[y_t | i_t, k_t\right], \mathbb{E}\left[l_t | i_t, k_t\right])$ can be calculated separately:

   (a) The adjustment for estimating $\mathbb{E}\left[y_t | i_t, k_t\right]$ is

   $$\Upsilon^{-1}\left(\frac{1}{\sqrt{n}}\sum_{i=1}^{n}\delta_1(i_{i,t}, k_{i,t})\left(y_{i,t} - \mathbb{E}\left[y_{i,t} | i_{i,t}, k_{i,t}\right]\right)\right),$$

   where

   $$\delta_1(i_t, k_t) \equiv -\frac{\partial g(\nu_t)}{\partial \nu_t}\left(\left(k_{t+1} - \mathbb{E}\left[k_{t+1} | \nu_t\right]\right) - \frac{\partial g(\nu_t)}{\partial \nu_t}\left(k_t - \mathbb{E}\left[k_t | \nu_t\right]\right)\right).$$

   (b) The adjustment for estimating $\mathbb{E}\left[l_t | i_t, k_t\right]$ is

   $$\Upsilon^{-1}\left(\frac{1}{\sqrt{n}}\sum_{i=1}^{n}\delta_2(i_{i,t}, k_{i,t})\left(l_{i,t} - \mathbb{E}\left[l_{i,t} | i_{i,t}, k_{i,t}\right]\right)\right),$$

   where

   $$\delta_2(i_t, k_t) \equiv -\beta_{l,0}\delta_1(i_t, k_t).$$

[17] All the details of derivations can be found in Section C.3.

(c) The adjustment for $\left(\widehat{\mathbb{E}}\left[\left.y_t\right|i_t,k_t\right],\widehat{\mathbb{E}}\left[\left.l_t\right|i_t,k_t\right]\right)$ is therefore the sum of the above two expressions:

$$-\Upsilon^{-1}\left(\frac{1}{\sqrt{n}}\sum_{i=1}^{n}\frac{\partial g\left(\nu_t\right)}{\partial\nu_t}\left(\begin{array}{c}\left(k_{i,t+1}-\mathbb{E}\left[\left.k_{i,t+1}\right|\nu_{i,t}\right]\right)\\-\frac{\partial g(\nu_t)}{\partial\nu_t}\left(k_{i,t}-\mathbb{E}\left[\left.k_{i,t}\right|\nu_{i,t}\right]\right)\end{array}\right)\left(\begin{array}{c}\left(y_{i,t}-\beta_{l,0}l_{i,t}\right)\\-\mathbb{E}\left[\left.y_{i,t}-\beta_{l,0}l_{i,t}\right|i_{i,t},k_{i,t}\right]\end{array}\right)\right),\tag{22}$$

which is equal to

$$-\Upsilon^{-1}\frac{1}{\sqrt{n}}\sum_{i=1}^{n}\Lambda_{1i}\left(\left(y_{i,t}-\beta_{l,0}l_{i,t}\right)-\mathbb{E}\left[\left.y_{i,t}-\beta_{l,0}l_{i,t}\right|i_{i,t},k_{i,t}\right]\right)$$

with

$$\Lambda_{1i}\equiv\frac{\partial g\left(\nu_t\right)}{\partial\nu_t}\left(\left(k_{i,t+1}-\mathbb{E}\left[\left.k_{i,t+1}\right|\nu_{i,t}\right]\right)-\frac{\partial g\left(\nu_t\right)}{\partial\nu_t}\left(k_{i,t}-\mathbb{E}\left[\left.k_{i,t}\right|\nu_t\right]\right)\right).$$

2. The impact of estimating $\widehat{\beta}_l$ is more convoluted, because we have to remember that $\beta_{l,0}$ appears in the moment function through $Y_1=y_{t+1}-\beta_{l,0}l_{t+1}$ and through $\nu_t=\mathbb{E}\left[\left.y_t\right|i_t,k_t\right]-\beta_{l,0}\mathbb{E}\left[\left.l_t\right|i_t,k_t\right]-\beta_{k,0}k_t$.

(a) We start with the contribution through ν_t. Because

$$\frac{\partial\nu_t}{\partial\beta_{l,0}}=-\mathbb{E}\left[\left.l_t\right|i_t,k_t\right]$$

we can see that $\widehat{\beta}_l$ impacts the influence function by

$$-\Upsilon^{-1}\mathbb{E}\left[\delta_1\left(i_t,k_t\right)\mathbb{E}\left[\left.l_t\right|i_t,k_t\right]\right]\sqrt{n}\left(\widehat{\beta}_l-\beta_{l,0}\right)$$

(b) The other impact is because $\beta_{l,0}$ appears in moment function through $Y_1=y_{t+1}-\beta_{l,0}l_{t+1}$. We take the derivative of the population moment function in (25) with respect to $\beta_{l,0}$:

$$-\Upsilon^{-1}\mathbb{E}\left[\left(l_{t+1}-\mathbb{E}\left[\left.l_{t+1}\right|\nu_t\right]\right)\left(\begin{array}{c}\left(k_{t+1}-\mathbb{E}\left[\left.k_{t+1}\right|\nu_t\right]\right)\\-\frac{\partial g(\nu_t)}{\partial\nu_t}\left(k_t-\mathbb{E}\left[\left.k_t\right|\nu_t\right]\right)\end{array}\right)\right]\sqrt{n}(\widehat{\beta}_l-\beta_{l,0})$$

This is a traditional two-step adjustment of the influence function.

(c) The adjustment for $\widehat{\beta}_l$ is therefore equal to the sum of the above two expressions:

$$-\Upsilon^{-1}\Gamma\sqrt{n}(\widehat{\beta}_l-\beta_{l,0}),$$

where we recall

$$\Gamma\equiv\mathbb{E}\left[\left(\begin{array}{c}\left(l_{t+1}-\mathbb{E}\left[\left.l_{t+1}\right|\nu_t\right]\right)\\-\frac{\partial g(\nu_t)}{\partial\nu_t}\mathbb{E}\left[\left.l_t\right|i_t,k_t\right]\end{array}\right)\left(\begin{array}{c}\left(k_{t+1}-\mathbb{E}\left[\left.k_{t+1}\right|\nu_t\right]\right)\\-\frac{\partial g(\nu_t)}{\partial\nu_t}\left(k_t-\mathbb{E}\left[\left.k_t\right|\nu_t\right]\right)\end{array}\right)\right]$$

This is the adjustment in Remark 1 after Proposition 3.

(d) In OP, $\widehat{\beta}_l$ is obtained by partially linear regression estimation. Using the standard results, we get

$$\begin{aligned}
&\sqrt{n}(\widehat{\beta}_l - \beta_{l,0}) \\
&= \frac{\frac{1}{\sqrt{n}} \sum_{i=1}^{n} \left(l_{i,t} - \mathbb{E}\left[l_{i,t} \mid i_{i,t}, k_{i,t}\right]\right) \left(\left(y_{i,t} - \beta_{l,0} l_{i,t}\right) - \mathbb{E}\left[y_{i,t} - \beta_{l,0} l_{i,t} \mid i_{i,t}, k_{i,t}\right]\right)}{\mathbb{E}\left[\left(l_{i,t} - \mathbb{E}\left[l_{i,t} \mid i_{i,t}, k_{i,t}\right]\right)^2\right]} + o_p(1),
\end{aligned}$$

we conclude that the adjustment for $\widehat{\beta}_l$ is

$$-\Upsilon^{-1} \Gamma \frac{1}{\sqrt{n}} \sum_{i=1}^{n} \frac{\left(l_{i,t} - \mathbb{E}\left[l_{i,t} \mid i_{i,t}, k_{i,t}\right]\right) \left(\left(y_{i,t} - \beta_{l,0} l_{i,t}\right) - \mathbb{E}\left[y_{i,t} - \beta_{l,0} l_{i,t} \mid i_{i,t}, k_{i,t}\right]\right)}{\mathbb{E}\left[\left(l_{i,t} - \mathbb{E}\left[l_{i,t} \mid i_{i,t}, k_{i,t}\right]\right)^2\right]}, \quad (23)$$

that is equal to

$$-\Upsilon^{-1} \frac{1}{\sqrt{n}} \sum_{i=1}^{n} \Lambda_{2i} \left(\left(y_{i,t} - \beta_{l,0} l_{i,t}\right) - \mathbb{E}\left[y_{i,t} - \beta_{l,0} l_{i,t} \mid i_{i,t}, k_{i,t}\right]\right)$$

$$\Lambda_{2i} \equiv \frac{\Gamma}{\mathbb{E}\left[\left(l_{i,t} - \mathbb{E}\left[l_{i,t} \mid i_{i,t}, k_{i,t}\right]\right)^2\right]} \left(l_{i,t} - \mathbb{E}\left[l_{i,t} \mid i_{i,t}, k_{i,t}\right]\right).$$

3. Combining the adjustments, we conclude that the adjustment for the first step estimation is (17).

C.3 Derivation of Adjustments

To derive the adjustments for estimating $\mathbb{E}\left[y_t \mid i_t, k_t\right]$ and $\mathbb{E}\left[l_t \mid i_t, k_t\right]$, we calculate the path-wise derivative of Newey (1994) for each non-parametric function that enters in the moment condition. The moment function involves non-parametric regressions $\gamma_j(\nu_t)$ on the generated regressor ν_t. Hahn and Ridder (2013) show that the path-derivative adjustment has three components: an adjustment for the estimation of γ_j, an adjustment for the estimation of ν_t, and an adjustment for the effect of the estimation of ν_t on γ_j. In this case only the second and third adjustments have to be made, because γ_j is estimated by minimization. So no adjustment is needed for the estimation of γ_j (and g below).

After linearization the moment function for β_k is the inverse of the second derivative of Proposition 2 times

$$h\left(w, \gamma\left(\nu_t\right)\right) \equiv \left(Y_1 - \gamma_1\left(\nu_t\right) - \beta_{k,0}\left(Y_2 - \gamma_2\left(\nu_t\right)\right)\right) \left(\left(Y_2 - \gamma_2\left(\nu_t\right)\right) - \frac{\partial g\left(\nu_t\right)}{\partial \nu_t}\left(Y_3 - \gamma_3\left(\nu_t\right)\right)\right) \quad (24)$$

with $w = (Y_1, Y_2, Y_3, i_t)$, where $Y_3 \equiv k_t$, and $\gamma_3(v_t) \equiv \mathbb{E}[Y_3 | v_t]$. The corresponding population moment condition is

$$
\begin{aligned}
0 &= -\mathbb{E}\left[\left. \frac{\partial M(Y_1, Y_2, \phi_t, k_t; \beta_k)}{\partial \beta_k} \right|_{\beta_k = \beta_{k,0}} \right] \\
&= \mathbb{E}\left[(Y_1 - \gamma_1(\nu_t) - \beta_{k,0}(Y_2 - \gamma_2(\nu_t))) \left((Y_2 - \gamma_2(\nu_t)) - \frac{\partial g(\nu_t)}{\partial \nu_t} (k_t - \mathbb{E}[k_t | \nu_t]) \right) \right]. \qquad (25)
\end{aligned}
$$

C.3.1 Adjustment for Estimation of $\mathbb{E}[y_t | i_t, k_t]$

To account for estimating $\mathbb{E}[y_t | i_t, k_t]$, we note that the estimation of $\mathbb{E}[y_t | i_t, k_t]$ induces sampling variation in the estimator of

$$
\nu_t = \phi_t(i_t, k_t) - \beta_{k,0} k_t = \mathbb{E}[y_t | i_t, k_t] - \beta_{l,0} \mathbb{E}[l_t | i_t, k_t] - \beta_{k,0} k_t,
$$

The generated regressor v_t is the conditioning variable in the non-parametric regressions $\mathbb{E}[Y_j | v_t] = \gamma_j(v_t)$ for $j = 1, 2, 3$. Their contribution to the influence function is calculated as in Theorem 5 in Hahn and Ridder (2013). We do the calculation for each γ_j and sum the adjustments. We first find the derivatives of h with respect to the γ_j

$$
\begin{aligned}
\Psi_1 &\equiv \frac{\partial h(w, \gamma(\nu_t))}{\partial \gamma_1} = -(Y_2 - \gamma_2(\nu_t)) + \frac{\partial g(\nu_t)}{\partial \nu_t} (Y_3 - \gamma_3(\nu_t)), \\
\Psi_2 &\equiv \frac{\partial h(w, \gamma(\nu_t))}{\partial \gamma_2(\nu_t)} = \beta_{k,0} \left((Y_2 - \gamma_2(\nu_t)) - \frac{\partial g(\nu_t)}{\partial \nu_t} (Y_3 - \gamma_3(\nu_t)) \right) \\
&\qquad - (Y_1 - \gamma_1(\nu_t) - \beta_{k,0}(Y_2 - \gamma_2(\nu_t))), \\
\Psi_3 &\equiv \frac{\partial h(w, \gamma(\nu_t))}{\partial \gamma_3(\nu_t)} = (Y_1 - \gamma_1(\nu_t) - \beta_{k,0}(Y_2 - \gamma_2(\nu_t))) \frac{\partial g(\nu_t)}{\partial \nu_t},
\end{aligned}
$$

with $w = (Y_1, Y_2, Y_3, i_t, k_t)$. By Theorem 5 of Hahn and Ridder (2013) the generated-regressor adjustment, i.e. the adjustment for the effect of the generated regressor on γ_j is

$$
\sum_{j=1}^{3} \mathbb{E}\left[(\Psi_j - \kappa_j(\nu_t)) \frac{\partial \gamma_j(\nu_t)}{\partial \nu_t} + \frac{\partial \kappa_j(\nu_t)}{\partial \nu_t} (\gamma_j(i_t, k_t) - \gamma_j(v_t)) \right| i_t, k_t \right] (y_t - \mathbb{E}[y_t | i_t, k_t]) \qquad (26)
$$

where $\gamma_j(i_t, k_t) \equiv \mathbb{E}[Y_j | i_t, k_t]$ and $\kappa_j(\nu_t) \equiv \mathbb{E}[\Psi_j | \nu_t]$ for $j = 1, 2, 3$.

The generated regressor also enters directly in h with derivative

$$
\Psi_4 \equiv \frac{\partial h(w, \gamma(\nu_t))}{\partial \nu_t} = -(Y_1 - \gamma_1(\nu_t) - \beta_k(Y_2 - \gamma_2(\nu_t^*))) \frac{\partial^2 g(\nu_t)}{\partial \nu_t^2} (Y_3 - \gamma_3(\nu_t)). \qquad (27)
$$

with the adjustment calculated by Newey (1994)

$$
\mathbb{E}[\Psi_4 | i_t, k_t] (y_t - \mathbb{E}[y_t | i_t, k_t]). \qquad (28)
$$

It is straightforward to show that

$$\kappa_j(\nu_t) = 0, \quad j = 1, 2, 3$$

Therefore (26) is equal to

$$\mathbb{E}\left[\Psi_1 \frac{\partial \gamma_1(\nu_t)}{\partial \nu_t} + \Psi_2 \frac{\partial \gamma_2(\nu_t)}{\partial \nu_t} + \Psi_3 \frac{\partial \gamma_3(\nu_t)}{\partial \nu_t} + \Psi_4 \middle| i_t, k_t\right] (y_t - \mathbb{E}[y_t | i_t, k_t]),$$

which we will simplify by using (6), which implies that $\mathbb{E}[\xi_{t+1} | i_t, k_t] = 0$.

Note that

$$\begin{aligned}
&\mathbb{E}\left[\Psi_1 \frac{\partial \gamma_1(\nu_t)}{\partial \nu_t} \middle| i_t, k_t\right] \\
&= \mathbb{E}\left[-\left((Y_2 - \gamma_2(\nu_t)) - \frac{\partial g(\nu_t)}{\partial \nu_t}(Y_3 - \gamma_3(\nu_t))\right) \frac{\partial \gamma_1(\nu_t)}{\partial \nu_t} \middle| i_t, k_t\right] \\
&= -\left((k_{t+1} - \mathbb{E}[k_{t+1} | \nu_t]) - \frac{\partial g(\nu_t)}{\partial \nu_t}(k_t - \mathbb{E}[k_t | \nu_t])\right) \frac{\partial \gamma_1(\nu_t)}{\partial \nu_t},
\end{aligned}$$

where if (OP 1) does not hold, the k_{t+1} in the last line becomes $\mathbb{E}[k_{t+1}|i_t, k_t]$.

$$\begin{aligned}
&\mathbb{E}\left[\Psi_2 \frac{\partial \gamma_2(\nu_t)}{\partial \nu_t} \middle| i_t, k_t\right] \\
&= \beta_{k,0}\mathbb{E}\left[\left((Y_2 - \gamma_2(\nu_t)) - \frac{\partial g(\nu_t)}{\partial \nu_t}(Y_3 - \gamma_3(\nu_t))\right) \frac{\partial \gamma_2(\nu_t)}{\partial \nu_t} \middle| i_t, k_t\right] \\
&- \mathbb{E}\left[(Y_1 - \gamma_1(\nu_t) - \beta_{k,0}(Y_2 - \gamma_2(\nu_t))) \frac{\partial \gamma_2(\nu_t)}{\partial \nu_t} \middle| i_t, k_t\right] \\
&= \beta_{k,0}\left((k_{t+1} - \mathbb{E}[k_{t+1} | \nu_t]) - \frac{\partial g(\nu_t)}{\partial \nu_t}(k_t - \mathbb{E}[k_t | \nu_t])\right) \frac{\partial \gamma_2(\nu_t)}{\partial \nu_t} \\
&- \mathbb{E}[\xi_{t+1} + \eta_{t+1} | i_t, k_t] \frac{\partial \gamma_2(\nu_t)}{\partial \nu_t} \\
&= \beta_{k,0}\left((k_{t+1} - \mathbb{E}[k_{t+1} | \nu_t]) - \frac{\partial g(\nu_t)}{\partial \nu_t}(k_t - \mathbb{E}[k_t | \nu_t])\right) \frac{\partial \gamma_2(\nu_t)}{\partial \nu_t},
\end{aligned}$$

because $\mathbb{E}[\xi_{t+1} + \eta_{t+1} | i_t, k_t] = 0$. If (OP 1) does not hold, the k_{t+1} in the last line and the second last line of the above expression becomes $\mathbb{E}[k_{t+1}|i_t, k_t]$.

$$\begin{aligned}
&\mathbb{E}\left[\Psi_3 \frac{\partial \gamma_3(\nu_t)}{\partial \nu_t} \middle| i_t, k_t\right] \\
&= \mathbb{E}\left[(Y_1 - \gamma_1(\nu_t) - \beta_k(Y_2 - \gamma_2(\nu_t))) \frac{\partial g(\nu_t)}{\partial \nu_t} \middle| i_t, k_t\right] \frac{\partial \gamma_3(\nu_t)}{\partial \nu_t} \\
&= \mathbb{E}[\xi_{t+1} + \eta_{t+1} | i_t, k_t] \frac{\partial g(\nu_t)}{\partial \nu_t} \frac{\partial \gamma_3(\nu_t)}{\partial \nu_t} = 0,
\end{aligned}$$

because $\mathbb{E}\left[\left.\xi_{t+1}+\eta_{t+1}\right| i_t,k_t\right]=0$, and

$$\begin{aligned}
&\mathbb{E}\left[\left.\Psi_4\right| i_t,k_t\right]\\
&=-\mathbb{E}\left[\left.\left(Y_1-\gamma_1\left(\nu_t\right)-\beta_{k,0}\left(Y_2-\gamma_2\left(\nu_t\right)\right)\right)\frac{\partial^2 g\left(\nu_t\right)}{\partial\nu_t^2}\left(k_t-\mathbb{E}\left[\left.k_t\right|\nu_t\right]\right)\right| i_t,k_t\right]\\
&=-\mathbb{E}\left[\left.\xi_{t+1}+\eta_{t+1}\right| i_t,k_t\right]\frac{\partial^2 g\left(\nu_t\right)}{\partial\nu_t^2}\left(k_t-\mathbb{E}\left[\left.k_t\right|\nu_t\right]\right)=0.
\end{aligned}$$

These calculations show that the adjustment for the estimation of of $\mathbb{E}\left[\left.y_t\right| i_t,k_t\right]$ is equal to

$$\begin{aligned}
&\mathbb{E}\left[\left.\Psi_1\frac{\partial\gamma_1\left(\nu_t\right)}{\partial\nu_t}+\Psi_2\frac{\partial\gamma_2\left(\nu_t\right)}{\partial\nu_t}+\Psi_3\frac{\partial\gamma_3\left(\nu_t\right)}{\partial\nu_t}+\Psi_4\right| i_t,k_t\right]\left(y_t-\mathbb{E}\left[\left.y_t\right| i_t,k_t\right]\right)\\
&=\left(\left(k_{t+1}-\mathbb{E}\left[\left.k_{t+1}\right|\nu_t\right]\right)-\frac{\partial g\left(\nu_t\right)}{\partial\nu_t}\left(k_t-\mathbb{E}\left[\left.k_t\right|\nu_t\right]\right)\right)\left(-\frac{\partial\gamma_1\left(\nu_t\right)}{\partial\nu_t}+\beta_{k,0}\frac{\partial\gamma_2\left(\nu_t\right)}{\partial\nu_t}\right)\left(y_t-\mathbb{E}\left[\left.y_t\right| i_t,k_t\right]\right)\\
&=-\left(\left(k_{t+1}-\mathbb{E}\left[\left.k_{t+1}\right|\nu_t\right]\right)-\frac{\partial g\left(\nu_t\right)}{\partial\nu_t}\left(k_t-\mathbb{E}\left[\left.k_t\right|\nu_t\right]\right)\right)\frac{\partial g(\nu_t)}{\partial\nu_t}\left(y_t-\mathbb{E}\left[\left.y_t\right| i_t,k_t\right]\right)\\
&=\delta_1\left(i_t,k_t\right)\left(y_t-\mathbb{E}\left[\left.y_t\right| i_t,k_t\right]\right), \qquad (29)
\end{aligned}$$

where if (OP 1) does not hold k_{t+1} becomes $\mathbb{E}[k_{t+1}|i_t,k_t]$.

C.3.2 Adjustment for Estimation of $\mathbb{E}\left[\left.l_t\right| i_t,k_t\right]$

To derive the adjustment for estimating $\mathbb{E}\left[\left.l_t\right| i_t,k_t\right]$, we work with the moment function $h\left(w,\gamma\left(\nu_t\right)\right)$ in (24). Note that the estimation of $\mathbb{E}\left[\left.l_t\right| i_t,k_t\right]$ induces sampling variation in

$$\nu_t=\mathbb{E}\left[\left.y_t\right| i_t,k_t\right]-\beta_{l,0}\mathbb{E}\left[\left.l_t\right| i_t,k_t\right]-\beta_{k,0}k_t.$$

so that the adjustment is

$$-\beta_{l,0}\delta_1\left(i_t,k_t\right)\left(l_t-\mathbb{E}\left[\left.l_t\right| i_t,k_t\right]\right) \qquad (30)$$

with $\delta_1\left(i_t,k_t\right)$ given in (29).

C.3.3 Adjustment for Estimation of $\widehat{\beta}_l$ through ν_t

To calculate the adjustment for estimating $\widehat{\beta}_l$ as a component of

$$\nu_t=\mathbb{E}\left[\left.y_t\right| i_t,k_t\right]-\beta_{l,0}\mathbb{E}\left[\left.l_t\right| i_t,k_t\right]-\beta_{k,0}k_t$$

we can use

$$\frac{\partial h\left(w,\gamma\left(\nu_t\right)\right)}{\partial\nu_t}\frac{\partial\nu_t}{\partial\beta_{l,0}}=-\frac{\partial h\left(w,\gamma\left(\nu_t\right)\right)}{\partial\nu_t}\mathbb{E}\left[\left.l_t\right| i_t,k_t\right]$$

and conclude that the adjustment for the estimation of $\widehat{\beta}_l$ is by Theorem 4 of Hahn and Ridder (2013)

$$-\Upsilon^{-1}\mathbb{E}\left[\delta_1\left(i_t,k_t\right)\mathbb{E}\left[\left.l_t\right| i_t,k_t\right]\right]\sqrt{n}(\widehat{\beta}_l-\beta_{l,0}).$$

Supplemental Appendix to: Identification and the Influence Function of Olley and Pakes' (1996) Production Function Estimator

Jinyong Hahn
Department of Economics, UCLA

Zhipeng Liao
Department of Economics, UCLA

Geert Ridder
Department of Economics, USC

March 19, 2021

Abstract

This supplemental appendix contains additional technical results of Hahn, Liao, and Ridder (2021). Section SA provides detailed description of the multi-step estimator of $\beta_{k,0}$ mentioned in Hahn, Liao, and Ridder (2021). Section SB derives the asymptotic properties of the multi-step estimator and provides consistent estimation of its asymptotic variance. Section SC contains assumptions and auxiliary lemmas used in Section SB.

SA The multi-step Series Estimator

In this section, we describe the multi-step procedure on estimating $\beta_{k,0}$. The model can be rewritten as

$$y_{1,i} = l_{1,i}\beta_{l,0} + \phi\left(i_{1,i}, k_{1,i}\right) + \eta_{1,i}, \tag{SA.1}$$

$$y^*_{2,i} = k_{2,i}\beta_{k,0} + g(\nu_{1,i}) + u_{2,i}, \tag{SA.2}$$

where $u_{2,i} \equiv \xi_{2,i} + \eta_{2,i}$, $y^*_{2,i} \equiv y_{2,i} - l_{2,i}\beta_{l,0}$ and $\nu_{1,i} \equiv \phi\left(i_{1,i}, k_{1,i}\right) - k_{1,i}\beta_{k,0}$, and $\xi_{2,i}$ is defined in equation (1) of the paper. The following restrictions are maintained throughout this appendix

$$\mathbb{E}\left[\eta_{1,i}\right| i_{1,i}, k_{1,i}] = 0 \quad \text{and} \quad \mathbb{E}\left[u_{2,i}\right| i_{1,i}, k_{1,i}] = 0. \tag{SA.3}$$

For any β_k, let

$$\nu_{1,i}(\beta_k) \equiv \phi\left(i_{1,i}, k_{1,i}\right) - k_{1,i}\beta_k \quad \text{and} \quad g(\nu_{1,i}(\beta_k); \beta_k) \equiv \mathbb{E}[y^*_{2,i} - \beta_k k_{2,i}|\nu_{1,i}(\beta_k)]. \tag{SA.4}$$

Then by definition

$$\nu_{1,i} = \nu_{1,i}(\beta_{k,0}) \quad \text{and} \quad g(\nu_{1,i}) = g(\nu_{1,i}(\beta_{k,0}); \beta_{k,0}). \tag{SA.5}$$

The unknown parameters are $\beta_{l,0}$, $\beta_{k,0}$, $\phi(\cdot)$ and $g(\cdot;\beta_k)$ for any β_k in Θ_k, where Θ_k is a compact subset of $\mathbb{R}$ which contains $\beta_{k,0}$ as an interior point.

Suppose that we have data $\{(y_{t,i}, i_{t,i}, k_{t,i}, l_{t,i})_{t=1,2}\}_{i=1}^n$ and a preliminary estimator $\hat{\beta}_l$ of $\beta_{l,0}$. The asymptotic theory established here allows for a generic estimator of $\beta_{l,0}$, as long as certain regularity conditions (i.e., Assumptions SC1(iii) and SC4(i) in Section SC) hold. For example, $\hat{\beta}_l$ may be obtained from the partially linear regression proposed in Olley and Pakes (1996), or from the GMM estimation proposed in Ackerberg, Caves, and Frazer (2015). The unknown parameters $\beta_{k,0}$, $\phi(\cdot)$ and $g(\cdot;\beta_k)$ for any $\beta_k \in \Theta_k$ are estimated through the following multi-step estimation procedure described in the paper.

Step 1. Estimating $\phi(\cdot)$. Let $P_1(x_{1,i}) \equiv (p_{1,1}(x_{1,i}), \ldots, p_{1,m_1}(x_{1,i}))'$ be an m_1-dimensional approximating functions of $x_{1,i}$ where $x_{1,i} \equiv (i_{1,i}, k_{1,i})$. Define $\hat{y}_{1,i} \equiv y_{1,i} - l_{1,i}\hat{\beta}_l$. Then the unknown function $\phi(\cdot)$ is estimated by

$$\hat{\phi}\left(\cdot\right) \equiv P_1\left(\cdot\right)' (\mathbf{P}_1'\mathbf{P}_1)^{-1}(\mathbf{P}_1'\hat{\mathbf{Y}}_1) \tag{SA.6}$$

where $\mathbf{P}_1 \equiv (P_1(x_{1,1}), \ldots, P_1(x_{1,n}))'$ and $\hat{\mathbf{Y}}_1 \equiv (\hat{y}_{1,1}, \ldots, \hat{y}_{1,n})'$.

Step 2. Estimating $g(\cdot;\beta_k)$ for any $\beta_k \in \Theta_k$. With $\hat{\beta}_l$ and $\hat{\phi}(\cdot)$ obtained in the first step, one can estimate $y^*_{2,i}$ by $\hat{y}^*_{2,i} \equiv y_{2,i} - \hat{\beta}_l l_{2,i}$ and estimate $\nu_{1,i}(\beta_k)$ by $\hat{\nu}_{1,i}(\beta_k) \equiv \hat{\phi}(x_{1,i}) - \beta_k k_{1,i}$. Let

$P_2(\nu) \equiv (p_{2,1}(\nu), \ldots, p_{2,m_2}(\nu))'$ be an m_2-dimensional approximating functions. Then $g(\cdot;\beta_k)$ is estimated by

$$\hat{g}(\cdot;\beta_k) \equiv P_2(\cdot)'\hat{\beta}_g(\beta_k), \text{ where } \hat{\beta}_g(\beta_k) \equiv (\hat{\mathbf{P}}_2(\beta_k)'\hat{\mathbf{P}}_2(\beta_k))^{-1}\hat{\mathbf{P}}_2(\beta_k)'\hat{\mathbf{Y}}_2^*(\beta_k) \tag{SA.7}$$

where $\hat{\mathbf{P}}_2(\beta_k) \equiv (P_2(\hat{\nu}_{1,1}(\beta_k)), \ldots, P_2(\hat{\nu}_{1,n}(\beta_k)))'$ and $\hat{\mathbf{Y}}_2^*(\beta_k) \equiv (\hat{y}_{2,1}^* - \beta_k k_{2,1}, \ldots, \hat{y}_{2,n}^* - \beta_k k_{2,n})'$.

Step 3. Estimating $\beta_{k,0}$. The finite dimensional parameter $\beta_{k,0}$ is estimated by $\hat{\beta}_k$ through the following semiparametric nonlinear regression

$$\hat{\beta}_k \equiv \arg\min_{\beta_k \in \Theta_k} n^{-1} \sum_{i=1}^n \hat{\ell}_i(\beta_k)^2, \text{ where } \hat{\ell}_i(\beta_k) \equiv \hat{y}_{2,i}^* - k_{2,i}\beta_k - \hat{g}(\hat{\nu}_{1,i}(\beta_k);\beta_k). \tag{SA.8}$$

We shall derive the root-n normality of $\hat{\beta}_k$ and provide asymptotically valid inference for $\beta_{k,0}$.

SB Asymptotic Properties of $\hat{\beta}_k$

In this section, we derive the asymptotic properties of $\hat{\beta}_k$. The consistency and the asymptotic distribution of $\hat{\beta}_k$ are presented in Subsection SB.1. In Subsection SB.2, we provide a consistent estimator of the asymptotic variance of $\hat{\beta}_k$ which can be used to construct confidence interval for $\beta_{k,0}$. Proofs of the consistency and the asymptotic normality of $\hat{\beta}_k$, and the consistency of the standard deviation estimator are included in Subsection SB.3.

SB.1 Consistency and asymptotic normality

To show the consistency of $\hat{\beta}_k$, we use the standard arguments for showing the consistency of the extremum estimator which requires two primitive conditions: (i) the identification uniqueness condition of the unknown parameter $\beta_{k,0}$; and (ii) the convergence of the estimation criterion function $n^{-1}\sum_{i=1}^n \hat{\ell}_i(\beta_k)^2$ to the population criterion function uniformly over $\beta_k \in \Theta_k$. We impose the identification uniqueness condition of $\beta_{k,0}$ in condition (SB.9) below, which can be verified under low-level sufficient conditions. The uniform convergence of the estimation criterion function is proved in Lemma SB1 in Subsection SB.3.

Lemma SB1. *Let $\ell_i(\beta_k) \equiv y_{2,i} - l_{2,i}\beta_{l,0} - \beta_k k_{2,i} - g(\nu_{1,i}(\beta_k);\beta_k)$ for any $\beta_k \in \Theta_k$. Suppose that for any $\varepsilon > 0$, there exists a constant $\delta_\varepsilon > 0$ such that*

$$\inf_{\{\beta_k \in \Theta_k : |\beta_k - \beta_{k,0}| \geq \varepsilon\}} \mathbb{E}\left[\ell_i(\beta_k)^2 - \ell_i(\beta_{k,0})^2\right] > \delta_\varepsilon. \tag{SB.9}$$

Then under Assumptions SC1 and SC2 in Section SC, we have $\hat{\beta}_k = \beta_{k,0} + o_p(1)$.

The asymptotic normality of $\hat{\beta}_k$ can be derived from its first-order condition:

$$n^{-1}\sum_{i=1}^{n}\hat{\ell}_i(\hat{\beta}_k)\left(k_{2,i}+\frac{\partial\hat{g}(\hat{\nu}_{1,i}(\hat{\beta}_k);\hat{\beta}_k)}{\partial\beta_k}\right)=0 \tag{SB.10}$$

where for any $\beta_k\in\Theta_k$

$$\frac{\partial\hat{g}(\hat{\nu}_{1,i}(\beta_k);\beta_k)}{\partial\beta_k}=\hat{\beta}_g(\beta_k)'\frac{\partial P_2(\hat{\nu}_{1,i}(\beta_k))}{\partial\beta_k}+P_2(\hat{\nu}_{1,i}(\beta_k))'\frac{\partial\hat{\beta}_g(\beta_k)}{\partial\beta_k}. \tag{SB.11}$$

By the definition of $\hat{g}(\hat{\nu}_{1,i}(\hat{\beta}_k);\hat{\beta}_k)$ in (SA.7), we can write

$$n^{-1}\sum_{i=1}^{n}P_2(\hat{\nu}_{1,i}(\hat{\beta}_k))\hat{g}(\hat{\nu}_{1,i}(\hat{\beta}_k);\hat{\beta}_k)=n^{-1}\sum_{i=1}^{n}P_2(\hat{\nu}_{1,i}(\hat{\beta}_k))(\hat{y}^*_{2,i}-k_{2,i}\hat{\beta}_k)$$

which implies that

$$n^{-1}\sum_{i=1}^{n}\hat{\ell}_i(\hat{\beta}_k)P_2(\hat{\nu}_{1,i}(\hat{\beta}_k))=0.$$

Therefore, the first-order condition (SB.10) can be reduced to

$$n^{-1}\sum_{i=1}^{n}\hat{\ell}_i(\hat{\beta}_k)\left(k_{2,i}-k_{1,i}\hat{\beta}_g(\hat{\beta}_k)'\frac{\partial P_2(\hat{\nu}_{1,i}(\hat{\beta}_k))}{\partial\nu}\right)=0 \tag{SB.12}$$

which slightly simplifies the derivation of the asymptotic normality of $\hat{\beta}_k$.

Theorem SB1. *Let* $\varsigma_{1,i}\equiv k_{2,i}-\mathbb{E}[k_{2,i}|\nu_{1,i}]-g_1(\nu_{1,i})(k_{1,i}-\mathbb{E}[k_{1,i}|\nu_{1,i}])$ *where*

$$g_1(\nu)\equiv\frac{\partial g(\nu)}{\partial\nu}. \tag{SB.13}$$

Suppose that

$$\Upsilon\equiv\mathbb{E}\left[\varsigma_{1,i}^2\right]>0,\ \textit{where}\ \varsigma_{1,i}\equiv k_{2,i}-\mathbb{E}[k_{2,i}|\nu_{1,i}]-g_1(\nu_{1,i})(k_{1,i}-\mathbb{E}[k_{1,i}|\nu_{1,i}]). \tag{SB.14}$$

Then under (SB.9) in Lemma SB1, and Assumptions SC1, SC2 and SC3 in Section SC

$$\begin{aligned}\hat{\beta}_k-\beta_{k,0} &= \Upsilon^{-1}n^{-1}\sum_{i=1}^{n}u_{2,i}\varsigma_{1,i}\\ &\quad -\Upsilon^{-1}n^{-1}\sum_{i=1}^{n}\eta_{1,i}g_1(\nu_{1,i})\left(\varsigma_{1,i}-\varsigma_{2,i}\right)\\ &\quad -\Upsilon^{-1}\Gamma(\hat{\beta}_l-\beta_{l,0})+o_p(n^{-1/2}),\end{aligned} \tag{SB.15}$$

where $\Gamma \equiv \mathbb{E}\left[(l_{2,i} - l_{1,i}g_1(\nu_{1,i}))\,\varsigma_{1,i} + l_{1,i}g_1(\nu_{1,i})\varsigma_{2,i}\right]$ *and* $\varsigma_{2,i} \equiv k_{2,i} - \mathbb{E}[k_{2,i}|x_{1,i}]$. *Moreover*

$$n^{1/2}(\hat{\beta}_k - \beta_{k,0}) \to_d N(0, \Upsilon^{-1}\Omega\Upsilon^{-1}) \tag{SB.16}$$

where $\Omega \equiv \mathbb{E}\left[((u_{2,i} - \eta_{1,i}g_1(\nu_{1,i}))\varsigma_{1,i} - \Gamma\varepsilon_{1,i} + \eta_{1,i}g_1(\nu_{1,i})\varsigma_{2,i})^2\right]$.

Remark 1. The local identification condition of $\beta_{k,0}$ is imposed in (SB.14) which is important to ensure the root-n consistency of $\hat{\beta}_k$. This condition is verified in Proposition 2 of the paper. □

Remark 2. The random variable $\varepsilon_{1,i}$ in the definition of Ω is from the linear representation of the estimator error in $\hat{\beta}_l$, i.e.,

$$\hat{\beta}_l - \beta_{l,0} = n^{-1}\sum_{i=1}^{n}\varepsilon_{1,i} + o_p(n^{-1/2}) \tag{SB.17}$$

which is maintained in Assumption SC1(iii) in Section SC. The explicit form of $\varepsilon_{1,i}$ depends on the estimation procedure of $\beta_{l,0}$. For example, when $\beta_{l,0}$ is estimated by the partially linear regression proposed in Olley and Pakes (1996),[1]

$$\varepsilon_{1,i} = \frac{l_{1,i} - \mathbb{E}[l_{1,i}|\,x_{1,i}]}{\mathbb{E}[|l_{1,i} - \mathbb{E}[l_{1,i}|\,x_{1,i}]|^2]}\eta_{1,i}.$$

On the other hand, $\varepsilon_{1,i}$ may take different forms in different estimation procedures (under different identification condition on $\beta_{l,0}$), such as the GMM procedure in Ackerberg, Caves, and Frazer (2015). □

Remark 3. Since $\mathbb{E}[\varsigma_{1,i}|\nu_{1,i}] = 0$ for $j = 1, 2$,

$$\mathbb{E}\left[l_{2,i}\varsigma_{1,i}\right] = \mathbb{E}\left[(l_{2,i} - \mathbb{E}[l_{2,i}|\nu_{1,i}])\,\varsigma_{1,i}\right]$$

by the law of iterated expectation. Similarly

$$\mathbb{E}\left[l_{1,i}g_1(\nu_{1,i})\,(\varsigma_{1,i} - \varsigma_{2,i})\right] = \mathbb{E}\left[\mathbb{E}[l_{1,i}|x_{1,i}]g_1(\nu_{1,i})\,(\varsigma_{1,i} - \varsigma_{2,i})\right]$$

Therefore we can write

$$\begin{aligned} \Gamma &= \mathbb{E}\left[l_{2,i}\varsigma_{1,i} - l_{1,i}g_1(\nu_{1,i})\,(\varsigma_{1,i} - \varsigma_{2,i})\right] \\ &= \mathbb{E}\left[(l_{2,i} - \mathbb{E}[l_{2,i}|\nu_{1,i}])\,\varsigma_{1,i} - \mathbb{E}[l_{1,i}|x_{1,i}]g_1(\nu_{1,i})\,(\varsigma_{1,i} - \varsigma_{2,i})\right]. \end{aligned} \tag{SB.18}$$

[1]See Assumption SC5 and Lemma SC26 in Subsection SC.5 for the regularity conditions and the derivation for (SB.17).

When the perpetual inventory method (PIM) i.e., $k_{2,i} = (1-\delta)\, k_{1,i} + i_{1,i}$ holds, we have $\varsigma_{2,i} = 0$ for any $i = 1, \ldots, n$. Therefore, we deduce that

$$\Gamma = \mathbb{E}\left[(l_{2,i} - \mathbb{E}[l_{2,i}|\nu_{1,i}] - g_1(\nu_{1,i})\mathbb{E}[l_{1,i}|x_{1,i}])\, \varsigma_{1,i}\right] \tag{SB.19}$$

which appears in Proposition 3 of the paper. Moreover

$$\Omega = \mathbb{E}\left[((u_{2,i} - \eta_{1,i} g_1(\nu_{1,i}))\, \varsigma_{1,i} - \Gamma \varepsilon_{1,i})^2\right] \tag{SB.20}$$

under PIM. □

Remark 4. From the asymptotic expansion in (SB.15), we see that the asymptotic variance of $\hat{\beta}_k$ is determined by three components. The first component, $n^{-1}\sum_{i=1}^{n} u_{2,i}\varsigma_{1,i}$ comes from the third-step estimation with known $\nu_{1,i}$. The second and the third components are from the first-step estimation. Specifically, the second one, $n^{-1}\sum_{i=1}^{n} \eta_{1,i} g_1(\nu_{1,i})\, (\varsigma_{1,i} - \varsigma_{2,i})$ is from estimating $\phi(\cdot)$ in the first step, while the third component $\Gamma(\hat{\beta}_l - \beta_{l,0})$ is due to the estimation error in $\hat{\beta}_l$. □

SB.2 Consistent variance estimation

The asymptotic variance of $\hat{\beta}_k$ can be estimated using the estimators of $\varsigma_{1,i}$, $\varsigma_{2,i}$, $\varepsilon_{1,i}$, $\eta_{1,i}$, $u_{2,i}$ and $g_1(\nu_{1,i})$. First, it is clear that $g_1(\nu_{1,i})$ can be estimated by $\hat{g}_1(\hat{\nu}_{1,i}(\hat{\beta}_k); \hat{\beta}_k)$ where

$$\hat{g}_1(\hat{\nu}_{1,i}(\beta_k); \beta_k) \equiv \hat{\beta}_g(\beta_k)' \frac{\partial P_2(\hat{\nu}_{1,i}(\beta_k))}{\partial \nu} \text{ for any } \beta_k \in \Theta_k. \tag{SB.21}$$

Second, the residual $\varsigma_{1,i}$ can be estimated by

$$\hat{\varsigma}_{1,i} \equiv \Delta\hat{k}_{2,i} - P_2(\hat{\nu}_{1,i}(\hat{\beta}_k))'(\hat{\mathbf{P}}_2(\hat{\beta}_k)'\hat{\mathbf{P}}_2(\hat{\beta}_k))^{-1} \sum_{i=1}^{n} P_2(\hat{\nu}_{1,i}(\hat{\beta}_k))\Delta\hat{k}_{2,i} \tag{SB.22}$$

where $\Delta\hat{k}_{2,i} \equiv k_{2,i} - k_{1,i}\hat{g}_1(\hat{\nu}_{1,i}(\hat{\beta}_k); \hat{\beta}_k)$. Third, the residual $\varsigma_{2,i}$ can be estimated by

$$\hat{\varsigma}_{2,i} \equiv k_{2,i} - P_1\left(x_{1,i}\right)' \left(\mathbf{P}_1'\mathbf{P}_1\right)^{-1} \sum_{i=1}^{n} P_1\left(x_{1,i}\right) k_{2,i}. \tag{SB.23}$$

Given the estimated residual $\hat{\varsigma}_{1,i}$, the Hessian term Υ can be estimated by

$$\hat{\Upsilon}_n \equiv n^{-1} \sum_{i=1}^{n} \hat{\varsigma}_{1,i}^2. \tag{SB.24}$$

Moreover the Jacobian term Γ can be estimated by

$$\hat{\Gamma}_n \equiv n^{-1}\sum_{i=1}^{n}\left[(l_{2,i}-l_{1,i}\hat{g}_1(\hat{\nu}_{1,i}(\hat{\beta}_k);\hat{\beta}_k))\hat{\varsigma}_{1,i}+l_{1,i}\hat{g}_1(\hat{\nu}_{1,i}(\hat{\beta}_k);\hat{\beta}_k)\hat{\varsigma}_{2,i}\right]. \tag{SB.25}$$

Define $\hat{u}_{2,i} \equiv \hat{y}_{2,i}-l_{2,i}\hat{\beta}_l-k_{2,i}\hat{\beta}_k-\hat{g}(\hat{\nu}_{1,i}(\hat{\beta}_k);\hat{\beta}_k)$ and $\hat{\eta}_{1,i} \equiv y_{1,i}-l_{1,i}\hat{\beta}_l-\hat{\phi}(x_{1,i})$. Then Ω is estimated by

$$\hat{\Omega}_n \equiv n^{-1}\sum_{i=1}^{n}\left((\hat{u}_{2,i}-\hat{\eta}_{1,i}\hat{g}_1(\hat{\nu}_{1,i}(\hat{\beta}_k);\hat{\beta}_k))\hat{\varsigma}_{1,i}-\hat{\Gamma}_n\hat{\varepsilon}_{1,i}+\hat{\eta}_{1,i}\hat{g}_1(\hat{\nu}_{1,i}(\hat{\beta}_k);\hat{\beta}_k)\hat{\varsigma}_{2,i}\right)^2 \tag{SB.26}$$

where $\hat{\varepsilon}_{1,i}$ denotes the estimator of $\varepsilon_{1,i}$ for $i=1,\ldots,n$.

Theorem SB2. *Suppose that the conditions in Theorem SB1 hold. Then under Assumption SC4 in Section SC, we have*

$$\hat{\Upsilon}_n = \Upsilon + o_p(1) \quad \textit{and} \quad \hat{\Omega}_n = \Omega + o_p(1) \tag{SB.27}$$

and moreover

$$\frac{n^{1/2}(\hat{\beta}_k-\beta_{k,0})}{(\hat{\Upsilon}_n^{-1}\hat{\Omega}_n\hat{\Upsilon}_n^{-1})^{1/2}} \to_d N(0,1). \tag{SB.28}$$

SB.3 Proof of the asymptotic properties

In this subsection, we prove the main results presented in the previous subsection. Throughout this subsection, we use $C>1$ to denote a generic finite constant which does not depend on n, m_1 or m_2 but whose value may change in different places.

Proof of Lemma SB1. By (SC.80) in the proof of Lemma SC8 and Assumption SC2(i)

$$\sup_{\beta_k\in\Theta_k}\mathbb{E}\left[\ell_i(\beta_k)^2\right] \leq C \tag{SB.29}$$

which together with Lemma SC8 implies that

$$\sup_{\beta_k\in\Theta_k} n^{-1}\sum_{i=1}^{n}\ell_i(\beta_k)^2 = O_p(1). \tag{SB.30}$$

By the Markov inequality, Assumptions SC1(i, iii) and SC2(i), we obtain

$$n^{-1}\sum_{i=1}^{n}(\hat{y}^*_{2,i}-y^*_{2,i})^2 = (\hat{\beta}_l-\beta_l)^2 n^{-1}\sum_{i=1}^{n}l^2_{2,i} = O_p(n^{-1}). \tag{SB.31}$$

By the definition of $\hat{\ell}_i(\beta_k)$ and $\ell_i(\beta_k)$, we can write

$$
\begin{aligned}
& n^{-1}\sum_{i=1}^{n}\hat{\ell}_i(\beta_k)^2 - \mathbb{E}\left[\ell_i(\beta_k)^2\right] \\
= \; & n^{-1}\sum_{i=1}^{n}\left(\ell_i(\beta_k)^2 - \mathbb{E}\left[\ell_i(\beta_k)^2\right]\right) + 2n^{-1}\sum_{i=1}^{n}\ell_i(\beta_k)(\hat{y}^*_{2,i} - y^*_{2,i}) \\
& -2n^{-1}\sum_{i=1}^{n}\ell_i(\beta_k)(\hat{g}(\hat{\nu}_{1,i}(\beta_k);\beta_k) - g(\nu_{1,i}(\beta_k);\beta_k)) \\
& -2n^{-1}\sum_{i=1}^{n}(\hat{y}^*_{2,i} - y^*_{2,i})(\hat{g}(\hat{\nu}_{1,i}(\beta_k);\beta_k) - g(\nu_{1,i}(\beta_k);\beta_k)) \\
& +n^{-1}\sum_{i=1}^{n}(\hat{y}^*_{2,i} - y^*_{2,i})^2 + n^{-1}\sum_{i=1}^{n}(\hat{g}(\hat{\nu}_{1,i}(\beta_k);\beta_k) - g(\nu_{1,i}(\beta_k);\beta_k))^2,
\end{aligned}
$$

which together with Assumption SC2(vi), Lemma SC7, Lemma SC8, (SB.30), (SB.31) and the Cauchy-Schwarz inequality implies that

$$
\sup_{\beta_k\in\Theta_k}\left|n^{-1}\sum_{i=1}^{n}\hat{\ell}_i(\beta_k)^2 - \mathbb{E}\left[\ell_i(\beta_k)^2\right]\right| = o_p(1). \quad \text{(SB.32)}
$$

The consistency of $\hat{\beta}_k$ follows from its definition in (SA.8), (SB.32), the identification uniqueness condition of $\beta_{k,0}$ assumed in (SB.9) and the standard arguments of showing the consistency of the extremum estimator. *Q.E.D.*

Lemma SB2. *Let $g_{1,i} \equiv g_1(\nu_{1,i})$ and $\hat{J}_i(\beta_k) \equiv \hat{\ell}_i(\beta_k)\left(k_{2,i} - k_{1,i}\hat{g}_1(\hat{\nu}_{1,i}(\beta_k);\beta_k)\right)$ for any $\beta_k \in \Theta_k$, where $\hat{g}_1(\hat{\nu}_{1,i}(\beta_k);\beta_k)$ is defined in (SB.21). Then under Assumptions SC1, SC2 and SC3, we have*

$$
n^{-1}\sum_{i=1}^{n}\hat{J}_i(\beta_{k,0}) = n^{-1}\sum_{i=1}^{n}\left(u_{2,i}\varsigma_{1,i} - \eta_{1,i}g_1(\nu_{1,i})\left(\varsigma_{1,i} - \varsigma_{2,i}\right)\right) - \Gamma(\hat{\beta}_l - \beta_{l,0}) + o_p(n^{-1/2}). \quad \text{(SB.33)}
$$

Proof of Lemma SB2. By the definition of $\hat{\ell}_i(\beta_{k,0})$ and Lemma SC10,

$$
\begin{aligned}
& n^{-1}\sum_{i=1}^{n}\hat{\ell}_i(\beta_{k,0})\left(k_{2,i} - k_{1,i}\hat{g}_1(\hat{\nu}_{1,i}(\beta_{k,0});\beta_{k,0})\right) \\
& = n^{-1}\sum_{i=1}^{n}(\hat{y}^*_{2,i}(\beta_{k,0}) - g(\nu_{1,i}))(k_{2,i} - k_{1,i}g_{1,i}) \\
& \quad - n^{-1}\sum_{i=1}^{n}(\hat{g}(\hat{\nu}_{1,i}(\beta_{k,0});\beta_{k,0}) - g(\nu_{1,i}))(k_{2,i} - k_{1,i}g_{1,i}) + o_p(n^{-1/2}) \quad \text{(SB.34)}
\end{aligned}
$$

where $\hat{y}^*_{2,i}(\beta_{k,0}) \equiv y_{2,i} - l_{2,i}\hat{\beta}_l - k_{2,i}\beta_{k,0}$, and by Lemma SC12

$$\begin{aligned} & n^{-1}\sum_{i=1}^{n}(\hat{g}(\hat{\nu}_{1,i}(\beta_{k,0});\beta_{k,0}) - g(\nu_{1,i}))(k_{2,i} - k_{1,i}g_{1,i}) \\ = \; & n^{-1}\sum_{i=1}^{n} u_{2,i}\varphi(\nu_{1,i}) - \mathbb{E}[l_{2,i}\varphi(\nu_{1,i})](\hat{\beta}_l - \beta_{l,0}) \\ & + n^{-1}\sum_{i=1}^{n} g_{1,i}(\hat{\phi}(x_{1,i}) - \phi(x_{1,i}))\varsigma_{1,i} + o_p(n^{-1/2}), \end{aligned} \tag{SB.35}$$

where $\varphi(\nu_{1,i}) \equiv \mathbb{E}[k_{2,i}|\nu_{1,i}] - \mathbb{E}[k_{1,i}|\nu_{1,i}]g_{1,i}$. By the definition of $\hat{y}^*_{2,i}(\beta_{k,0})$, we get

$$\begin{aligned} & n^{-1}\sum_{i=1}^{n}(\hat{y}^*_{2,i}(\beta_{k,0}) - g(\nu_{1,i}))(k_{2,i} - k_{1,i}g_{1,i}) \\ = \; & n^{-1}\sum_{i=1}^{n} u_{2,i}(k_{2,i} - k_{1,i}g_{1,i}) - (\hat{\beta}_l - \beta_{l,0})n^{-1}\sum_{i=1}^{n} l_{2,i}(k_{2,i} - k_{1,i}g_{1,i}) \\ = \; & n^{-1}\sum_{i=1}^{n} u_{2,i}(k_{2,i} - k_{1,i}g_{1,i}) - (\hat{\beta}_l - \beta_{l,0})\mathbb{E}[l_{2,i}(k_{2,i} - k_{1,i}g_{1,i})] + o_p(n^{-1/2}) \end{aligned} \tag{SB.36}$$

where the second equality is by Assumption SC1(iii) and

$$n^{-1}\sum_{i=1}^{n} l_{2,i}(k_{2,i} - k_{1,i}g_{1,i}) = \mathbb{E}[l_{2,i}(k_{2,i} - k_{1,i}g_{1,i})] + O_p(n^{-1/2})$$

which holds by the Markov inequality, Assumptions SC1(i) and SC2(i, ii). Therefore by (SB.34), (SB.35) and (SB.36), we obtain

$$\begin{aligned} & n^{-1}\sum_{i=1}^{n}\hat{\ell}_i(\beta_{k,0})\left(k_{2,i} - k_{1,i}\hat{g}_1(\hat{\nu}_{1,i}(\beta_{k,0});\beta_{k,0})\right) \\ = \; & n^{-1}\sum_{i=1}^{n} u_{2,i}\varsigma_{1,i} - (\hat{\beta}_l - \beta_{l,0})\mathbb{E}[l_{2,i}\varsigma_{1,i}] - n^{-1}\sum_{i=1}^{n} g_{1,i}(\hat{\phi}(x_{1,i}) - \phi(x_{1,i}))\varsigma_{1,i} + o_p(n^{-1/2}) \end{aligned} \tag{SB.37}$$

The claim of the lemma follows from (SB.37) and Lemma SC13. *Q.E.D.*

Lemma SB3. *Under Assumptions SC1, SC2 and SC3, we have*

$$n^{-1}\sum_{i=1}^{n}(\hat{J}_i(\hat{\beta}_k) - \hat{J}_i(\beta_{k,0})) = -(\hat{\beta}_k - \beta_{k,0})\left(\mathbb{E}[\varsigma^2_{1,i}] + o_p(1)\right) + o_p(n^{-1/2}).$$

Proof of Lemma SB3. First note that by the definition of $\hat{J}_i(\beta_k)$ and $\hat{\ell}_i(\beta_k)$, we can write

$$
\begin{aligned}
& n^{-1}\sum_{i=1}^{n}(\hat{J}_i(\hat{\beta}_k)-\hat{J}_i(\beta_{k,0})) \\
= \quad & -(\hat{\beta}_k-\beta_{k,0})n^{-1}\sum_{i=1}^{n}k_{2,i}(k_{2,i}-k_{1,i}\hat{g}_1(\hat{\nu}_{1,i}(\hat{\beta}_k);\hat{\beta}_k)) \\
& -n^{-1}\sum_{i=1}^{n}(\hat{g}(\hat{\nu}_{1,i}(\hat{\beta}_k);\hat{\beta}_k)-\hat{g}(\hat{\nu}_{1,i}(\beta_{k,0});\beta_{k,0}))(k_{2,i}-k_{1,i}\hat{g}_1(\hat{\nu}_{1,i}(\beta_{k,0});\beta_{k,0})) \\
& -n^{-1}\sum_{i=1}^{n}u_{2,i}k_{1,i}(\hat{g}_1(\hat{\nu}_{1,i}(\hat{\beta}_k);\hat{\beta}_k)-\hat{g}_1(\hat{\nu}_{1,i}(\beta_{k,0});\beta_{k,0})) \\
& +(\hat{\beta}_l-\beta_{l,0})n^{-1}\sum_{i=1}^{n}l_{2,i}k_{1,i}(\hat{g}_1(\hat{\nu}_{1,i}(\hat{\beta}_k);\hat{\beta}_k)-\hat{g}_1(\hat{\nu}_{1,i}(\beta_{k,0});\beta_{k,0})) \qquad \text{(SB.38)}
\end{aligned}
$$

which together with Assumption SC1(iii), Lemma SC17, Lemma SC19 and Lemma SC23 implies that

$$
\begin{aligned}
n^{-1}\sum_{i=1}^{n}(\hat{J}_i(\hat{\beta}_k)-\hat{J}_i(\beta_{k,0})) &= -(\hat{\beta}_k-\beta_{k,0})\mathbb{E}[k_{2,i}(k_{2,i}-k_{1,i}g_{1,i})] \\
&\quad +(\hat{\beta}_k-\beta_{k,0})\left[\mathbb{E}[k_{1,i}g_{1,i}\varsigma_{1,i}]+\mathbb{E}[k_{2,i}\varphi(\nu_{1,i})]\right] \\
&\quad +(\hat{\beta}_k-\beta_{k,0})o_p(1)+o_p(n^{-1/2}) \\
&= -(\hat{\beta}_k-\beta_{k,0})\left(\mathbb{E}[\varsigma_{1,i}^2]+o_p(1)\right)+o_p(n^{-1/2})
\end{aligned}
$$

which finishes the proof. *Q.E.D.*

Proof of Theorem SB1. By Assumptions SC1(ii, iii) and SC2(i, ii), and Hölder's inequality

$$
|\Gamma| \leq \mathbb{E}\left[|(l_{2,i}-l_{1,i}g_1(\nu_{1,i}))\,\varsigma_{1,i}+l_{1,i}g_1(\nu_{1,i})\varsigma_{2,i}|\right] \leq C \qquad \text{(SB.39)}
$$

and

$$
\begin{aligned}
\Omega &= \mathbb{E}\left[\left((u_{2,i}-\eta_{1,i}g_1(\nu_{1,i}))\varsigma_{1,i}-\Gamma\varepsilon_{1,i}+\eta_{1,i}g_1(\nu_{1,i})\varsigma_{2,i}\right)^2\right] \\
&\leq C\mathbb{E}[u_{2,i}^4+\eta_{1,i}^4+k_{1,i}^4+k_{2,i}^4+\varepsilon_{1,i}^2] \leq C. \qquad \text{(SB.40)}
\end{aligned}
$$

By Assumption SC1(i), (SB.40) and the Lindeberg–Lévy central limit theorem,

$$
n^{-1/2}\sum_{i=1}^{n}\left((u_{2,i}-\eta_{1,i}g_1(\nu_{1,i}))\varsigma_{1,i}-\Gamma\varepsilon_{1,i}+\eta_{1,i}g_1(\nu_{1,i})\varsigma_{2,i}\right) \to_d N(0,\Omega). \qquad \text{(SB.41)}
$$

By (SB.12), Assumption SC1(iii), Lemma SB2 and Lemma SB3, we can write

$$
\begin{aligned}
0 &= n^{-1}\sum_{i=1}^{n}\hat{J}_i(\beta_{k,0}) + n^{-1}\sum_{i=1}^{n}(\hat{J}_i(\hat{\beta}_k) - \hat{J}_i(\beta_{k,0})) \\
&= n^{-1}\sum_{i=1}^{n}\left(u_{2,i}\varsigma_{1,i} - \eta_{1,i}g_1(\nu_{1,i})\left(\varsigma_{1,i} - \varsigma_{2,i}\right)\right) - \Gamma n^{1/2}(\hat{\beta}_l - \beta_{l,0}) \\
&\quad -(\hat{\beta}_k - \beta_{k,0})\left(\mathbb{E}[\varsigma_{1,i}^2] + o_p(1)\right) + o_p(n^{-1/2}) \qquad \text{(SB.42)}
\end{aligned}
$$

which together with (SB.14) and (SB.41) implies that

$$
\hat{\beta}_k - \beta_{k,0} = \Upsilon^{-1}n^{-1}\sum_{i=1}^{n}\left(u_{2,i}\varsigma_{1,i} - \eta_{1,i}g_1(\nu_{1,i})\left(\varsigma_{1,i} - \varsigma_{2,i}\right)\right) - \Upsilon^{-1}\Gamma(\hat{\beta}_l - \beta_{l,0}) + o_p(n^{-1/2}). \quad \text{(SB.43)}
$$

This proves (SB.15). The claim in (SB.16) follows from Assumption SC1(iii), (SB.41) and (SB.43). *Q.E.D.*

Proof of Theorem SB2. The results in (SB.27) are proved in Lemma SC25(i, iii), which together with Theorem SB1, Assumption SC4(iii) and the Slutsky Theorem proves the claim in (SB.28). *Q.E.D.*

SC Auxiliary Results

In this section, we provide the auxiliary results which are used to show Lemma SB1, Theorem SB1 and Theorem SB2. The conditions (SB.9) and (SB.14) are assumed throughout this section. The following notations are used throughout this section. We use $\|\cdot\|_2$ to denote the L_2-norm under the joint distribution of $(y_{t,i}, i_{t,i}, k_{t,i}, l_{t,i})_{t=1,2}$, $\|\cdot\|$ to denote the Euclidean norm and $\|\cdot\|_S$ to denote the matrix operator norm. For any real symmetric square matrix A, we use $\lambda_{\min}(A)$ and $\lambda_{\max}(A)$ to denote the smallest and largest eigenvalues of A respectively. Throughout this appendix, we use $C > 1$ to denote a generic finite constant which does not depend on n, m_1 or m_2 but whose value may change in different places.

SC.1 The asymptotic properties of the first-step estimators

Let $Q_{m_1} \equiv \mathbb{E}\left[P_1(x_{1,i})P_1(x_{1,i})'\right]$. The following assumptions are needed for studying the first-step estimator $\hat{\phi}(\cdot)$.

Assumption SC1. *(i) The data* $\{(y_{t,i}, i_{t,i}, k_{t,i}, l_{t,i})_{t=1,2}\}_{i=1}^{n}$ *are i.i.d.; (ii)* $\mathbb{E}[\eta_{1,i}|\, x_{1,i}] = 0$ *and*

$\mathbb{E}[l_{1,i}^2 + \eta_{1,i}^4 \big| x_{1,i}] \leq C$; *(iii) there exist i.i.d. random variables* $\varepsilon_{1,i}$ *with* $\mathbb{E}[\varepsilon_{1,i}^4] \leq C$ *such that*

$$\hat{\beta}_l - \beta_{l,0} = n^{-1} \sum_{i=1}^{n} \varepsilon_{1,i} + o_p(n^{-1/2});$$

(iv) there exist $r_\phi > 0$ *and* $\beta_{\phi,m} \in \mathbb{R}^m$ *such that* $\sup_{x \in \mathcal{X}} |\phi_m(x) - \phi(x)| = O(m^{-r_\phi})$ *where* $\phi_m(x) \equiv P_1(x)' \beta_{\phi,m}$ *and* $\mathcal{X}$ *denotes the support of* $x_{1,i}$ *which is compact; (v)* $C^{-1} \leq \lambda_{\min}(Q_{m_1}) \leq \lambda_{\max}(Q_{m_1}) \leq C$ *uniformly over* m_1*; (vi)* $m_1^2 n^{-1} + n^{1/2} m_1^{-r_\phi} = o(1)$ *and* $\log(m_1)\xi_{0,m_1}^2 n^{-1} = o(1)$ *where* ξ_{0,m_1} *is a nondecreasing sequence such that* $\sup_{x \in \mathcal{X}} \|P_1(x)\| \leq \xi_{0,m_1}$.

Assumption SC1(iii) assumes that there exists a root-n consistent estimator $\hat{\beta}_l$ of $\beta_{l,0}$. Different estimation procedures of $\hat{\beta}_l$ may give different forms for $\varepsilon_{1,i}$. For example, $\hat{\beta}_l$ may be obtained together with the nonparametric estimator of $\phi(\cdot)$ in the partially linear regression proposed in Olley and Pakes (1996), or from the GMM estimation proposed in Ackerberg, Caves, and Frazer (2015). Therefore, the specific form of $\varepsilon_{1,i}$ has to be derived case by case.[2] The rest conditions in Assumption SC1 are fairly standard in series estimation; see, for example, Andrews (1991), Newey (1997) and Chen (2007).[3] In particular, condition (iv) specifies the precision for approximating the unknown function $\phi(\cdot)$ via approximating functions, for which comprehensive results are available from numerical approximation theory.

The properties of the first-step estimator $\hat{\phi}(\cdot)$ are presented in the following lemma.

Lemma SC4. *Under Assumption SC1, we have*

$$n^{-1} \sum_{i=1}^{n} |\hat{\phi}(x_{1,i}) - \phi(x_{1,i})|^2 = O_p(m_1 n^{-1}) \tag{SC.44}$$

and moreover

$$\sup_{x_1 \in \mathcal{X}} |\hat{\phi}(x_1) - \phi(x_1)| = O_p(\xi_{0,m_1} m_1^{1/2} n^{-1/2}). \tag{SC.45}$$

Proof of Lemma SC4. Under Assumption SC1(i, v, vi), we can invoke Lemma 6.2 in Belloni, Chernozhukov, Chetverikov, and Kato (2015) to obtain

$$\left\| n^{-1} \mathbf{P}_1' \mathbf{P}_1 - Q_{m_1} \right\|_S = O_p((\log m_1)^{1/2} \xi_{0,m_1} n^{-1/2}) = o_p(1) \tag{SC.46}$$

[2]See (SC.270) in Subsection SC.5 for the form of $\varepsilon_{1,i}$ when $\beta_{l,0}$ is estimated by the partially linear regression proposed in Olley and Pakes (1996).

[3]For some approximating functions such as power series, Assumptions SC1(v, vi) hold under certain nonsingular transformation on the vector approximating functions $P_1(\cdot)$, i.e., $BP_1(\cdot)$, where B is some non-singular constant matrix. Since the nonparametric series estimator is invariant to any nonsingular transformation of $P_1(\cdot)$, we do not distinguish between $BP_1(\cdot)$ and $P_1(\cdot)$ throughout this appendix.

which together with Assumption SC1(v) implies that

$$C^{-1} \leq \lambda_{\min}(n^{-1}\mathbf{P}_1'\mathbf{P}_1) \leq \lambda_{\max}(n^{-1}\mathbf{P}_1'\mathbf{P}_1) \leq C \tag{SC.47}$$

uniformly over m_1 with probability approaching 1 (wpa1). Since $\hat{y}_{1,i} = y_{1,i} - l_{1,i}\hat{\beta}_l = \phi(x_{1,i}) + \eta_{1,i} - l_{1,i}(\hat{\beta}_l - \beta_{l,0})$, we can write

$$\begin{aligned} \hat{\beta}_\phi - \beta_{\phi,m_1} &= (\mathbf{P}_1'\mathbf{P}_1)^{-1}\sum_{i=1}^n P_1(x_{1,i})\eta_{1,i} \\ &\quad +(\mathbf{P}_1'\mathbf{P}_1)^{-1}\sum_{i=1}^n P_1(x_{1,i})(\phi(x_{1,i}) - \phi_{m_1}(x_{1,i})) \\ &\quad -(\hat{\beta}_l - \beta_{l,0})(\mathbf{P}_1'\mathbf{P}_1)^{-1}\sum_{i=1}^n P_1(x_{1,i})l_{1,i}. \end{aligned} \tag{SC.48}$$

By Assumption SC1(i, ii, v) and the Markov inequality

$$n^{-1}\sum_{i=1}^n P_1(x_{1,i})\eta_{1,i} = O_p(m_1^{1/2}n^{-1/2}) \tag{SC.49}$$

which together with Assumption SC1(vi), (SC.46) and (SC.47) implies that

$$\left[(n^{-1}\mathbf{P}_1'\mathbf{P}_1)^{-1} - Q_{m_1}^{-1}\right] n^{-1}\sum_{i=1}^n P_1(x_{1,i})\eta_{1,i} = O_p((\log m_1)^{1/2}\xi_{0,m_1}m_1^{1/2}n^{-1}) = o_p(n^{-1/2}). \tag{SC.50}$$

By Assumption SC1(iv, vi) and (SC.47)

$$(\mathbf{P}_1'\mathbf{P}_1)^{-1}\sum_{i=1}^n P_1(x_{1,i})(\phi(x_{1,i}) - \phi_{m_1}(x_{1,i})) = O_p(m^{-r_\phi}) = O_p(n^{-1/2}). \tag{SC.51}$$

Under Assumption SC1(i, ii, v, vi), we can use similar arguments in showing (SC.49) to get

$$n^{-1}\sum_{i=1}^n P_1(x_{1,i})l_{1,i} - \mathbb{E}[P_1(x_{1,i})l_{1,i}] = O_p(m_1^{1/2}n^{-1/2}) = o_p(1). \tag{SC.52}$$

By Assumption SC1(i, ii, v),

$$\|\mathbb{E}[l_{1,i}P_1(x_{1,i})]\|^2 \leq \lambda_{\max}(Q_{m_1})\mathbb{E}[l_{1,i}P_1(x_{1,i})']Q_{m_1}^{-1}\mathbb{E}[P_1(x_{1,i})l_{1,i}] \leq C\mathbb{E}[l_{1,i}^2] \leq C \tag{SC.53}$$

which combined with (SC.52) implies that

$$n^{-1}\sum_{i=1}^{n}P_1(x_{1,i})l_{1,i}=O_p(1). \quad \text{(SC.54)}$$

By Assumption SC1(iii, v, vi), (SC.46), (SC.52), (SC.53) and (SC.54),

$$(\hat{\beta}_l-\beta_{l,0})(\mathbf{P}_1'\mathbf{P}_1)^{-1}\sum_{i=1}^{n}P_1(x_{1,i})l_{1,i}=Q_{m_1}^{-1}\mathbb{E}[P_1(x_{1,i})l_{1,i}](\hat{\beta}_l-\beta_{l,0})+O_p(n^{-1/2})$$

which combined with Assumption SC1(vi), (SC.48), (SC.50) and (SC.51) shows that

$$\hat{\beta}_\phi-\beta_{\phi,m_1}=Q_{m_1}^{-1}\left(\sum_{i=1}^{n}P_1(x_{1,i})\eta_{1,i}-\mathbb{E}[P_1(x_{1,i})l_{1,i}](\hat{\beta}_l-\beta_{l,0})\right)+O_p(n^{-1/2})=O_p(m_1^{1/2}n^{-1/2}) \quad \text{(SC.55)}$$

where the second equality follows from Assumptions SC1(iii, v), (SC.49) and (SC.53). By the Cauchy-Schwarz inequality

$$\begin{aligned} n^{-1}\sum_{i=1}^{n}|\hat{\phi}(x_{1,i})-\phi(x_{1,i})|^2 &\leq 2n^{-1}\sum_{i=1}^{n}|\hat{\phi}(x_{1,i})-\phi_{m_1}(x_{1,i})|^2+2n^{-1}\sum_{i=1}^{n}|\phi_{m_1}(x_{1,i})-\phi(x_{1,i})|^2 \\ &\leq 2\lambda_{\max}(n^{-1}\mathbf{P}_1'\mathbf{P}_1)\left\|\hat{\beta}_\phi-\beta_{\phi,m_1}\right\|^2+2\sup_{x\in\mathcal{X}_1}|\phi_{m_1}(x)-\phi(x)|=O_p(m_1^{1/2}n^{-1/2}) \quad \text{(SC.56)}\end{aligned}$$

where the equality is by Assumptions SC1(iv, vi), (SC.47) and (SC.55), which proves (SC.44). By the triangle inequality, the Cauchy-Schwarz inequality, Assumption SC1(iv, vi) and (SC.55)

$$\begin{aligned}\sup_{x_1\in\mathcal{X}}|\hat{\phi}(x_1)-\phi(x_1)| &\leq \sup_{x_1\in\mathcal{X}}|\hat{\phi}(x_1)-\phi_{m_1}(x_1)|+\sup_{x_1\in\mathcal{X}}|\phi_{m_1}(x_1)-\phi(x_1)| \\ &\leq \xi_{0,m_1}\left\|\hat{\beta}_\phi-\beta_{\phi,m_1}\right\|+O(m_1^{-r_\phi})=O_p(\xi_{0,m_1}m_1^{1/2}n^{-1/2}) \quad \text{(SC.57)}\end{aligned}$$

which proves the claim in (SC.44). *Q.E.D.*

SC.2 Auxiliary results for the consistency of $\hat{\beta}_k$

Recall that $\nu_{1,i}(\beta_k)\equiv\phi(x_{1,i})-\beta_k k_{1,i}$ and $g(\nu;\beta_k)\equiv\mathbb{E}[y_{2,i}^*-\beta_k k_{2,i}|\nu_{1,i}(\beta_k)=\nu]$. For any $\beta_k\in\Theta_k$, let $\Omega(\beta_k)\equiv[a_{\beta_k},b_{\beta_k}]$ denote the support of $\nu_{1,i}(\beta_k)$ with $c_\nu<a_{\beta_k}<b_{\beta_k}<C_\nu$, where c_ν and C_ν are finite constants. Define $\Omega_\varepsilon(\beta_k)\equiv[a_{\beta_k}-\varepsilon,b_{\beta_k}+\varepsilon]$ for any constant $\varepsilon>0$. For an integer $d\geq 0$, let $|g(\beta_k)|_d=\max_{0\leq j\leq d}\sup_{\nu\in\Omega(\beta_k)}\left|\partial^j g(\nu;\beta_k)/\partial\nu^j\right|$.

Assumption SC2. *(i)* $\mathbb{E}[(y_{2,i}^*)^4+l_{2,i}^4+k_{2,i}^4|x_{1,i}]\leq C$; *(ii)* $g(\nu;\beta_k)$ *is twice continuously differentiable with uniformly bounded derivatives; (iii) for some* $d\geq 1$ *there exist* $\beta_{g,m_2}(\beta_k)\in\mathbb{R}^{m_2}$ *and* $r_g>0$ *such that* $\sup_{\beta_k\in\Theta_k}|g(\beta_k)-g_{m_2}(\beta_k)|_d=O(m_2^{-r_g})$ *where* $g_{m_2}(\nu;\beta_k)\equiv P_2(\nu)'\beta_{g,m_2}(\beta_k)$*;*

(iv) for any $\beta_k \in \Theta_k$ *there exists a nonsingular matrix* $B(\beta_k)$ *such that for* $\tilde{P}_2(\nu_1(\beta_k); \beta_k) \equiv B(\beta_k)P_2(\nu_1(\beta_k))$,

$$C^{-1} \leq \lambda_{\min}(Q_{m_2}(\beta_k)) \leq \lambda_{\max}(Q_{m_2}(\beta_k)) \leq C$$

uniformly over $\beta_k \in \Theta_k$*, where* $Q_{m_2}(\beta_k) \equiv \mathbb{E}[\tilde{P}_2(\nu_1(\beta_k); \beta_k)\, \tilde{P}_2(\nu_1(\beta_k); \beta_k)']$*; (v) for* $j = 0, 1, 2, 3$*, there exist sequences* ξ_{j,m_2} *such that* $\sup_{\beta_k \in \Theta_k} \sup_{\nu \in \Omega_\varepsilon(\beta_k)} \left\| \partial^j \tilde{P}_2(\nu; \beta_k) / \partial \nu^{j_1} \partial \beta_k^{j-j_1} \right\| \leq \xi_{j,m_2}$ *where* $j_1 \leq j$ *and* $\varepsilon = m_2^{-2}$*; (vi)* $\xi_{j,m_2} \leq C m_2^{j+1}$ *and* $\xi_{0,m_1}(m_1^{1/2} m_2^3 + (\log(n))^{1/2}) n^{-1/2} + n^{1/2} m_2^{-r_g} = o(1)$*.*

Assumption SC2(i) imposes upper bound on the conditional moments of $y_{2,i}^*$, $l_{2,i}$ and $k_{2,i}$ given $x_{1,i}$. Assumptions SC2(ii, iii) require that the conditional moment function $g(\nu; \beta_k)$ is smooth and can be well approximated by linear combinations of $P_2(\nu)$. Assumption SC2(iv) imposes normalization on the approximating functions $P_2(\nu)$, and uniform lower and upper bounds on the eigenvalues of $Q_{m_2}(\beta_k)$. Assumption SC2(v, vi) restrict the magnitudes of the normalized approximating functions and their derivatives, and the convergence rate of the series approximation error.

Let $\tilde{P}_2(\nu_{1,i}(\beta_{k,1}); \beta_{k,2}) \equiv B(\beta_{k,2})P_2(\nu_{1,i}(\beta_{k,1}))$ for any $\beta_{k,1}, \beta_{k,2} \in \Theta_k$. Since the series estimator $\hat{g}(\hat{\nu}_{1,i}(\beta_k); \beta_k) = P_2(\hat{\nu}_{1,i}(\beta_k))'\hat{\beta}_g(\beta_k)$ is invariant to any non-singular transformation on $P_2(\nu)$, throughout the rest of the Appendix we let

$$\tilde{\mathbf{P}}_2(\beta_k) \equiv (\tilde{P}_{2,1}(\beta_k), \ldots, \tilde{P}_{2,n}(\beta_k))' \quad \text{and} \quad \hat{\mathbf{P}}_2(\beta_k) \equiv (\hat{P}_{2,1}(\beta_k), \ldots, \hat{P}_{2,n}(\beta_k))'$$

where $\tilde{P}_{2,i}(\beta_k) \equiv \tilde{P}_2(\nu_{1,i}(\beta_k); \beta_k)$, $\hat{P}_{2,i}(\beta_k) \equiv B(\beta_k)P_2(\hat{\nu}_{1,i}(\beta_k))$ and $\hat{\nu}_{1,i}(\beta_k) \equiv \hat{\phi}(x_{1,i}) - k_{1,i}\beta_k$.[4] Define

$$\partial^j \tilde{P}_2(\nu; \beta_k) \equiv \frac{\partial^j \tilde{P}_2(\nu; \beta_k)}{\partial \nu^j} \quad \text{and} \quad \partial^j \tilde{P}_{2,i}(\beta_k) \equiv \partial^j \tilde{P}_2(\nu_{1,i}(\beta_k); \beta_k)$$

for $j = 1, 2, 3$ and $i = 1, \ldots, n$.

Lemma SC5. *Under Assumptions SC1 and SC2, we have*

$$\sup_{\beta_k \in \Theta_k} \left\| n^{-1}\hat{\mathbf{P}}_2(\beta_k)'\hat{\mathbf{P}}_2(\beta_k) - n^{-1}\tilde{\mathbf{P}}_2(\beta_k)'\tilde{\mathbf{P}}_2(\beta_k) \right\|_S = O_p(\xi_{1,m_2} m_1^{1/2} n^{-1/2}).$$

PROOF OF LEMMA SC5. Since $\hat{\nu}_{1,i}(\beta_k) = \hat{\phi}(x_{1,i}) - \beta_k k_{1,i}$, by Lemma SC4

$$\sup_{\beta_k \in \Theta_k} \max_{i \leq n} |\hat{\nu}_{1,i}(\beta_k) - \nu_{1,i}(\beta_k)| = \max_{i \leq n} |\hat{\phi}(x_{1,i}) - \phi(x_{1,i})| = O_p(\xi_{0,m_1} m_1^{1/2} n^{-1/2}) = o_p(1) \quad \text{(SC.58)}$$

[4] Note that we define $\hat{P}_{2,i}(\beta_k) \equiv P_2(\hat{\omega}_{1,i}(\beta_k))$ in Section SA. We change its definition here since the asymptotic properties of the sereis estimator $\hat{g}(\hat{\omega}_{1,i}(\beta_k); \beta_k) = P_2(\hat{\omega}_{1,i}(\beta_k))'\hat{\beta}_g(\beta_k)$ shall be investigated under the new definition $\hat{P}_{2,i}(\beta_k) \equiv B(\beta_k)P_2(\hat{\omega}_{1,i}(\beta_k))$.

which together with Assumption SC2(vi) implies that

$$\hat{\nu}_{1,i}(\beta_k) \in \Omega_\varepsilon(\beta_k) \text{ wpa1} \quad \text{(SC.59)}$$

for any $i \leq n$ and uniformly over $\beta_k \in \Theta_k$. By the mean value expansion, we have for any $b \in \mathbb{R}^{m_2}$

$$\left| b'(\tilde{P}_{2,i}(\beta_k) - \hat{P}_{2,i}(\beta_k)) \right| = \left| b'\partial^1 \tilde{P}_2\left(\tilde{\nu}_{1,i}(\beta_k);\beta_k\right)\left(\hat{\nu}_{1,i}(\beta_k) - \nu_{1,i}(\beta_k)\right) \right| \quad \text{(SC.60)}$$

where $\tilde{\nu}_{1,i}(\beta_k)$ lies between $\nu_{1,i}(\beta_k)$ and $\hat{\nu}_{1,i}(\beta_k)$. Since $\nu_{1,i}(\beta_k)$ and $\hat{\nu}_{1,i}(\beta_k)$ are in $\Omega_\varepsilon(\beta_k)$ uniformly over $\beta_k \in \Theta_k$ and for any $i = 1, \ldots, n$ wpa1, the same property holds for $\tilde{\nu}_{1,i}(\beta_k)$. By the Cauchy-Schwarz inequality, Assumption SC2(v) and (SC.60)

$$\left| b'(\tilde{P}_{2,i}(\beta_k) - \hat{P}_{2,i}(\beta_k)) \right| \leq \|b\| \, \xi_{1,m_2} |\hat{\phi}(x_{1,i}) - \phi(x_{1,i})| \text{ wpa1.}$$

Therefore,

$$\begin{aligned} & b'(\hat{\mathbf{P}}_2(\beta_k) - \tilde{\mathbf{P}}_2(\beta_k))'(\hat{\mathbf{P}}_2(\beta_k) - \tilde{\mathbf{P}}_2(\beta_k))b \\ &= \sum_{i=1}^{n} (b'(\tilde{P}_{2,i}(\beta_k) - \hat{P}_{2,i}(\beta_k)))^2 \leq \|b\|^2 \, \xi_{1,m_2}^2 \sum_{i=1}^{n} |\hat{\phi}(x_{1,i}) - \phi(x_{1,i})|^2 \end{aligned}$$

wpa1, which together with Lemma SC4 implies that

$$\sup_{\beta_k \in \Theta_k} ||\hat{\mathbf{P}}_2(\beta_k) - \tilde{\mathbf{P}}_2(\beta_k)||_S = O_p(\xi_{1,m_2} m_1^{1/2}). \quad \text{(SC.61)}$$

By Lemma SC32 and Assumption SC2(iv, vi), we have uniformly over $\beta_k \in \Theta_k$

$$C^{-1} \leq \lambda_{\min}(n^{-1}\tilde{\mathbf{P}}_2(\beta_k)'\tilde{\mathbf{P}}_2(\beta_k)) \leq \lambda_{\max}(n^{-1}\tilde{\mathbf{P}}_2(\beta_k)'\tilde{\mathbf{P}}_2(\beta_k)) \leq C \text{ wpa1.} \quad \text{(SC.62)}$$

By the triangle inequality, Assumption SC2(vi), (SC.61) and (SC.62), we get

$$\begin{aligned} & \sup_{\beta_k \in \Theta_k} \left\| n^{-1}\hat{\mathbf{P}}_2(\beta_k)'\hat{\mathbf{P}}_2(\beta_k) - n^{-1}\tilde{\mathbf{P}}_2(\beta_k)'\tilde{\mathbf{P}}_2(\beta_k) \right\|_S \\ \leq & \sup_{\beta_k \in \Theta_k} n^{-1} \left\| (\hat{\mathbf{P}}_2(\beta_k) - \tilde{\mathbf{P}}_2(\beta_k))'(\hat{\mathbf{P}}_2(\beta_k) - \tilde{\mathbf{P}}_2(\beta_k)) \right\|_S \\ & + \sup_{\beta_k \in \Theta_k} n^{-1} \left\| (\hat{\mathbf{P}}_2(\beta_k) - \tilde{\mathbf{P}}_2(\beta_k))'\tilde{\mathbf{P}}_2(\beta_k) \right\|_S \\ & + \sup_{\beta_k \in \Theta_k} n^{-1} \left\| \tilde{\mathbf{P}}_2(\beta_k)'(\hat{\mathbf{P}}_2(\beta_k) - \tilde{\mathbf{P}}_2(\beta_k)) \right\|_S = O_p(\xi_{1,m_2} m_1^{1/2} n^{-1/2}) \end{aligned}$$

which proves the claim of the lemma. *Q.E.D.*

Lemma SC6. *Under Assumptions SC1 and SC2, we have*

$$\sup_{\beta_k \in \Theta_k} n^{-1} \sum_{i=1}^{n} \left| \tilde{P}_{2,i}(\beta_k)' \hat{\beta}_g(\beta_k) - g(\nu_{1,i}(\beta_k); \beta_k) \right|^2 = O_p((m_2^{5/2} + \xi_{1,m_2}^2 m_1)n^{-1}) = o_p(1)$$

where $\hat{\beta}_g(\beta_k) \equiv (\hat{\mathbf{P}}_2(\beta_k)'\hat{\mathbf{P}}_2(\beta_k))^{-1}\hat{\mathbf{P}}_2(\beta_k)'\hat{\mathbf{Y}}_2^*(\beta_k)$.

Proof of Lemma SC6. Let $\tilde{\beta}_{g,m_2}(\beta_k) \equiv (B(\beta_k)')^{-1}\beta_{g,m_2}(\beta_k)$ and $\beta_{g,m_2}(\beta_k)$ is defined in Assumption SC2(iii). By the Cauchy-Schwarz inequality and Assumption SC2(iii)

$$\begin{aligned}
& n^{-1} \sum_{i=1}^{n} \left| \tilde{P}_{2,i}(\beta_k)' \hat{\beta}_g(\beta_k) - g(\nu_{1,i}(\beta_k); \beta_k) \right|^2 \\
\leq\ & 2n^{-1} \sum_{i=1}^{n} \left| \tilde{P}_{2,i}(\beta_k)' \hat{\beta}_g(\beta_k) - g_{m_2}(\nu_{1,i}(\beta_k); \beta_k) \right|^2 \\
& +2n^{-1} \sum_{i=1}^{n} |g_{m_2}(\nu_{1,i}(\beta_k); \beta_k) - g(\nu_{1,i}(\beta_k); \beta_k)|^2 \\
\leq\ & 2\lambda_{\max}(n^{-1}\tilde{\mathbf{P}}_2(\beta_k)'\tilde{\mathbf{P}}_2(\beta_k))||\hat{\beta}_g(\beta_k) - \tilde{\beta}_{g,m_2}(\beta_k)||^2 + Cm_2^{-2r_g} \qquad \text{(SC.63)}
\end{aligned}$$

uniformly over $\beta_k \in \Theta_k$, where $g_{m_2}(\nu_{1,i}(\beta_k); \beta_k) \equiv \tilde{P}_{2,i}(\beta_k)'\tilde{\beta}_{g,m_2}(\beta_k)$ for any $\beta_k \in \Theta_k$. We next show that

$$\sup_{\beta_k \in \Theta_k} \left\| \hat{\beta}_g(\beta_k) - \tilde{\beta}_{g,m_2}(\beta_k) \right\|^2 = O_p((m_2^{5/2} + \xi_{1,m_2}^2 m_1)n^{-1}) = o_p(1) \qquad \text{(SC.64)}$$

which together with (SC.62) and (SC.63) proves the claim of the lemma.

Let $u_{2,i}(\beta_k) \equiv y_{2,i}^* - k_{2,i}\beta_k - g(\nu_{1,i}(\beta_k), \beta_k)$. Then we can write

$$\begin{aligned}
\hat{\beta}_g(\beta_k) - \tilde{\beta}_{g,m_2}(\beta_k) &= (\hat{\mathbf{P}}_2(\beta_k)'\hat{\mathbf{P}}_2(\beta_k))^{-1}\hat{\mathbf{P}}_2(\beta_k)'(\hat{\mathbf{Y}}_2^*(\beta_k) - \hat{\mathbf{P}}_2(\beta_k)'\tilde{\beta}_{g,m_2}(\beta_k)) \\
&= (\hat{\mathbf{P}}_2(\beta_k)'\hat{\mathbf{P}}_2(\beta_k))^{-1} \sum_{i=1}^{n} \hat{P}_{2,i}(\beta_k)(g(\nu_{1,i}(\beta_k), \beta_k) - g_{m_2}(\hat{\nu}_{1,i}(\beta_k), \beta_k)) \\
&\quad -(\hat{\beta}_l - \beta_{l,0})(\hat{\mathbf{P}}_2(\beta_k)'\hat{\mathbf{P}}_2(\beta_k))^{-1} \sum_{i=1}^{n} \hat{P}_{2,i}(\beta_k) l_{2,i} \\
&\quad +(\hat{\mathbf{P}}_2(\beta_k)'\hat{\mathbf{P}}_2(\beta_k))^{-1} \sum_{i=1}^{n} \hat{P}_{2,i}(\beta_k) u_{2,i}(\beta_k) \qquad \text{(SC.65)}
\end{aligned}$$

where $g_{m_2}(\hat{\nu}_{1,i}(\beta_k), \beta_k) \equiv \hat{P}_{2,i}(\beta_k)'\tilde{\beta}_{g,m_2}(\beta_k)$. By Assumption SC2(vi), Lemma SC5 and (SC.62), we have uniformly over $\beta_k \in \Theta_k$

$$C^{-1} \leq \lambda_{\min}(n^{-1}\hat{\mathbf{P}}_2(\beta_k)'\hat{\mathbf{P}}_2(\beta_k)) \leq \lambda_{\max}(n^{-1}\hat{\mathbf{P}}_2(\beta_k)'\hat{\mathbf{P}}_2(\beta_k)) \leq C \text{ wpa1} \qquad \text{(SC.66)}$$

which implies that $\hat{\mathbf{P}}_2(\beta_k)(\hat{\mathbf{P}}_2(\beta_k)'\hat{\mathbf{P}}_2(\beta_k))^{-1}\hat{\mathbf{P}}_2(\beta_k)'$ is an idempotent matrix uniformly over $\beta_k \in$

Θ_k wpa1. Therefore,

$$\begin{aligned} & \left\| (\hat{\mathbf{P}}_2(\beta_k)'\hat{\mathbf{P}}_2(\beta_k))^{-1} \sum_{i=1}^n \hat{P}_{2,i}(\beta_k)(g(\nu_{1,i}(\beta_k),\beta_k) - g_{m_2}(\hat{\nu}_{1,i}(\beta_k),\beta_k)) \right\|^2 \\ \leq \quad & O_p(1)n^{-1}\sum_{i=1}^n (g(\nu_{1,i}(\beta_k),\beta_k) - g_{m_2}(\hat{\nu}_{1,i}(\beta_k),\beta_k))^2. \end{aligned} \tag{SC.67}$$

uniformly over $\beta_k \in \Theta_k$. Since $\nu_{1,i}(\beta_k) = \phi(x_{1,i}) - k_{1,i}\beta_k$, we can use Assumptions SC1(i) and SC2(i) to deduce

$$\sup_{\beta_k \in \Theta_k} |g(\nu_{1,i}(\beta_k);\beta_k)| \leq C. \tag{SC.68}$$

Therefore,

$$\begin{aligned} \sup_{\beta_k \in \Theta_k} \left\| \tilde{\beta}_{g,m_2}(\beta_k) \right\|^2 & \leq \sup_{\beta_k \in \Theta_k} (\lambda_{\min}(Q_{m_2}(\beta_k)))^{-1} \left\| \tilde{P}_{2,i}(\beta_k)'\tilde{\beta}_{g,m_2}(\beta_k) \right\|_2^2 \\ & \leq C \sup_{\beta_k \in \Theta_k} \|g(\nu_{1,i}(\beta_k);\beta_k) - g_{m_2}(\nu_{1,i}(\beta_k);\beta_k)\|_2^2 \\ & \quad + C \sup_{\beta_k \in \Theta_k} \|g(\nu_{1,i}(\beta_k);\beta_k)\|_2^2 \leq C. \end{aligned} \tag{SC.69}$$

By the second order expansion, Assumption SC2(iii, v, vi), Lemma SC4, (SC.68) and (SC.69), we have uniformly over $\beta_k \in \Theta_k$,

$$\begin{aligned} & n^{-1}\sum_{i=1}^n (g_{m_2}(\nu_{1,i}(\beta_k),\beta_k) - g_{m_2}(\hat{\nu}_{1,i}(\beta_k),\beta_k))^2 \\ \leq \quad & 2n^{-1}\sum_{i=1}^n (\partial^1 \tilde{P}_{2,i}(\beta_k)'\tilde{\beta}_{g,m_2}(\beta_k)(\hat{\phi}(x_{1,i}) - \phi(x_{1,i}))^2 \\ & + 2n^{-1}\sum_{i=1}^n (\partial^2 \tilde{P}_2\left(\tilde{\nu}_{1,i}(\beta_k);\beta_k\right)'\tilde{\beta}_{g,m_2}(\beta_k)(\hat{\phi}(x_{1,i}) - \phi(x_{1,i})^2)^2 \\ = \quad & O_p(m_1 n^{-1}) + O_p(\xi_{2,m_2}^2 \xi_{0,m_1}^2 m_1^2 n^{-2}) = O_p(m_1 n^{-1}) \end{aligned}$$

where $\tilde{\nu}_{1,i}(\beta_k)$ is between $\nu_{1,i}(\beta_k)$ and $\hat{\nu}_{1,i}(\beta_k)$ and it lies in $\Omega_\varepsilon(\beta_k)$ uniformly over $\beta_k \in \Theta_k$ wpa1 by (SC.59), which together with Assumption SC2(iii, vi) implies that

$$\begin{aligned} & n^{-1}\sum_{i=1}^{n}(g(\nu_{1,i}(\beta_k),\beta_k) - g_{m_2}(\hat{\nu}_{1,i}(\beta_k),\beta_k))^2 \\ \leq \; & Cn^{-1}\sum_{i=1}^{n}(g(\nu_{1,i}(\beta_k),\beta_k) - g_{m_2}(\nu_{1,i}(\beta_k),\beta_k))^2 \\ & +Cn^{-1}\sum_{i=1}^{n}(g_{m_2}(\nu_{1,i}(\beta_k),\beta_k) - g_{m_2}(\hat{\nu}_{1,i}(\beta_k),\beta_k))^2 \\ = \; & O_p(m_1 n^{-1} + m_2^{-2r_g}) = O_p(m_1 n^{-1}). \end{aligned} \tag{SC.70}$$

From (SC.67) and (SC.70), we get uniformly over $\beta_k \in \Theta_k$

$$(\hat{\mathbf{P}}_2(\beta_k)'\hat{\mathbf{P}}_2(\beta_k))^{-1}\sum_{i=1}^{n}\hat{P}_{2,i}(\beta_k)(g(\nu_{1,i}(\beta_k),\beta_k) - g_{m_2}(\hat{\nu}_{1,i}(\beta_k),\beta_k)) = O_p(m_1^{1/2}n^{-1/2}). \tag{SC.71}$$

By Assumptions SC1(i) and SC2(i), and the Markov inequality,

$$n^{-1}\sum_{i=1}^{n} l_{2,i}^2 = O_p(1) \tag{SC.72}$$

which together with Assumption SC1(iii) and (SC.66) implies that

$$(\hat{\beta}_l - \beta_{l,0})(\hat{\mathbf{P}}_2(\beta_k)'\hat{\mathbf{P}}_2(\beta_k))^{-1}\sum_{i=1}^{n}\hat{P}_{2,i}(\beta_k) l_{2,i} = O_p(n^{-1/2}) \tag{SC.73}$$

uniformly over $\beta_k \in \Theta_k$. By the mean value expansion, the Cauchy-Schwarz inequality and the triangle inequality, we have for any $b \in \mathbb{R}^{m_2}$

$$\begin{aligned} & \left| n^{-1}\sum_{i=1}^{n} b'(\hat{P}_{2,i}(\beta_k) - \tilde{P}_{2,i}(\beta_k))u_{2,i}(\beta_k) \right| \\ = \; & \left| n^{-1}\sum_{i=1}^{n} b'\partial^1\tilde{P}_2\left(\tilde{\nu}_{1,i}(\beta_k);\beta_k\right)\left(\hat{\nu}_{1,i}(\beta_k) - \nu_{1,i}(\beta_k)\right)u_{2,i}(\beta_k) \right| \\ \leq \; & \|b\|\,\xi_{1,m_2} n^{-1}\sum_{i=1}^{n}\left|(\hat{\phi}(x_{1,i}) - \phi(x_{1,i}))u_{2,i}(\beta_k)\right|. \end{aligned} \tag{SC.74}$$

By the definition of $u_{2,i}(\beta_k)$, we can use Assumptions SC1(i) and SC2(i), (SC.68) and the Markov inequality to deduce

$$\sup_{\beta_k \in \Theta_k} n^{-1}\sum_{i=1}^{n}(u_{2,i}(\beta_k))^2 = O_p(1). \tag{SC.75}$$

Thus by the Cauchy-Schwarz inequality, Lemma SC4 and (SC.75),

$$\sup_{\beta_k\in\Theta_k} n^{-1}\sum_{i=1}^{n}\left|(\hat{\phi}(x_{1,i})-\phi(x_{1,i}))u_{2,i}(\beta_k)\right| = O_p(m_1^{1/2}n^{-1/2})$$

which together with (SC.66) and (SC.74) implies that

$$(\hat{\mathbf{P}}_2(\beta_k)'\hat{\mathbf{P}}_2(\beta_k))^{-1}\sum_{i=1}^{n}(\hat{P}_{2,i}(\beta_k)-\tilde{P}_{2,i}(\beta_k))u_{2,i}(\beta_k) = O_p(\xi_{1,m_2}m_1^{1/2}n^{-1/2}) \quad \text{(SC.76)}$$

uniformly over $\beta_k\in\Theta_k$. Applying Lemma SC33 and (SC.66) yields

$$(\hat{\mathbf{P}}_2(\beta_k)'\hat{\mathbf{P}}_2(\beta_k))^{-1}\sum_{i=1}^{n}\tilde{P}_{2,i}(\beta_k)u_{2,i}(\beta_k) = O_p(m_2^{5/4}n^{-1/2}) \quad \text{(SC.77)}$$

uniformly over $\beta_k\in\Theta_k$. The claim in (SC.64) then follows from Assumption SC2(vi), (SC.65), (SC.71), (SC.73), (SC.76) and (SC.77). *Q.E.D.*

Lemma SC7. *Under Assumptions SC1 and SC2, we have*

$$\sup_{\beta_k\in\Theta_k} n^{-1}\sum_{i=1}^{n}|\hat{g}(\hat{\nu}_{1,i}(\beta_k);\beta_k)-g(\nu_{1,i}(\beta_k);\beta_k)|^2 = O_p((m_2^{5/2}+\xi_{1,m_2}^2 m_1)n^{-1}) = o_p(1).$$

Proof of Lemma SC7. By the triangle inequality, (SC.64) and (SC.69)

$$\sup_{\beta_k\in\Theta_k}\left\|\hat{\beta}_g(\beta_k)\right\| \le \sup_{\beta_k\in\Theta_k}\left\|\tilde{\beta}_{g,m_2}(\beta_k)\right\| + \sup_{\beta_k\in\Theta_k}\left\|\hat{\beta}_g(\beta_k)-\tilde{\beta}_{g,m_2}(\beta_k)\right\| = O_p(1). \quad \text{(SC.78)}$$

By the mean value expansion, the Cauchy-Schwarz inequality, Assumption SC2(v, vi), Lemma SC4 and (SC.78),

$$\begin{aligned}
&\sup_{\beta_k\in\Theta_k} n^{-1}\sum_{i=1}^{n}\left|(\hat{P}_{2,i}(\beta_k)-\tilde{P}_{2,i}(\beta_k))'\hat{\beta}_g(\beta_k)\right|^2 \\
=\;& \sup_{\beta_k\in\Theta_k} n^{-1}\sum_{i=1}^{n}\left|\partial^1\tilde{P}_2\left(\tilde{\nu}_{1,i}(\beta_k);\beta_k\right)'\hat{\beta}_g(\beta_k)(\hat{\nu}_{1,i}(\beta_k)-\nu_{1,i}(\beta_k)\right|^2 \\
\le\;& \xi_{1,m_2}^2 n^{-1}\sum_{i=1}^{n}(\hat{\phi}(x_{1,i})-\phi(x_{1,i}))^2\sup_{\beta_k\in\Theta_k}\left\|\hat{\beta}_g(\beta_k)\right\| = O_p(\xi_{1,m_2}^2 m_1 n^{-1}) = o_p(1) \quad \text{(SC.79)}
\end{aligned}$$

where $\tilde{\nu}_{1,i}(\beta_k)$ is between $\hat{\nu}_{1,i}(\beta_k)$ and $\nu_{1,i}(\beta_k)$ and hence by (SC.59) it lies in $\Omega_\varepsilon(\beta_k)$ wpa1 for any $i\le n$ and uniformly over $\beta_k\in\Theta_k$. The claim of the lemma directly follows from Lemma SC6 and (SC.79). *Q.E.D.*

Lemma SC8. *Under Assumptions SC1 and SC2, we have*

$$\sup_{\beta_k \in \Theta_k} n^{-1} \sum_{i=1}^{n} (\ell_i(\beta_k)^2 - \mathbb{E}\left[\ell_i(\beta_k)^2\right]) = O_p(n^{-1/2}).$$

Proof of Lemma SC8. For any $\beta_k \in \Theta_k$, by the Cauchy-Schwarz inequality and (SC.68),

$$\ell_i(\beta_k)^2 \leq C\left[(y_{2,i}^*)^2 + k_{2,i}^2\beta_k^2 + g(\nu_{1,i}(\beta_k);\beta_k)^2\right] \leq C(1 + (y_{2,i}^*)^2 + k_{2,i}^2). \tag{SC.80}$$

For any $\beta_{k,1}$, $\beta_{k,2} \in \Theta_k$, by the triangle inequality and Assumption SC2(ii),

$$|\ell_i(\beta_{k,1}) - \ell_i(\beta_{k,2})| \leq (C + k_{2,i}) |\beta_{k,1} - \beta_{k,2}|. \tag{SC.81}$$

By Assumption SC2(ii), (SC.80) and (SC.81), we get

$$\mathbb{E}\left[\ell_i(\beta_k)^2 |\ell_i(\beta_{k,1}) - \ell_i(\beta_{k,2})|^2\right] \leq C(\beta_{k,1} - \beta_{k,2})^2 \tag{SC.82}$$

for any $\beta_k \in \Theta_k$, which implies that

$$\mathbb{E}\left[\left|\ell_i(\beta_{k,1})^2 - \ell_i(\beta_{k,2})^2\right|^2\right] \leq C(\beta_{k,1} - \beta_{k,2})^2.$$

Therefore we have for any $\beta_{k,1}, \beta_{k,2} \in \Theta_k$,

$$\left\|\ell_i(\beta_{k,1})^2 - \ell_i(\beta_{k,2})^2\right\|_2 \leq C |\beta_{k,1} - \beta_{k,2}|. \tag{SC.83}$$

By Assumptions SC1(i) and SC2(i), and (SC.68),

$$\begin{aligned}
&\mathbb{E}\left[\left|n^{-1/2} \sum_{i=1}^{n} \left(\ell_i(\beta_k)^2 - \mathbb{E}\left[\ell_i(\beta_k)^2\right]\right)\right|^2\right] \\
&= \mathbb{E}\left[\ell_i(\beta_k)^4\right] - \left(\mathbb{E}\left[\ell_i(\beta_k)^2\right]\right)^2 \leq C\left(\mathbb{E}\left[(y_{2,i}^*)^4 + k_{2,i}^4 + (g(\nu_{1,i}(\beta_k);\beta_k))^4\right]\right) \leq C
\end{aligned}$$

for any $\beta_k \in \Theta_k$, which implies that

$$n^{-1/2} \sum_{i=1}^{n} \left(\ell_i(\beta_k)^2 - \mathbb{E}\left[\ell_i(\beta_k)^2\right]\right) = O_p(1) \tag{SC.84}$$

for any $\beta_k \in \Theta_k$. Moreover, by Assumption SC1(i) and (SC.83)

$$\mathbb{E}\left[\left|n^{-1/2}\sum_{i=1}^{n}\left(\ell_i(\beta_{k,1})^2-\ell_i(\beta_{k,2})^2-\mathbb{E}\left[\ell_i(\beta_{k,1})^2-\ell_i(\beta_{k,1})^2\right]\right)\right|^2\right]$$
$$\leq \mathbb{E}\left[\left|\ell_i(\beta_{k,1})^2-\ell_i(\beta_{k,2})^2\right|^2\right] \leq C\left|\beta_{k,1}-\beta_{k,2}\right|^2. \quad \text{(SC.85)}$$

Collecting the results in (SC.84) and (SC.85), we can invoke Theorem 2.2.4 in van der Vaart and Wellner (1996) to deduce that

$$\left\|\sup_{\beta_k\in\Theta_k}\left|n^{-1/2}\sum_{i=1}^{n}\left(\ell_i(\beta_k)^2-\mathbb{E}\left[\ell_i(\beta_k)^2\right]\right)\right|\right\|_2 \leq C$$

which together with the Markov inequality finishes the proof. *Q.E.D.*

SC.3 Auxiliary results for the asymptotic normality of $\hat{\beta}_k$

Let $\varphi(\nu) \equiv \gamma_2(\nu) - \gamma_1(\nu)g_1(\nu)$ where $g_1(\nu) \equiv \partial g(\nu)/\partial\nu$ and $\gamma_j(\nu) \equiv \mathbb{E}[k_{j,i}|\nu_{1,i} = \nu]$ for $j = 1, 2$. For any $\beta_k \in \Theta_k$ and $i = 1, \ldots, n$, let

$$\hat{g}_i(\beta_k) \equiv \hat{g}(\hat{\nu}_{1,i}(\beta_k); \beta_k) \quad \text{and} \quad \hat{g}_{1,i}(\beta_k) \equiv \hat{g}_1(\hat{\nu}_{1,i}(\beta_k); \beta_k).$$

The following assumptions are needed for showing the asymptotic normality of $\hat{\beta}_k$.

Assumption SC3. *(i) $\varphi(\nu)$ is continuously differentiable with uniformly bounded derivatives over $\nu \in \Omega(\beta_{k,0})$; (ii) there exist $\beta_{\varphi,m_2} \in \mathbb{R}^{m_2}$ and $r_\varphi > 0$ such that*

$$\sup_{\nu\in\Omega(\beta_{k,0})}\left|\varphi(\nu)-\varphi_{m_2}(\nu)\right| = O(m_2^{-r_\varphi})$$

where $\varphi_{m_2}(\nu) \equiv P_2(\nu)'\beta_{\varphi,m_2}$; (iii) there exists $\beta_{\varsigma,m_1} \in \mathbb{R}^{m_2}$ such that

$$\left\|(\varsigma_{1,i}-\varsigma_{2,i})g_1(\nu_{1,i})-P_1(x_{1,i})'\beta_{\varsigma,m_1}\right\|_2 \to 0 \text{ as } m_1 \to \infty;$$

(iv) $n^{1/2}m_2^{-r_\varphi} + m_1 m_2^4 n^{-1/2} = o(1)$.

Assumptions SC3(i, ii) require that the function $\varphi(\nu)$ is smooth and can be well approximated by the approximating functions $P_2(\nu)$. By the definition of $\varsigma_{1,i}$ and $\varsigma_{2,i}$, we can write

$$\varsigma_{1,i} - \varsigma_{2,i} = \mathbb{E}[k_{2,i}|x_{1,i}] - \mathbb{E}[k_{2,i}|\nu_{1,i}] - (k_{1,i} - \mathbb{E}[k_{1,i}|\nu_{1,i}])\, g_1(\nu_{1,i})$$

which combined with Assumption SC2(i, ii) implies that $(\varsigma_{1,i}-\varsigma_{2,i})g_1(\nu_{1,i})$ is a function of $x_{1,i}$ with finite L_2-norm. Assumption SC3(iii) requires that $(\varsigma_{1,i}-\varsigma_{2,i})g_1(\nu_{1,i})$ can be approximated by the approximating functions $P_1(x_{1,i})$. Assumption SC3(iv) restricts the numbers of the approximating functions and the smoothness of $\varphi(\nu)$.

Lemma SC9. *Under Assumptions SC1, SC2 and SC3(iv), we have*

$$\left\|\hat{\beta}_g(\beta_{k,0}) - \tilde{\beta}_{g,m_2}(\beta_{k,0})\right\| = O_p((m_1^{1/2} + m_2^{1/2})n^{-1/2})$$

where $\tilde{\beta}_{g,m_2}(\beta_{k,0}) \equiv (B(\beta_{k,0})')^{-1}\beta_{g,m_2}(\beta_{k,0})$ *and* $\beta_{g,m_2}(\beta_{k,0})$ *is defined in Assumption SC2(iii).*

Proof of Lemma SC9. By the definition of $\hat{\beta}_g(\beta_k)$, we can utilize the decomposition in (SC.65), and the results in (SC.71) and (SC.73) to get

$$\hat{\beta}_g(\beta_{k,0}) - \tilde{\beta}_{g,m_2}(\beta_{k,0}) = (\hat{\mathbf{P}}_2(\beta_{k,0})'\hat{\mathbf{P}}_2(\beta_{k,0}))^{-1}\sum_{i=1}^{n}\hat{P}_{2,i}(\beta_{k,0})u_{2,i} + O_p(m_1^{1/2}n^{-1/2}). \qquad \text{(SC.86)}$$

By the second order expansion, we have for any $b \in \mathbb{R}^{m_2}$

$$\begin{aligned} n^{-1}\sum_{i=1}^{n} b'(\hat{P}_{2,i}(\beta_{k,0}) - \tilde{P}_{2,i}(\beta_{k,0}))u_{2,i} &= n^{-1}\sum_{i=1}^{n} b'\partial^1\tilde{P}_{2,i}(\beta_{k,0})(\hat{\phi}_i - \phi_i)u_{2,i} \qquad \text{(SC.87)} \\ &\quad + n^{-1}\sum_{i=1}^{n} b'\partial^2\tilde{P}_2(\tilde{\nu}_{1,i};\beta_{k,0})(\hat{\phi}_i - \phi_i)^2 u_{2,i} \end{aligned}$$

where $\tilde{\nu}_{1,i}$ is between $\hat{\nu}_{1,i}(\beta_{k,0})$ and $\nu_{1,i}(\beta_{k,0})$. By (SC.59), $\tilde{\nu}_{1,i} \in \Omega_\varepsilon(\beta_{k,0})$ for any $i = 1,\ldots,n$ wpa1. By Assumption SC2(i) and (SC.68),

$$\mathbb{E}\left[u_{2,i}^4 | x_{1,i}\right] \leq C. \qquad \text{(SC.88)}$$

By Assumption SC1(i, v, vi), (SC.88) and the Markov inequality

$$\left\|n^{-1}\sum_{i=1}^{n}|u_{2,i}|\,P_1(x_{1,i})P_1(x_{1,i})' - \mathbb{E}\left[|u_{2,i}|\,P_1(x_{1,i})P_1(x_{1,i})'\right]\right\| = o_p(1). \qquad \text{(SC.89)}$$

Since $\lambda_{\max}(\mathbb{E}[|u_{2,i}|\,P_1(x_{1,i})P_1(x_{1,i})']) \leq C$ by Assumption SC1(v) and (SC.88), from (SC.89) we deduce that

$$\lambda_{\max}\left(n^{-1}\sum_{i=1}^{n}|u_{2,i}|\,P_1(x_{1,i})P_1(x_{1,i})'\right) \leq C \text{ wpa1}. \qquad \text{(SC.90)}$$

By (SC.55) and (SC.90), we get

$$n^{-1}\sum_{i=1}^{n}\left|u_{2,i}(\hat{\phi}_i - \phi_{m_1,i})^2\right| = O_p(m_1 n^{-1}) \qquad \text{(SC.91)}$$

where $\phi_{m_1,i} \equiv \phi_{m_1}(x_{1,i})$. By Assumption SC1(i, iv) and (SC.88), and the Markov inequality

$$n^{-1}\sum_{i=1}^{n}\left|u_{2,i}(\phi_{m_1,i}-\phi_i)^2\right| = O_p(m_1^{-2r_\phi})$$

which together with (SC.91) and Assumption SC1(vi) implies that

$$n^{-1}\sum_{i=1}^{n}|u_{2,i}|\,(\hat{\phi}_i-\phi_i)^2 = O_p(m_1 n^{-1}). \tag{SC.92}$$

By the Cauchy-Schwarz inequality and the triangle inequality, Assumption SC2(v) and (SC.92)

$$\left|n^{-1}\sum_{i=1}^{n} b'\partial^2\tilde{P}_2\,(\tilde{\nu}_{1,i};\beta_{k,0})\,(\hat{\phi}_i-\phi_i)^2 u_{2,i}\right| \leq \|b\|\,O_p(\xi_{2,m_2}m_1 n^{-1}). \tag{SC.93}$$

By Assumptions SC1(i, v) and SC2(v), and (SC.88),

$$\mathbb{E}\left[\left\|n^{-1}\sum_{i=1}^{n} u_{2,i}\partial^1\tilde{P}_{2,i}\,(\beta_{k,0})\,P_1(x_{1,i})'\right\|^2\right] \leq C\xi_{1,m_2}^2 m_1 n^{-1}$$

which together with the Cauchy-Schwarz inequality, the Markov inequality and (SC.55) implies that

$$\begin{aligned}&\left|n^{-1}\sum_{i=1}^{n} u_{2,i}b'\partial^1\tilde{P}_{2,i}\,(\beta_{k,0})\,P_1(x_{1,i})'(\hat{\beta}_\phi-\beta_{\phi,m_1})\right| \qquad \text{(SC.94)}\\ \leq\;& \|b\|\left\|\hat{\beta}_\phi-\beta_{\phi,m_1}\right\|\left\|n^{-1}\sum_{i=1}^{n} u_{2,i}\partial^1\tilde{P}_{2,i}\,(\beta_{k,0})\,P_1(x_{1,i})'\right\| = \|b\|\,O_p(\xi_{1,m_2}m_1 n^{-1}).\end{aligned}$$

By Assumptions SC1(i, iv, vi) and SC2(v), and (SC.88),

$$\mathbb{E}\left[\left\|n^{-1}\sum_{i=1}^{n}\partial^1\tilde{P}_{2,i}\,(\beta_{k,0})\,(\phi_{m_2,i}-\phi_i)u_{2,i}\right\|^2\right] \leq C\xi_{1,m_2}^2 n^{-2}$$

which together with the Cauchy-Schwarz inequality and the Markov inequality implies that

$$\left|n^{-1}\sum_{i=1}^{n} b'\partial^1\tilde{P}_{2,i}\,(\beta_{k,0})\,(\phi_{m_2,i}-\phi_i)u_{2,i}\right| \leq \|b\|\,O_p(\xi_{1,m_2}n^{-1}). \tag{SC.95}$$

Collecting the results in (SC.94) and (SC.95) obtains

$$\left| n^{-1} \sum_{i=1}^{n} b' \partial^1 \tilde{P}_{2,i}\left(\beta_{k,0}\right)(\hat{\phi}_i - \phi_i) u_{2,i} \right| \leq \|b\| \, O_p(\xi_{1,m_2} m_1 n^{-1}). \tag{SC.96}$$

Therefore, from Assumptions SC2(vi) and SC3(iv), (SC.66), (SC.87), (SC.93) and (SC.96) we can deduce

$$(\hat{\mathbf{P}}_2(\beta_{k,0})' \hat{\mathbf{P}}_2(\beta_{k,0}))^{-1} \sum_{i=1}^{n} (\hat{P}_{2,i}(\beta_{k,0}) - \tilde{P}_{2,i}(\beta_{k,0})) u_{2,i} = O_p(m_1^{1/2} n^{-1/2}). \tag{SC.97}$$

By Assumptions SC1(i) and SC2(v), and (SC.88),

$$n^{-1} \sum_{i=1}^{n} \tilde{P}_{2,i}(\beta_{k,0}) u_{2,i} = O_p(m_2^{1/2} n^{-1/2})$$

which together with (SC.66) implies that

$$(\hat{\mathbf{P}}_2(\beta_{k,0})' \hat{\mathbf{P}}_2(\beta_{k,0}))^{-1} \sum_{i=1}^{n} \tilde{P}_{2,i}(\beta_{k,0}) u_{2,i} = O_p(m_2^{1/2} n^{-1/2}). \tag{SC.98}$$

The claim of the lemma follows from (SC.86), (SC.97) and (SC.98). *Q.E.D.*

Lemma SC10. *Under Assumptions SC1, SC2 and SC3, we have:*

$$n^{-1} \sum_{i=1}^{n} \hat{\ell}_i(\beta_{k,0}) k_{1,i} (\hat{g}_{1,i}(\beta_{k,0}) - g_1(\nu_{1,i})) = o_p(n^{-1/2}).$$

Proof of Lemma SC10. By the definition of $\hat{\ell}_i(\beta_{k,0})$, we can write

$$\begin{aligned}
& n^{-1} \sum_{i=1}^{n} \hat{\ell}_i(\beta_{k,0}) k_{1,i} (\hat{g}_{1,i}(\beta_{k,0}) - g_1(\nu_{1,i})) \\
= \; & n^{-1} \sum_{i=1}^{n} (g(\nu_{1,i}) - \hat{g}_i(\beta_{k,0})) k_{1,i} (\hat{g}_{1,i}(\beta_{k,0}) - g_1(\nu_{1,i})) \\
& + n^{-1} \sum_{i=1}^{n} (\hat{y}^*_{2,i}(\beta_{k,0}) - g(\nu_{1,i})) k_{1,i} (\hat{g}_{1,i}(\beta_{k,0}) - g_1(\nu_{1,i})).
\end{aligned} \tag{SC.99}$$

We shall show that both terms in the right hand side of the above equation are $o_p(n^{-1/2})$. By the

Cauchy-Schwarz inequality, Lemma SC9, (SC.66) and (SC.70)

$$\begin{aligned} n^{-1}\sum_{i=1}^{n}(\hat{g}_i(\beta_{k,0})-g(\nu_{1,i}))^2 &\leq Cn^{-1}\sum_{i=1}^{n}(\hat{P}_{2,i}(\beta_{k,0})'(\hat{\beta}_g(\beta_{k,0})-\tilde{\beta}_{g,m_2}(\beta_{k,0})))^2 \\ &\quad +Cn^{-1}\sum_{i=1}^{n}(\hat{P}_{2,i}(\beta_{k,0})'\tilde{\beta}_{g,m_2}(\beta_{k,0})-g(\nu_{1,i}))^2 \\ &\leq C\left\|\hat{\beta}_g(\beta_{k,0})-\tilde{\beta}_{g,m_2}(\beta_{k,0})\right\|^2\lambda_{\max}(n^{-1}\hat{\mathbf{P}}_2(\beta_{k,0})'\hat{\mathbf{P}}_2(\beta_{k,0}))+O_p(m_1n^{-1}) \\ &= O_p((m_1+m_2)n^{-1}) \qquad \text{(SC.100)} \end{aligned}$$

Similarly, we can show that

$$\begin{aligned} n^{-1}\sum_{i=1}^{n}(\hat{g}_{1,i}(\beta_{k,0})-g_1(\nu_{1,i}))^2 &\leq Cn^{-1}\sum_{i=1}^{n}(\partial^1\hat{P}_{2,i}(\beta_{k,0})'(\hat{\beta}_g(\beta_{k,0})-\tilde{\beta}_{g,m_2}(\beta_{k,0})))^2 \\ &\quad +Cn^{-1}\sum_{i=1}^{n}((\partial^1\hat{P}_{2,i}(\beta_{k,0})-\partial^1\tilde{P}_{2,i}(\beta_{k,0}))'\tilde{\beta}_{g,m_2}(\beta_{k,0}))^2 \\ &\quad +Cn^{-1}\sum_{i=1}^{n}(\partial^1\tilde{P}_{2,i}(\beta_{k,0})'\tilde{\beta}_{g,m_2}(\beta_{k,0})-g_1(\nu_{1,i}))^2 \\ &\leq C\xi_{1,m_2}^2\left\|\hat{\beta}_g(\beta_{k,0})-\tilde{\beta}_{g,m_2}(\beta_{k,0})\right\|^2+O_p(\xi_{2,m_2}^2m_1n^{-1}) \\ &= O_p(\xi_{1,m_2}^2(m_1+m_2)n^{-1}+\xi_{2,m_2}^2m_1n^{-1}). \qquad \text{(SC.101)} \end{aligned}$$

Therefore, by the Cauchy-Schwarz inequality, Assumption SC3(iv), (SC.100) and (SC.101),

$$n^{-1}\sum_{i=1}^{n}(\hat{g}_i(\beta_{k,0})-g(\nu_{1,i}))k_{1,i}(\hat{g}_{1,i}(\beta_{k,0})-g_1(\nu_{1,i}))=o_p(n^{-1/2}). \qquad \text{(SC.102)}$$

Since $\hat{y}^*_{2,i}(\beta_{k,0})-g(\nu_{1,i})=u_{2,i}-l_{1,i}(\hat{\beta}_l-\beta_{l,0})$, we can write

$$\begin{aligned} &n^{-1}\sum_{i=1}^{n}(\hat{y}^*_{2,i}(\beta_{k,0})-g(\nu_{1,i}))k_{1,i}(\hat{g}_{1,i}(\beta_{k,0})-g_1(\nu_{1,i})) \\ &= n^{-1}\sum_{i=1}^{n}u_{2,i}k_{1,i}(\hat{g}_{1,i}(\beta_{k,0})-g_1(\nu_{1,i})) \\ &\quad -(\hat{\beta}_l-\beta_{l,0})n^{-1}\sum_{i=1}^{n}l_{1,i}k_{1,i}(\hat{g}_{1,i}(\beta_{k,0})-g_1(\nu_{1,i})). \qquad \text{(SC.103)} \end{aligned}$$

Since $k_{1,i}$ has bounded support, by Assumptions SC1(i, ii, iii), SC2(vi) and SC3(iv), (SC.101) and the Markov inequality,

$$(\hat{\beta}_l - \beta_{l,0})n^{-1}\sum_{i=1}^{n} l_{1,i}k_{1,i}(\hat{g}_{1,i}(\beta_{k,0}) - g_1(\nu_{1,i})) = o_p(n^{-1/2}). \tag{SC.104}$$

Let

$$\partial^1 \hat{P}_{2,i}(\beta_k) \equiv \partial^1 \tilde{P}_2(\hat{\nu}_{1,i}(\beta_k);\beta_k) \text{ for any } \beta_k \in \Theta_k.$$

Then we can write

$$\begin{aligned}
& n^{-1}\sum_{i=1}^{n} u_{2,i}k_{1,i}(\hat{g}_{1,i}(\beta_{k,0}) - g_1(\nu_{1,i})) \\
= \; & n^{-1}\sum_{i=1}^{n} u_{2,i}k_{1,i}(\partial^1 \hat{P}_{2,i}(\beta_{k,0}) - \partial^1 \tilde{P}_{2,i}(\beta_{k,0}))'\hat{\beta}_g(\beta_{k,0}) \\
& + n^{-1}\sum_{i=1}^{n} u_{2,i}k_{1,i}\partial^1 \tilde{P}_{2,i}(\beta_{k,0})'(\hat{\beta}_g(\beta_{k,0}) - \tilde{\beta}_{g,m_2}(\beta_{k,0})) \\
& + n^{-1}\sum_{i=1}^{n} u_{2,i}k_{1,i}\left(\partial^1 \tilde{P}_{2,i}(\beta_{k,0})'\tilde{\beta}_{g,m_2}(\beta_{k,0}) - g_1(\nu_{1,i})\right).
\end{aligned} \tag{SC.105}$$

By Assumptions SC1(i) and SC2(iii), (SC.88) and the Markov inequality, we have

$$n^{-1}\sum_{i=1}^{n} u_{2,i}k_{1,i}\left(\partial^1 \tilde{P}_{2,i}(\beta_{k,0})'\tilde{\beta}_{g,m_2}(\beta_{k,0}) - g_1(\nu_{1,i})\right) = o_p(n^{-1/2}). \tag{SC.106}$$

Similarly,

$$n^{-1}\sum_{i=1}^{n} u_{2,i}k_{1,i}\partial^1 \tilde{P}_{2,i}(\beta_{k,0}) = O_p(\xi_{1,m_2}n^{-1/2})$$

which together with Assumptions SC2(vi) and SC3(iv), and Lemma SC9 implies that

$$n^{-1}\sum_{i=1}^{n} u_{2,i}k_{1,i}\partial^1 \tilde{P}_{2,i}(\beta_{k,0})'(\hat{\beta}_g(\beta_{k,0}) - \tilde{\beta}_{g,m_2}(\beta_{k,0})) = o_p(n^{-1/2}). \tag{SC.107}$$

By Assumption SC1(i), (SC.88) and the Markov inequality

$$n^{-1}\sum_{i=1}^{n} u_{2,i}^2 k_{1,i}^2 = O_p(1). \tag{SC.108}$$

Let $\hat{\phi}_i \equiv \hat{\phi}(x_{1,i})$ and $\phi_i \equiv \phi(x_{1,i})$. By the second order expansion,

$$\begin{aligned} & n^{-1}\sum_{i=1}^{n} u_{2,i}k_{1,i}(\partial^1 \hat{P}_{2,i}(\beta_{k,0}) - \partial^1 \tilde{P}_{2,i}(\beta_{k,0}))'\hat{\beta}_g(\beta_{k,0}) \\ = \; & n^{-1}\sum_{i=1}^{n} u_{2,i}k_{1,i}(\hat{\phi}_i - \phi_i)\partial^2 \tilde{P}_{2,i}(\beta_{k,0})'\hat{\beta}_g(\beta_{k,0}) \\ & + n^{-1}\sum_{i=1}^{n} u_{2,i}k_{1,i}(\hat{\phi}_i - \phi_i)^2 \partial^3 \tilde{P}_2(\tilde{\nu}_{1,i}(\beta_{k,0}); \beta_{k,0})'\hat{\beta}_g(\beta_{k,0}) \end{aligned} \tag{SC.109}$$

where $\tilde{\nu}_{1,i}(\beta_{k,0})$ is between $\hat{\nu}_{1,i}(\beta_{k,0})$ and $\nu_{1,i}(\beta_{k,0})$. Using similar arguments for proving (SC.92), we can show that

$$n^{-1}\sum_{i=1}^{n} |u_{2,i}k_{1,i}| \, (\hat{\phi}_i - \phi_i)^2 = O_p(m_1 n^{-1}). \tag{SC.110}$$

By the Cauchy-Schwarz inequality, Assumption SC2(v), Lemma SC4, (SC.78) and (SC.110)

$$n^{-1}\sum_{i=1}^{n} u_{2,i}k_{1,i}(\hat{\phi}_i - \phi_i)^2 \partial^3 \tilde{P}_2(\tilde{\nu}_{1,i}(\beta_{k,0}); \beta_{k,0})'\hat{\beta}_g(\beta_{k,0}) = O_p(\xi_{3,m_2} m_1 n^{-1}) = o_p(n^{-1/2}) \tag{SC.111}$$

where the second equality is by Assumptions SC2(vi) and SC3(iv). By Assumptions SC1(i, v) and SC2(v), and (SC.88)

$$n^{-1}\sum_{i=1}^{n} u_{2,i}k_{1,i}P_1(x_{1,i})\partial^2 \tilde{P}_{2,i}(\beta_{k,0})' = O_p(\xi_{2,m_2} m_1^{1/2} n^{-1/2})$$

which together with Lemma SC4 and (SC.78) implies that

$$n^{-1}\sum_{i=1}^{n} u_{2,i}k_{1,i}(\hat{\phi}_i - \phi_i)\partial^2 \tilde{P}_{2,i}(\beta_{k,0})'\hat{\beta}_g(\beta_{k,0}) = O_p(\xi_{2,m_2} m_1 n^{-1}) = o_p(n^{-1/2}) \tag{SC.112}$$

where the second equality is by Assumptions SC2(vi) and SC3(iv). Similarly, we can show that

$$n^{-1}\sum_{i=1}^{n} u_{2,i}k_{1,i}(\phi_{m_1}(x_{1,i}) - \phi(x_{1,i}))\partial^2 \tilde{P}_{2,i}(\beta_{k,0})'\hat{\beta}_g(\beta_{k,0}) = o_p(n^{-1/2})$$

which together with (SC.112) implies that

$$n^{-1}\sum_{i=1}^{n} u_{2,i}k_{1,i}(\hat{\phi}_i - \phi_i)\partial^2 \tilde{P}_{2,i}(\beta_{k,0})'\hat{\beta}_g(\beta_{k,0}) = o_p(n^{-1/2}). \tag{SC.113}$$

Collecting the results in (SC.109), (SC.111) and (SC.113) we get

$$n^{-1}\sum_{i=1}^{n}u_{2,i}k_{1,i}(\partial^{1}\hat{P}_{2,i}(\beta_{k,0})-\partial^{1}\tilde{P}_{2,i}(\beta_{k,0}))'\hat{\beta}_{g}(\beta_{k,0})=o_{p}(n^{-1/2}). \tag{SC.114}$$

By (SC.103), (SC.104), (SC.105), (SC.106), (SC.107) and (SC.114),

$$n^{-1}\sum_{i=1}^{n}(\hat{y}^{*}_{2,i}(\beta_{k,0})-g(\nu_{1,i}))k_{1,i}(\hat{g}_{1,i}(\beta_{k,0})-g_{1}(\nu_{1,i}))=o_{p}(n^{-1/2}). \tag{SC.115}$$

The claim of the lemma follows from (SC.99), (SC.102) and (SC.115). *Q.E.D.*

Lemma SC11. *Under Assumptions SC1, SC2 and SC3, we have*

$$\begin{aligned}&n^{-1}\sum_{i=1}^{n}(\hat{P}_{2,i}(\beta_{k,0})'\tilde{\beta}_{g,m_2}(\beta_{k,0})-g(\nu_{1,i}))(k_{2,i}-k_{1,i}g_{1}(\nu_{1,i}))\\ =\;&n^{-1}\sum_{i=1}^{n}g_{1}(\nu_{1,i})(\hat{\phi}(x_{1,i})-\phi(x_{1,i}))(k_{2,i}-k_{1,i}g_{1}(\nu_{1,i}))+o_{p}(n^{-1/2}).\end{aligned}$$

Proof of Lemma SC11. First we write

$$\begin{aligned}&n^{-1}\sum_{i=1}^{n}(\hat{P}_{2,i}(\beta_{k,0})'\tilde{\beta}_{g,m_2}(\beta_{k,0})-g(\nu_{1,i}))(k_{2,i}-k_{1,i}g_{1}(\nu_{1,i}))\\ =\;&n^{-1}\sum_{i=1}^{n}(\hat{P}_{2,i}(\beta_{k,0})-\tilde{P}_{2,i}(\beta_{k,0}))'\tilde{\beta}_{g,m_2}(\beta_{k,0})(k_{2,i}-k_{1,i}g_{1}(\nu_{1,i}))\\ &+n^{-1}\sum_{i=1}^{n}(\tilde{P}_{2,i}(\beta_{k,0})'\tilde{\beta}_{g,m_2}(\beta_{k,0})-g(\nu_{1,i}))(k_{2,i}-k_{1,i}g_{1}(\nu_{1,i})).\end{aligned} \tag{SC.116}$$

By Assumptions SC1(i) and SC2(i, ii), and the Markov inequality

$$n^{-1}\sum_{i=1}^{n}(k_{2,i}-k_{1,i}g_{1}(\nu_{1,i}))^{2}=O_{p}(1). \tag{SC.117}$$

Therefore by Assumption SC2(iii, vi) and (SC.117), we have

$$n^{-1}\sum_{i=1}^{n}(\tilde{P}_{2,i}(\beta_{k,0})'\tilde{\beta}_{g,m_2}(\beta_{k,0})-g(\nu_{1,i}))(k_{2,i}-k_{1,i}g_{1}(\nu_{1,i}))=o_{p}(n^{-1/2}). \tag{SC.118}$$

Recall that $\hat{\phi}_i \equiv \hat{\phi}(x_{1,i})$ and $\phi_i \equiv \phi(x_{1,i})$. By the second order expansion,

$$
\begin{aligned}
& n^{-1}\sum_{i=1}^{n}(\hat{P}_{2,i}(\beta_{k,0}) - \tilde{P}_{2,i}(\beta_{k,0}))'\tilde{\beta}_{g,m_2}(\beta_{k,0})(k_{2,i} - k_{1,i}g_1(\nu_{1,i})) \\
= \; & n^{-1}\sum_{i=1}^{n}\partial^1\tilde{P}_{2,i}(\beta_{k,0})'\tilde{\beta}_{g,m_2}(\beta_{k,0})(\hat{\phi}_i - \phi_i)(k_{2,i} - k_{1,i}g_1(\nu_{1,i})) \\
& + n^{-1}\sum_{i=1}^{n}\partial^2\tilde{P}_2(\tilde{\nu}_{1,i}(\beta_{k,0});\beta_{k,0})'\tilde{\beta}_{g,m_2}(\beta_{k,0})(\hat{\phi}_i - \phi_i)^2(k_{2,i} - k_{1,i}g_1(\nu_{1,i})). \quad \text{(SC.119)}
\end{aligned}
$$

By the Cauchy-Schwarz inequality and the triangle inequality, Assumption SC2(v), (SC.59) and (SC.69)

$$
\begin{aligned}
& \left| n^{-1}\sum_{i=1}^{n}\partial^2\tilde{P}_2(\tilde{\nu}_{1,i}(\beta_{k,0});\beta_{k,0})'\tilde{\beta}_{g,m_2}(\beta_{k,0})(\hat{\phi}_i - \phi_i)^2(k_{2,i} - k_{1,i}g_1(\nu_{1,i})) \right| \\
\leq \; & O_p(\xi_{2,m_2})n^{-1}\sum_{i=1}^{n}|k_{2,i} - k_{1,i}g_1(\nu_{1,i})|\,(\hat{\phi}_i - \phi_i)^2. \quad \text{(SC.120)}
\end{aligned}
$$

Since $\mathbb{E}[|k_{2,i} - k_{1,i}g_1(\nu_{1,i})|^2\,|x_{1,i}] \leq C$ by Assumption SC2(i, ii), we can use the similar arguments for showing (SC.92) to get

$$
n^{-1}\sum_{i=1}^{n}|k_{2,i} - k_{1,i}g_1(\nu_{1,i})|\,(\hat{\phi}_i - \phi_i)^2 = O_p(m_1 n^{-1})
$$

which combined with Assumption SC3(iv) and (SC.120) implies that

$$
n^{-1}\sum_{i=1}^{n}\partial^2\tilde{P}_2(\tilde{\nu}_{1,i}(\beta_{k,0});\beta_{k,0})'\tilde{\beta}_{g,m_2}(\beta_{k,0})(\hat{\phi}_i - \phi_i)^2(k_{2,i} - k_{1,i}g_1(\nu_{1,i})) = o_p(n^{-1/2}). \quad \text{(SC.121)}
$$

By the Cauchy-Schwarz inequality, Assumption SC2(ii, iii, vi), Lemma SC4 and (SC.117)

$$
\begin{aligned}
& n^{-1}\sum_{i=1}^{n}\partial^1\tilde{P}_{2,i}(\beta_{k,0})'\tilde{\beta}_{g,m_2}(\beta_{k,0})(\hat{\phi}_i - \phi_i)(k_{2,i} - k_{1,i}g_1(\nu_{1,i})) \\
= \; & n^{-1}\sum_{i=1}^{n}g_1(\nu_{1,i})(\hat{\phi}_i - \phi_i)(k_{2,i} - k_{1,i}g_1(\nu_{1,i})) + o_p(n^{-1/2})
\end{aligned}
$$

which together with (SC.119) and (SC.121) shows that

$$\begin{aligned}
& n^{-1}\sum_{i=1}^{n}(\hat{P}_{2,i}(\beta_{k,0})-\tilde{P}_{2,i}(\beta_{k,0}))'\tilde{\beta}_{g,m_2}(\beta_{k,0})(k_{2,i}-k_{1,i}g_1(\nu_{1,i})) \\
= \;& n^{-1}\sum_{i=1}^{n}g_1(\nu_{1,i})(\hat{\phi}_i-\phi_i)(k_{2,i}-k_{1,i}g_1(\nu_{1,i}))+o_p(n^{-1/2}). \qquad \text{(SC.122)}
\end{aligned}$$

The claim of the lemma follows from (SC.116), (SC.118) and (SC.122). *Q.E.D.*

Lemma SC12. *Under Assumptions SC1, SC2 and SC3, we have*

$$\begin{aligned}
& n^{-1}\sum_{i=1}^{n}(\hat{g}_i(\beta_{k,0})-g(\nu_{1,i}))(k_{2,i}-k_{1,i}g_1(\nu_{1,i})) \\
& = n^{-1}\sum_{i=1}^{n}u_{2,i}\varphi(\nu_{1,i})-\mathbb{E}[l_{2,i}\varphi(\nu_{1,i})](\hat{\beta}_l-\beta_{l,0}) \\
& \quad + n^{-1}\sum_{i=1}^{n}g_1(\nu_{1,i})(\hat{\phi}(x_{1,i})-\phi(x_{1,i}))\varsigma_{1,i}+o_p(n^{-1/2})
\end{aligned}$$

where $\varphi(\nu_{1,i})\equiv\mathbb{E}[k_{2,i}-k_{1,i}g_1(\nu_{1,i})|\nu_{1,i}]$ *and* $\varsigma_{1,i}$ *is defined in (SB.14).*

Proof of Lemma SC12. By the definition of $\hat{g}_i(\beta_{k,0})$, we can write

$$\begin{aligned}
& n^{-1}\sum_{i=1}^{n}(\hat{g}_i(\beta_{k,0})-g(\nu_{1,i}))(k_{2,i}-k_{1,i}g_1(\nu_{1,i})) \\
= \;& (\hat{\beta}_g(\beta_{k,0})-\tilde{\beta}_{g,m_2}(\beta_{k,0}))'n^{-1}\sum_{i=1}^{n}\hat{P}_{2,i}(\beta_{k,0})(k_{2,i}-k_{1,i}g_1(\nu_{1,i})) \\
& +n^{-1}\sum_{i=1}^{n}(\hat{P}_{2,i}(\beta_{k,0})'\tilde{\beta}_{g,m_2}(\beta_{k,0})-g(\nu_{1,i}))(k_{2,i}-k_{1,i}g_1(\nu_{1,i})). \qquad \text{(SC.123)}
\end{aligned}$$

In view of Lemma SC11 and (SC.123), the claim of the lemma follows if

$$\begin{aligned}
& (\hat{\beta}_g(\beta_{k,0})-\tilde{\beta}_{g,m_2}(\beta_{k,0}))'n^{-1}\sum_{i=1}^{n}\hat{P}_{2,i}(\beta_{k,0})(k_{2,i}-k_{1,i}g_1(\nu_{1,i})) \\
= \;& n^{-1}\sum_{i=1}^{n}u_{2,i}\varphi(\nu_{1,i})-\mathbb{E}[l_{2,i}\varphi(\nu_{1,i})](\hat{\beta}_l-\beta_{l,0}) \\
& -n^{-1}\sum_{i=1}^{n}g_1(\nu_{1,i})(\hat{\phi}(x_{1,i})-\phi(x_{1,i}))\varphi(\nu_{1,i})+o_p(n^{-1/2}). \qquad \text{(SC.124)}
\end{aligned}$$

We next prove (SC.124).

Let $\hat{\beta}_{\varphi}(\beta_{k,0}) \equiv (\hat{\mathbf{P}}_2(\beta_{k,0})'\hat{\mathbf{P}}_2(\beta_{k,0}))^{-1}\sum_{i=1}^{n}\hat{P}_{2,i}(\beta_{k,0})(k_{2,i}-k_{1,i}g_1(\nu_{1,i}))$. Then we can use the decomposition in (SC.65) to write

$$\begin{aligned}
&(\hat{\beta}_g(\beta_{k,0})-\tilde{\beta}_{g,m_2}(\beta_{k,0}))'n^{-1}\sum_{i=1}^{n}\hat{P}_{2,i}(\beta_{k,0})(k_{2,i}-k_{1,i}g_1(\nu_{1,i}))\\
=\;&(\hat{\beta}_g(\beta_{k,0})-\tilde{\beta}_{g,m_2}(\beta_{k,0}))'(n^{-1}\hat{\mathbf{P}}_2(\beta_{k,0})'\hat{\mathbf{P}}_2(\beta_{k,0}))\hat{\beta}_{\varphi}(\beta_{k,0})\\
=\;&n^{-1}\sum_{i=1}^{n}\hat{\beta}_{\varphi}(\beta_{k,0})'\hat{P}_{2,i}(\beta_{k,0})(g(\nu_{1,i}(\beta_{k,0}),\beta_{k,0})-g_{m_2}(\hat{\nu}_{1,i}(\beta_{k,0}),\beta_{k,0}))\\
&-(\hat{\beta}_l-\beta_{l,0})n^{-1}\sum_{i=1}^{n}\hat{\beta}_{\varphi}(\beta_{k,0})'\hat{P}_{2,i}(\beta_{k,0})l_{2,i}+n^{-1}\sum_{i=1}^{n}\hat{\beta}_{\varphi}(\beta_{k,0})'\hat{P}_{2,i}(\beta_{k,0})u_{2,i} \qquad \text{(SC.125)}
\end{aligned}$$

where $g_{m_2}(\hat{\nu}_{1,i}(\beta_{k,0}),\beta_{k,0}) \equiv \hat{P}_{2,i}(\beta_{k,0})'\tilde{\beta}_{g,m_2}(\beta_{k,0})$. Under Assumptions SC1, SC2 and SC3, we can use the same arguments for proving Lemma SC9 to show that

$$\hat{\beta}_{\varphi}(\beta_{k,0})-\tilde{\beta}_{\varphi,m_2}(\beta_{k,0}) = O_p((m_1^{1/2}+m_2^{1/2})n^{-1/2}) = o_p(1) \qquad \text{(SC.126)}$$

where $\tilde{\beta}_{\varphi,m_2}(\beta_{k,0}) \equiv (B(\beta_{k,0})')^{-1}\beta_{\varphi,m_2}$ and β_{φ,m_2} is defined in Assumption SC3(ii). By Assumptions SC1(i, v) and SC3(ii, iv), and (SC.126), we can use similar arguments for showing (SC.69) and (SC.78) to deduce

$$\left\|\tilde{\beta}_{\varphi,m_2}(\beta_{k,0})\right\| = O(1) \text{ and } \left\|\hat{\beta}_{\varphi}(\beta_{k,0})\right\| = O_p(1). \qquad \text{(SC.127)}$$

Moreover, we can use similar arguments for proving (SC.100) to show that

$$n^{-1}\sum_{i=1}^{n}(\hat{\beta}_{\varphi}(\beta_{k,0})'\hat{P}_{2,i}(\beta_{k,0})-\varphi(\nu_{1,i}))^2 = O_p((m_1+m_2)n^{-1}). \qquad \text{(SC.128)}$$

The rest of the proof is divided into 3 steps. The claim in (SC.124) follows from (SC.125), (SC.129), (SC.131) and (SC.133) below.

Step 1. In this step, we show that

$$\begin{aligned}
&n^{-1}\sum_{i=1}^{n}\hat{\beta}_{\varphi}(\beta_{k,0})'\hat{P}_{2,i}(\beta_{k,0})(g(\nu_{1,i}(\beta_{k,0}),\beta_{k,0})-g_{m_2}(\hat{\nu}_{1,i}(\beta_{k,0}),\beta_{k,0}))\\
=\;&-n^{-1}\sum_{i=1}^{n}g_1(\nu_{1,i})(\hat{\phi}(x_{1,i})-\phi(x_{1,i}))\varphi(\nu_{1,i})+o_p(n^{-1/2}). \qquad \text{(SC.129)}
\end{aligned}$$

Recall that $\hat{\phi}_i \equiv \hat{\phi}(x_{1,i})$ and $\phi_i \equiv \phi(x_{1,i})$. By Assumptions SC2(iii, vi) and SC3(i, iv), (SC.70) and

(SC.128), we get

$$
\begin{aligned}
& n^{-1}\sum_{i=1}^{n}\hat{\beta}_{\varphi}(\beta_{k,0})'\hat{P}_{2,i}(\beta_{k,0})(g_{m_2}(\hat{\nu}_{1,i}(\beta_{k,0}),\beta_{k,0})-g(\nu_{1,i})) \\
= \;& n^{-1}\sum_{i=1}^{n}\varphi(\nu_{1,i})(g_{m_2}(\hat{\nu}_{1,i}(\beta_{k,0}),\beta_{k,0})-g(\nu_{1,i}))+o_p(n^{-1/2}) \\
= \;& n^{-1}\sum_{i=1}^{n}\varphi(\nu_{1,i})(\hat{P}_{2,i}(\beta_{k,0})-\tilde{P}_{2,i}(\beta_{k,0}))'\tilde{\beta}_{g,m_2}(\beta_{k,0})+o_p(n^{-1/2}). \qquad \text{(SC.130)}
\end{aligned}
$$

By the second order expansion, Assumptions SC2(ii, iii, v, vi) and SC3(i, iv), Lemma SC4, (SC.59) and (SC.69)

$$
\begin{aligned}
& n^{-1}\sum_{i=1}^{n}\varphi(\nu_{1,i})(\hat{P}_{2,i}(\beta_{k,0})-\tilde{P}_{2,i}(\beta_{k,0}))'\tilde{\beta}_{g,m_2}(\beta_{k,0}) \\
= \;& n^{-1}\sum_{i=1}^{n}\varphi(\nu_{1,i})(\hat{\phi}_i-\phi_i)\partial^1\tilde{P}_{2,i}(\beta_{k,0})'\tilde{\beta}_{g,m_2}(\beta_{k,0}) \\
& +n^{-1}\sum_{i=1}^{n}\varphi(\nu_{1,i})(\hat{\phi}_i-\phi_i)^2\partial^2\tilde{P}_2(\tilde{\nu}_{1,i}(\beta_{k,0});\beta_{k,0})'\tilde{\beta}_{g,m_2}(\beta_{k,0})+o_p(n^{-1/2}) \\
= \;& n^{-1}\sum_{i=1}^{n}\varphi(\nu_{1,i})(\hat{\phi}_i-\phi_i)g_1(\nu_{1,i})+o_p(n^{-1/2})
\end{aligned}
$$

which together with (SC.130) proves (SC.129).

Step 2. In this step, we show that

$$
(\hat{\beta}_l-\beta_{l,0})n^{-1}\sum_{i=1}^{n}\hat{\beta}_{\varphi}(\beta_{k,0})'\hat{P}_{2,i}(\beta_{k,0})l_{2,i}=\mathbb{E}[l_{2,i}\varphi(\nu_{1,i})](\hat{\beta}_l-\beta_{l,0})+o_p(n^{-1/2}). \qquad \text{(SC.131)}
$$

By the Cauchy-Schwarz inequality, (SC.72) and (SC.128)

$$
\begin{aligned}
n^{-1}\sum_{i=1}^{n}\hat{\beta}_{\varphi}(\beta_{k,0})'\hat{P}_{2,i}(\beta_{k,0})l_{2,i} &= n^{-1}\sum_{i=1}^{n}\varphi(\nu_{1,i})l_{2,i}+O_p((m_1^{1/2}+m_2^{1/2})n^{-1/2}) \\
&= \mathbb{E}[l_{2,i}\varphi(\nu_{1,i})]+O_p((m_1^{1/2}+m_2^{1/2})n^{-1/2}) \qquad \text{(SC.132)}
\end{aligned}
$$

where the second equality is by the Markov inequality, Assumptions SC1(i), SC2(i) and SC3(i). The claim in (SC.131) follows by Assumptions SC1(iii), SC2(i) and SC3(ii, vi), and (SC.132).

Step 3. In this step, we show that

$$
n^{-1}\sum_{i=1}^{n}\hat{\beta}_{\varphi}(\beta_{k,0})'\hat{P}_{2,i}(\beta_{k,0})u_{2,i}=n^{-1}\sum_{i=1}^{n}u_{2,i}\varphi(\nu_{1,i})+o_p(n^{-1/2}). \qquad \text{(SC.133)}
$$

By the second order expansion,

$$\begin{aligned} n^{-1}\sum_{i=1}^{n}\hat{\beta}_{\varphi}(\beta_{k,0})'\hat{P}_{2,i}(\beta_{k,0})u_{2,i} &= n^{-1}\sum_{i=1}^{n}\hat{\beta}_{\varphi}(\beta_{k,0})'\tilde{P}_{2,i}(\beta_{k,0})u_{2,i} \\ &\quad +n^{-1}\sum_{i=1}^{n}\hat{\beta}_{\varphi}(\beta_{k,0})'\partial^{1}\tilde{P}_{2,i}(\beta_{k,0})(\hat{\phi}_i-\phi_i)u_{2,i} \\ &\quad +n^{-1}\sum_{i=1}^{n}\hat{\beta}_{\varphi}(\beta_{k,0})'\partial^{2}\tilde{P}_{2}(\tilde{\nu}_{1,i};\beta_{k,0})(\hat{\phi}_i-\phi_i)^2u_{2,i} \end{aligned} \tag{SC.134}$$

which together with Assumption SC3(vi), (SC.93), (SC.96) and (SC.127) implies that

$$n^{-1}\sum_{i=1}^{n}\hat{\beta}_{\varphi}(\beta_{k,0})'\hat{P}_{2,i}(\beta_{k,0})u_{2,i}=n^{-1}\sum_{i=1}^{n}\hat{\beta}_{\varphi}(\beta_{k,0})'\tilde{P}_{2,i}(\beta_{k,0})u_{2,i}+o_p(n^{-1/2}). \tag{SC.135}$$

Since by the Markov inequality, Assumptions SC1(i) and SC2(iv)

$$n^{-1}\sum_{i=1}^{n}\tilde{P}_{2,i}(\beta_{k,0})u_{2,i}=O_p(m_1^{1/2}n^{-1/2}) \tag{SC.136}$$

we deduce that

$$\begin{aligned} n^{-1}\sum_{i=1}^{n}\hat{\beta}_{\varphi}(\beta_{k,0})'\tilde{P}_{2,i}(\beta_{k,0})u_{2,i} &= n^{-1}\sum_{i=1}^{n}\tilde{\beta}_{\varphi,m_2}(\beta_{k,0})'\tilde{P}_{2,i}(\beta_{k,0})u_{2,i}+o_p(n^{-1/2}) \\ &= n^{-1}\sum_{i=1}^{n}\varphi(\nu_{1,i})u_{2,i}+o_p(n^{-1/2}) \end{aligned}$$

where the first equality is by (SC.126), (SC.136) and Assumption SC3(vi), the second equality is by Assumptions SC1(i) and SC3(ii), (SC.88) and the Markov inequality. *Q.E.D.*

Lemma SC13. *Under Assumptions SC1, SC2 and SC3, we have*

$$\begin{aligned} & n^{-1}\sum_{i=1}^{n}g_1(\nu_{1,i})(\hat{\phi}(x_{1,i})-\phi(x_{1,i}))\varsigma_{1,i} \\ = \;& n^{-1}\sum_{i=1}^{n}\eta_{1,i}g_1(\nu_{1,i})(\varsigma_{1,i}-\varsigma_{2,i})-\mathbb{E}[l_{1,i}g_1(\nu_{1,i})(\varsigma_{1,i}-\varsigma_{2,i})](\hat{\beta}_l-\beta_{l,0})+o_p(n^{-1/2}) \end{aligned}$$

where $\varsigma_{2,i}\equiv k_{2,i}-\mathbb{E}[k_{2,i}|x_{1,i}]$.

Proof of Lemma SC13. Since $\hat{\phi}(x_{1,i})-\phi(x_{1,i})=(\hat{\beta}_{\phi}-\beta_{\phi,m_1})'P_1(x_{1,i})+\phi_{m_1}(x_{1,i})-\phi(x_{1,i})$, we can write

$$
\begin{aligned}
&n^{-1}\sum_{i=1}^{n} g_{1,i}(\hat{\phi}(x_{1,i})-\phi(x_{1,i}))\varsigma_{1,i} \\
&= (\hat{\beta}_{\phi}-\beta_{\phi,m_1})' n^{-1}\sum_{i=1}^{n} P_1(x_{1,i}) g_{1,i}\varsigma_{1,i} + n^{-1}\sum_{i=1}^{n} g_{1,i}(\phi_{m_1}(x_{1,i})-\phi(x_{1,i}))\varsigma_{1,i} \qquad \text{(SC.137)}
\end{aligned}
$$

where $g_{1,i} \equiv g_1(\nu_{1,i})$. By Assumptions SC1(i, iv, vi) and SC2(ii), and the Markov inequality

$$
n^{-1}\sum_{i=1}^{n} g_{1,i}\left(\phi_{m_1}(x_{1,i})-\phi(x_{1,i})\right)\varsigma_{1,i} = o_p(n^{-1/2}). \qquad \text{(SC.138)}
$$

By the definition of $\varsigma_{1,i}$ and $\varsigma_{2,i}$, we can write

$$
\varsigma_{1,i} = \mathbb{E}[\varsigma_{1,i}|x_{1,i}] + \varsigma_{2,i}. \qquad \text{(SC.139)}
$$

By Assumptions SC1(i, v, vi) and SC2(i, ii), and the Markov inequality

$$
n^{-1}\sum_{i=1}^{n} P_1(x_{1,i}) g_{1,i}\varsigma_{1,i} - \mathbb{E}\left[P_1(x_{1,i}) g_{1,i}\varsigma_{1,i}\right] = O_p(m_1^{1/2} n^{-1/2})
$$

which together with Assumption SC3(iv), (SC.55) and (SC.139) implies that

$$
\begin{aligned}
& (\hat{\beta}_{\phi}-\beta_{\phi,m_1})' n^{-1}\sum_{i=1}^{n} P_1(x_{1,i}) g_{1,i}\varsigma_{1,i} \\
= \; & n^{-1}\sum_{i=1}^{n} \eta_{1,i} P_1(x_{1,i})' Q_{m_1}^{-1}\mathbb{E}\left[P_1(x_{1,i}) g_{1,i}(\varsigma_{1,i}-\varsigma_{2,i})\right] \\
& -(\hat{\beta}_{l}-\beta_{l,0})\mathbb{E}[l_{1,i}P_1(x_{1,i})'] Q_{m_1}^{-1}\mathbb{E}\left[P_1(x_{1,i}) g_{1,i}(\varsigma_{1,i}-\varsigma_{2,i})\right] + o_p(n^{-1/2}). \qquad \text{(SC.140)}
\end{aligned}
$$

By Assumptions SC1(i, ii, v), SC2(i, ii) and SC3(iii)

$$
\begin{aligned}
& \mathbb{E}\left[\left| n^{-1}\sum_{i=1}^{n} \eta_{1,i}\left[P_1(x_{1,i})' Q_{m_1}^{-1}\mathbb{E}[P_1(x_{1,i}) g_{1,i}(\varsigma_{1,i}-\varsigma_{2,i})] - g_{1,i}(\varsigma_{1,i}-\varsigma_{2,i})\right]\right|^2\right] \\
& \leq C n^{-1}\mathbb{E}\left[\left| P_1(x_{1,i})' Q_{1,m_1}^{-1}\mathbb{E}[P_1(x_{1,i}) g_{1,i}(\varsigma_{1,i}-\varsigma_{2,i})] - g_{1,i}(\varsigma_{1,i}-\varsigma_{2,i})\right|^2\right] = o(n^{-1})
\end{aligned}
$$

which together with the Markov inequality implies that

$$
n^{-1}\sum_{i=1}^{n} \eta_{1,i} P_1(x_{1,i})' Q_{m_1}^{-1}\mathbb{E}[P_{1,i} g_{1,i}(\varsigma_{1,i}-\varsigma_{2,i})] = n^{-1}\sum_{i=1}^{n} \eta_{1,i} g_{1,i}(\varsigma_{1,i}-\varsigma_{2,i}) + o_p(n^{-1/2}). \qquad \text{(SC.141)}
$$

By Hölder's inequality, Assumptions SC1(ii, v), SC2(ii) and SC3(iii)

$$\begin{aligned}
& \left|\mathbb{E}\left[l_{1,i}P_1(x_{1,i})'\right]Q_{m_1}^{-1}\mathbb{E}\left[P_1(x_{1,i})g_{1,i}(\varsigma_{1,i}-\varsigma_{2,i})\right]-\mathbb{E}[l_{1,i}g_{1,i}(\varsigma_{1,i}-\varsigma_{2,i})]\right|^2 \\
= & \left|\mathbb{E}\left[l_{1,i}\left(P_1(x_{1,i})'Q_{m_1}^{-1}\mathbb{E}\left[P_1(x_{1,i})g_{1,i}(\varsigma_{1,i}-\varsigma_{2,i})\right]-g_{1,i}(\varsigma_{1,i}-\varsigma_{2,i})\right)\right]\right|^2 \\
\leq & \ \mathbb{E}\left[l_{1,i}^2\right]\mathbb{E}\left[\left(P_1(x_{1,i})'Q_{m_1}^{-1}\mathbb{E}\left[P_1(x_{1,i})g_{1,i}(\varsigma_{1,i}-\varsigma_{2,i})\right]-g_{1,i}(\varsigma_{1,i}-\varsigma_{2,i})\right)^2\right]=o(1)
\end{aligned}$$

which combined with Assumption SC1(iii) implies that

$$\begin{aligned}
& (\hat{\beta}_l-\beta_{l,0})\mathbb{E}[l_{1,i}P_1(x_{1,i})']Q_{m_1}^{-1}\mathbb{E}\left[P_1(x_{1,i})g_{1,i}(\varsigma_{1,i}-\varsigma_{2,i})\right] \\
= & \ (\hat{\beta}_l-\beta_{l,0})\mathbb{E}[l_{1,i}g_{1,i}(\varsigma_{1,i}-\varsigma_{2,i})]+o_p(n^{-1/2}) \qquad \text{(SC.142)}
\end{aligned}$$

The claim of the lemma follows from (SC.137), (SC.138), (SC.140), (SC.141) and (SC.142).*Q.E.D.*

Lemma SC14. *Under Assumptions SC1, SC2 and SC3, we have*

$$\hat{\beta}_g(\hat{\beta}_k)-\tilde{\beta}_{g,m_2}(\hat{\beta}_k)=(\hat{\beta}_k-\beta_{k,0})O_p(\xi_{1,m_2}m_1^{1/2}n^{-1/2})+O_p((m_1^{1/2}+m_2)n^{-1/2}).$$

Proof of Lemma SC14. Using the decomposition in (SC.65), and applying the results in (SC.71), (SC.73) and (SC.77), we have

$$\hat{\beta}_g(\hat{\beta}_k)-\tilde{\beta}_{g,m_2}(\hat{\beta}_k)=(\hat{\mathbf{P}}_2(\hat{\beta}_k)'\hat{\mathbf{P}}_2(\hat{\beta}_k))^{-1}\sum_{i=1}^n(\hat{P}_{2,i}(\hat{\beta}_k)-\tilde{P}_{2,i}(\hat{\beta}_k))u_{2,i}(\hat{\beta}_k)+O_p((m_1^{1/2}+m_2)n^{-1/2}). \qquad \text{(SC.143)}$$

By the second-order expansion, we have for any $b\in\mathbb{R}^{m_2}$

$$\begin{aligned}
& n^{-1}\sum_{i=1}^n b'(\hat{P}_{2,i}(\hat{\beta}_k)-\tilde{P}_{2,i}(\hat{\beta}_k))u_{2,i}(\hat{\beta}_k) \qquad \text{(SC.144)} \\
= & \ n^{-1}\sum_{i=1}^n b'\partial^1\tilde{P}_{2,i}(\hat{\beta}_k)(\hat{\phi}_i-\phi_i)u_{2,i}(\hat{\beta}_k)+n^{-1}\sum_{i=1}^n b'\partial^2\hat{P}_2(\tilde{\nu}_{1,i}(\hat{\beta}_k);\hat{\beta}_k)(\hat{\phi}_i-\phi_i)^2u_{2,i}(\hat{\beta}_k)
\end{aligned}$$

where $\tilde{\nu}_{1,i}(\hat{\beta}_k)$ lies between $\hat{\nu}_{1,i}(\hat{\beta}_k)$ and $\nu_{1,i}(\hat{\beta}_k)$. By (SC.68) and the compactness of Θ_k,

$$\sup_{\beta_k\in\Theta_k}|u_{2,i}(\beta_k)|\leq C+|y_{2,i}^*|+|k_{2,i}|. \qquad \text{(SC.145)}$$

Using similar arguments in showing (SC.92), we have

$$n^{-1}\sum_{i=1}^n(\hat{\phi}_i-\phi_i)^2(C+|y_{2,i}^*|+|k_{2,i}|)=O_p(m_1n^{-1})$$

which together with the Cauchy-Schwarz inequality, the triangle inequality, Assumptions SC2(vi) and SC3(iv), and (SC.145) implies that

$$\begin{aligned} & \left| n^{-1}\sum_{i=1}^{n} b'\partial^2 \hat{P}_{2,i}(\tilde{\nu}_{1,i}(\hat{\beta}_k);\hat{\beta}_k)(\hat{\phi}_i-\phi_i)^2 u_{2,i}(\hat{\beta}_k)\right| \\ \leq \;& \|b\|\,\xi_{2,m_2} n^{-1}\sum_{i=1}^{n}(\hat{\phi}_i-\phi_i)^2(C+|y^*_{2,i}|) = \|b\|\,o_p(m_1^{1/2}n^{-1/2}). \end{aligned} \quad \text{(SC.146)}$$

Since $u_{2,i}(\hat{\beta}_k) = u_{2,i} - k_{2,i}(\hat{\beta}_k-\beta_{k,0}) - (g(\nu_{1,i}(\hat{\beta}_k),\hat{\beta}_k) - g(\nu_{1,i}))$, we can write

$$\begin{aligned} & n^{-1}\sum_{i=1}^{n} b'\partial^1 \tilde{P}_{2,i}(\hat{\beta}_k)(\hat{\phi}_i-\phi_i)u_{2,i}(\hat{\beta}_k) \\ = \;& n^{-1}\sum_{i=1}^{n} b'\partial^1 \tilde{P}_{2,i}(\hat{\beta}_k)(\hat{\phi}_i-\phi_i)u_{2,i} \\ & -(\hat{\beta}_k-\beta_{k,0})n^{-1}\sum_{i=1}^{n} b'\partial^1 \tilde{P}_{2,i}(\hat{\beta}_k)k_{2,i}(\hat{\phi}_i-\phi_i) \\ & -n^{-1}\sum_{i=1}^{n} b'\partial^1 \tilde{P}_{2,i}(\hat{\beta}_k)(\hat{\phi}_i-\phi_i)(g(\nu_{1,i}(\hat{\beta}_k),\hat{\beta}_k)-g(\nu_{1,i})). \end{aligned} \quad \text{(SC.147)}$$

By the Cauchy-Schwarz inequality, the triangle inequality, Assumption SC2(i, v) and Lemma SC4

$$\left| n^{-1}\sum_{i=1}^{n} b'\partial^1 \tilde{P}_{2,i}(\hat{\beta}_k)k_{2,i}(\hat{\phi}_i-\phi_i)\right| \leq \|b\|\,O_p(\xi_{1,m_2}m_1^{1/2}n^{-1/2}). \quad \text{(SC.148)}$$

Similarly we can show that

$$\begin{aligned} & \left| n^{-1}\sum_{i=1}^{n} b'\partial^1 \tilde{P}_{2,i}(\hat{\beta}_k)(\hat{\phi}_i-\phi_i)(g(\nu_{1,i}(\hat{\beta}_k),\hat{\beta}_k)-g(\nu_{1,i}))\right| \\ \leq \;& \|b\|\,|\hat{\beta}_k-\beta_{k,0}|O_p(\xi_{1,m_2}m_1^{1/2}n^{-1/2}). \end{aligned} \quad \text{(SC.149)}$$

By the Cauchy-Schwarz inequality, the triangle inequality, Assumption SC3(iv), Lemma SC35, Lemma SC36 and (SC.55),

$$\left| n^{-1}\sum_{i=1}^{n} b'\partial^1 \tilde{P}_{2,i}(\hat{\beta}_k)(\hat{\phi}_i-\phi_i)u_{2,i}\right| \leq \|b\|\,O_p(m_2^{5/2}m_1 n^{-1}) \leq \|b\|\,O_p((m_1^{1/2}+m_2)n^{-1/2}) \quad \text{(SC.150)}$$

Collecting the results in (SC.144), (SC.146), (SC.147), (SC.148), (SC.149) and (SC.150), we have

$$n^{-1}\sum_{i=1}^{n}(\hat{P}_{2,i}(\hat{\beta}_k)-\tilde{P}_{2,i}(\hat{\beta}_k))u_{2,i}(\hat{\beta}_k)=(\hat{\beta}_k-\beta_{k,0})O_p(\xi_{1,m_2}m_1^{1/2}n^{-1/2})+O_p((m_1^{1/2}+m_2)n^{-1/2})$$

which together with (SC.143) proves the claim of the lemma. *Q.E.D.*

Lemma SC15. *Under Assumptions SC1, SC2 and SC3, we have*

$$n^{-1}\sum_{i=1}^{n}\left|\hat{g}_i(\hat{\beta}_k)-g(\nu_{1,i}(\hat{\beta}_k);\hat{\beta}_k)\right|^2=(\hat{\beta}_k-\beta_{k,0})^2O_p(\xi_{1,m_2}^2m_1n^{-1})+O_p\left((m_1+m_2^2)n^{-1}\right)$$

where $\hat{g}_i(\hat{\beta}_k)\equiv\hat{g}(\hat{\nu}_{1,i}(\hat{\beta}_k);\hat{\beta}_k)$.

Proof of Lemma SC15. First note that by (SC.70),

$$n^{-1}\sum_{i=1}^{n}\left|g_{m_2}(\hat{\nu}_{1,i}(\hat{\beta}_k);\hat{\beta}_k)-g(\nu_{1,i}(\hat{\beta}_k),\hat{\beta}_k)\right|^2=O_p(m_1n^{-1}) \tag{SC.151}$$

where $g_{m_2}(\hat{\nu}_{1,i}(\hat{\beta}_k);\hat{\beta}_k)\equiv\hat{P}_{2,i}(\hat{\beta}_k)'\tilde{\beta}_{g,m_2}(\hat{\beta}_k)$. By Lemma SC14 and (SC.66)

$$\begin{aligned}&n^{-1}\sum_{i=1}^{n}\left|\hat{P}_{2,i}(\hat{\beta}_k)'\hat{\beta}_g(\hat{\beta}_k)-g_{m_2}(\hat{\nu}_{1,i}(\hat{\beta}_k);\hat{\beta}_k)\right|^2\\ \leq\;&\lambda_{\max}(n^{-1}\hat{\mathbf{P}}_2(\hat{\beta}_k)'\hat{\mathbf{P}}_2(\hat{\beta}_k))||\hat{\beta}_g(\hat{\beta}_k)-\tilde{\beta}_{g,m_2}(\hat{\beta}_k)||^2\\ =\;&(\hat{\beta}_k-\beta_{k,0})^2O_p(\xi_{1,m_2}^2m_1n^{-1})+O_p\left((m_1+m_2^2)n^{-1}\right)\end{aligned}$$

which together with (SC.151) finishes the proof. *Q.E.D.*

Lemma SC16. *Under Assumptions SC1, SC2 and SC3, we have*

$$n^{-1}\sum_{i=1}^{n}\left|\hat{g}_{1,i}(\hat{\beta}_k)-g_1(\nu_{1,i}(\hat{\beta}_k);\hat{\beta}_k)\right|^2=o_p(1)$$

where $\hat{g}_{1,i}(\hat{\beta}_k)\equiv\partial^1\hat{P}_{2,i}(\hat{\beta}_k)'\hat{\beta}_g(\hat{\beta}_k)$.

Proof of Lemma SC16. First, we can use similar arguments for showing (SC.70) to get

$$n^{-1}\sum_{i=1}^{n}\left|\partial^1\hat{P}_{2,i}(\hat{\beta}_k)'\tilde{\beta}_{g,m_2}(\hat{\beta}_k)-g_1(\nu_{1,i}(\hat{\beta}_k),\hat{\beta}_k)\right|^2=O_p(\xi_{2,m_2}^2m_1n^{-1}). \tag{SC.152}$$

By Assumption SC2(v), Lemma SC14 and the consistency of $\hat{\beta}_k$

$$\begin{aligned}
& n^{-1}\sum_{i=1}^{n}\left|\partial^1\hat{P}_{2,i}(\hat{\beta}_k)'\hat{\beta}_g(\hat{\beta}_k)-\partial^1\hat{P}_{2,i}(\hat{\beta}_k)'\tilde{\beta}_{g,m_2}(\hat{\beta}_k)\right|^2 \\
\leq\ & \xi_{1,m_2}^2||\hat{\beta}_g(\hat{\beta}_k)-\tilde{\beta}_{g,m_2}(\hat{\beta}_k)||^2 \\
=\ & o_p(\xi_{1,m_2}^4 m_1 n^{-1})+O_p\left(\xi_{1,m_2}^2(m_1+m_2^2)n^{-1}\right) \qquad \text{(SC.153)}
\end{aligned}$$

which together with Assumption SC3(iv) and (SC.152) proves the claim of the lemma. *Q.E.D.*

Lemma SC17. *Under Assumptions SC1, SC2 and SC3, we have*

$$n^{-1}\sum_{i=1}^{n}k_{2,i}(k_{2,i}-k_{1,i}\hat{g}_{1,i}(\hat{\beta}_k))=\mathbb{E}[k_{2,i}(k_{2,i}-k_{1,i}g_1(\nu_{1,i}))]+o_p(1) \qquad \text{(SC.154)}$$

and

$$n^{-1}\sum_{i=1}^{n}l_{2,i}k_{1,i}(\hat{g}_{1,i}(\hat{\beta}_k)-\hat{g}_{1,i}(\beta_{k,0})=o_p(1). \qquad \text{(SC.155)}$$

Proof of Lemma SC17. By the Cauchy-Schwarz inequality, Assumptions SC2(i, ii, vi) and SC3(iv), Lemma SC16 and the consistency of $\hat{\beta}_k$, we have

$$\begin{aligned}
n^{-1}\sum_{i=1}^{n}k_{2,i}(k_{2,i}-k_{1,i}\hat{g}_{1,i}(\hat{\beta}_k)) &= n^{-1}\sum_{i=1}^{n}k_{2,i}(k_{2,i}-k_{1,i}g_{1,i}(\hat{\beta}_k))+o_p(1) \\
&= n^{-1}\sum_{i=1}^{n}k_{2,i}(k_{2,i}-k_{1,i}g_1(\nu_{1,i}))+o_p(1) \\
&= \mathbb{E}[k_{2,i}(k_{2,i}-k_{1,i}g_1(\nu_{1,i}))]+o_p(1)
\end{aligned}$$

where the third equality is by the Markov inequality. This proves the claim in (SC.154). Similarly, by Assumption SC2(ii), Lemma SC16 and the consistency of $\hat{\beta}_k$, we have

$$\begin{aligned}
& n^{-1}\sum_{i=1}^{n}\left|\hat{g}_{1,i}(\hat{\beta}_k)-\hat{g}_{1,i}(\beta_{k,0})\right|^2 \\
& \leq 2n^{-1}\sum_{i=1}^{n}\left|g_1(\nu_{1,i}(\hat{\beta}_k);\hat{\beta}_k)-g_1(\nu_{1,i}(\beta_{k,0});\beta_{k,0})\right|^2+o_p(1) \\
& \leq C(\hat{\beta}_k-\beta_{k,0})^2+o_p(1)=o_p(1). \qquad \text{(SC.156)}
\end{aligned}$$

By the Markov inequality and Assumption SC2(i), $n^{-1}\sum_{i=1}^{n}l_{2,i}^2k_{1,i}^2=O_p(1)$ which together with (SC.156) proves the claim in (SC.155). *Q.E.D.*

Lemma SC18. *Under Assumptions SC1, SC2 and SC3, we have*

$$\begin{aligned} & n^{-1}\sum_{i=1}^{n}(\hat{g}_i(\hat{\beta}_k)-\hat{g}_i(\beta_{k,0}))(k_{2,i}-k_{1,i}\hat{g}_{1,i}(\beta_{k,0})) \\ = & -(\hat{\beta}_k-\beta_{k,0})\left(\mathbb{E}[(\gamma_{2,i}+\epsilon_{1,i}g_{1,i})(k_{2,i}-k_{1,i}g_{1,i})]+o_p(1)\right)+O_p((m_2+m_1^{1/2})n^{-1/2}) \end{aligned}$$

where $g_{1,i}\equiv g_1(\nu_{1,i})$, $\epsilon_{1,i}\equiv k_{1,i}-\mathbb{E}[k_{1,i}|\nu_{1,i}]$, $\gamma_{2,i}\equiv\mathbb{E}[k_{2,i}|\nu_{1,i}]$, $\nu_{1,i}$ *and* $g_1(\cdot)$ *are defined in (SA.5) and (SB.13) respectively.*

Proof of Lemma SC18. First note that

$$\begin{aligned} & n^{-1}\sum_{i=1}^{n}(\hat{g}_i(\hat{\beta}_k)-\hat{g}_i(\beta_{k,0}))(k_{2,i}-k_{1,i}\hat{g}_{1,i}(\beta_{k,0})) \\ = & -n^{-1}\sum_{i=1}^{n}k_{1,i}(\hat{g}_i(\hat{\beta}_k)-\hat{g}_i(\beta_{k,0}))(\hat{g}_{1,i}(\beta_{k,0})-g_{1,i}) \\ & +n^{-1}\sum_{i=1}^{n}(\hat{g}_i(\hat{\beta}_k)-g(\nu_{1,i}(\hat{\beta}_k);\hat{\beta}_k)-\hat{g}_i(\beta_{k,0})+g(\nu_{1,i}))(k_{2,i}-k_{1,i}g_{1,i}) \\ & +n^{-1}\sum_{i=1}^{n}(g(\nu_{1,i}(\hat{\beta}_k);\hat{\beta}_k)-g(\nu_{1,i}))(k_{2,i}-k_{1,i}g_{1,i}). \end{aligned} \quad \text{(SC.157)}$$

By the Cauchy-Schwarz inequality, Assumption SC3(iv), Lemma SC15 and (SC.101),

$$n^{-1}\sum_{i=1}^{n}k_{1,i}(\hat{g}_i(\hat{\beta}_k)-\hat{g}_i(\beta_{k,0}))(\hat{g}_{1,i}(\beta_{k,0})-g_{1,i})=(\hat{\beta}_k-\beta_{k,0})o_p(1)+o_p(n^{-1/2}). \quad \text{(SC.158)}$$

Similarly, we can use the Cauchy-Schwarz inequality, Lemma SC15, the consistency of $\hat{\beta}_k$ and (SC.117) to get

$$\begin{aligned} & n^{-1}\sum_{i=1}^{n}(\hat{g}_i(\hat{\beta}_k)-g(\nu_{1,i}(\hat{\beta}_k);\hat{\beta}_k)-\hat{g}_i(\beta_{k,0})+g(\nu_{1,i}))(k_{2,i}-k_{1,i}g_{1,i}) \\ = & (\hat{\beta}_k-\beta_{k,0})o_p(1)+O_p((m_2+m_1^{1/2})n^{-1/2}). \end{aligned} \quad \text{(SC.159)}$$

Moreover, by Assumptions SC2(ii) and the consistency of $\hat{\beta}_k$

$$\begin{aligned} & n^{-1}\sum_{i=1}^{n}(g(\nu_{1,i}(\hat{\beta}_k);\hat{\beta}_k)-g(\nu_{1,i}(\beta_{k,0});\beta_{k,0}))(k_{2,i}-k_{1,i}g_{1,i}) \\ = & (\hat{\beta}_k-\beta_{k,0})n^{-1}\sum_{i=1}^{n}\frac{\partial g(\nu_{1,i}(\beta_{k,0});\beta_{k,0})}{\partial\beta_k}(k_{2,i}-k_{1,i}g_{1,i})+(\hat{\beta}_k-\beta_{k,0})o_p(1). \end{aligned} \quad \text{(SC.160)}$$

Since

$$\begin{aligned}
\frac{\partial g(\nu_{1,i}(\beta_{k,0});\beta_{k,0})}{\partial \beta_k} &= \left.\frac{\partial(\gamma_1(\nu_{1,i}(\beta_k)) - \beta_k\gamma_2(\nu_{1,i}(\beta_k)))}{\partial \beta_k}\right|_{\beta_k=\beta_{k,0}} \\
&= -\gamma_2(\nu_{1,i}) + \left.\frac{\partial(\gamma_1(\nu_{1,i}(\beta_k)) - \beta_{k,0}\gamma_2(\nu_{1,i}(\beta_k)))}{\partial \beta_k}\right|_{\beta_k=\beta_{k,0}} \\
&= -\gamma_2(\nu_{1,i}) - g_{1,i}(k_{1,i} - \mathbb{E}[k_{1,i}|\nu_{1,i}])
\end{aligned}$$

where the third equality is by the derivative formula in Newey (1994) (Example 1 Continued, p.1358), by Assumptions SC1(i) and SC2(i, ii), and the Markov inequality,

$$\begin{aligned}
&n^{-1}\sum_{i=1}^{n}\frac{\partial g(\nu_{1,i}(\beta_{k,0});\beta_{k,0})}{\partial \beta_k}(k_{2,i} - k_{1,i}g_{1,i}) \\
&= -n^{-1}\sum_{i=1}^{n}(\gamma_{2,i} + \epsilon_{1,i}g_{1,i})(k_{2,i} - k_{1,i}g_{1,i}) = -\mathbb{E}[(\gamma_{2,i} + \epsilon_{1,i}g_{1,i})(k_{2,i} - k_{1,i}g_{1,i})] + O_p(n^{-1/2})
\end{aligned}$$

which together with (SC.160) implies that

$$\begin{aligned}
& n^{-1}\sum_{i=1}^{n}(g(\nu_{1,i}(\hat{\beta}_k);\hat{\beta}_k) - g(\nu_{1,i}(\beta_{k,0});\beta_{k,0}))(k_{2,i} - k_{1,i}g_{1,i}) \\
= \;& -(\hat{\beta}_k - \beta_{k,0})\left(\mathbb{E}[(\gamma_{2,i} + \epsilon_{1,i}g_{1,i})(k_{2,i} - k_{1,i}g_{1,i})] + o_p(1)\right) + o_p(n^{-1/2}). \qquad \text{(SC.161)}
\end{aligned}$$

The claim of the lemma follows from (SC.157), (SC.158), (SC.159) and (SC.161). *Q.E.D.*

Lemma SC19. *Under Assumptions SC1, SC2 and SC3, we have*

$$n^{-1}\sum_{i=1}^{n}u_{2,i}k_{1,i}(\hat{g}_{1,i}(\hat{\beta}_k) - \hat{g}_{1,i}(\beta_{k,0})) = (\hat{\beta}_k - \beta_{k,0})o_p(1) + o_p(n^{-1/2}).$$

Proof of Lemma SC19. By the second order expansion,

$$\begin{aligned}
n^{-1}\sum_{i=1}^{n}u_{2,i}k_{1,i}\hat{g}_{1,i}(\hat{\beta}_k) &= n^{-1}\sum_{i=1}^{n}u_{2,i}k_{1,i}\partial^1\hat{P}_{2,i}(\hat{\beta}_k)'\hat{\beta}_g(\hat{\beta}_k) \\
&= n^{-1}\sum_{i=1}^{n}u_{2,i}k_{1,i}\partial^1\tilde{P}_{2,i}(\hat{\beta}_k)'\hat{\beta}_g(\hat{\beta}_k) \\
&\quad + n^{-1}\sum_{i=1}^{n}u_{2,i}k_{1,i}(\hat{\phi}_i - \phi_i)\partial^2\tilde{P}_{2,i}(\hat{\beta}_k)'\hat{\beta}_g(\hat{\beta}_k) \\
&\quad + n^{-1}\sum_{i=1}^{n}u_{2,i}k_{1,i}(\hat{\phi}_i - \phi_i)^2\partial^3\tilde{P}_2(\tilde{\nu}_{1,i};\hat{\beta}_k)'\hat{\beta}_g(\hat{\beta}_k) \qquad \text{(SC.162)}
\end{aligned}$$

where $\tilde{\nu}_{1,i}$ is between $\hat{\nu}_{1,i}(\hat{\beta}_k)$ and $\nu_{1,i}(\hat{\beta}_k)$. By (SC.59), $\tilde{\nu}_{1,i} \in \Omega_\varepsilon(\beta_k)$ for any $i = 1, \ldots, n$ wpa1. By Assumption SC3(iv), Lemma SC14 and Lemma SC34

$$n^{-1}\sum_{i=1}^{n} u_{2,i}k_{1,i}\partial^1 \tilde{P}_{2,i}(\hat{\beta}_k)'(\hat{\beta}_g(\hat{\beta}_k) - \tilde{\beta}_{g,m_2}(\hat{\beta}_k)) = (\hat{\beta}_k - \beta_{k,0})o_p(1) + o_p(n^{-1/2})$$

which together with Assumption SC2(iii, vi) implies that

$$\begin{aligned} & n^{-1}\sum_{i=1}^{n} u_{2,i}k_{1,i}\partial^1 \tilde{P}_{2,i}(\hat{\beta}_k)'\hat{\beta}_g(\hat{\beta}_k) \\ = \; & n^{-1}\sum_{i=1}^{n} u_{2,i}k_{1,i}g_1(\nu_{1,i}(\hat{\beta}_k);\hat{\beta}_k) + (\hat{\beta}_k - \beta_{k,0})o_p(1) + o_p(n^{-1/2}). \end{aligned} \tag{SC.163}$$

Using similar arguments for proving (SC.149) we can show that

$$\left| n^{-1}\sum_{i=1}^{n} u_{2,i}k_{1,i}(\hat{\phi}_i - \phi_i)\partial^2 \tilde{P}_{2,i}(\hat{\beta}_k)'b \right| \leq \|b\| \, O_p(m_2^4 m_1 n^{-1}) \tag{SC.164}$$

for any $b \in \mathbb{R}^{m_2}$. By the Cauchy-Schwarz inequality, Assumption SC3(iv), (SC.78) and (SC.164)

$$\left| n^{-1}\sum_{i=1}^{n} u_{2,i}k_{1,i}(\hat{\phi}_i - \phi_i)\partial^2 \tilde{P}_{2,i}(\hat{\beta}_k)'\hat{\beta}_g(\hat{\beta}_k) \right| \leq \left\| \hat{\beta}_g(\hat{\beta}_k) \right\| O_p(m_2^{7/2} m_1 n^{-1}) = o_p(n^{-1/2}) \tag{SC.165}$$

By the Cauchy-Schwarz inequality and the triangle inequality, Assumption SC2(v), (SC.59), (SC.78) and (SC.92)

$$\begin{aligned} & \left| n^{-1}\sum_{i=1}^{n} u_{2,i}k_{1,i}(\hat{\phi}_i - \phi_i)^2\partial^3 \tilde{P}_2(\tilde{\nu}_{1,i};\hat{\beta}_k)'\hat{\beta}_g(\hat{\beta}_k) \right| \\ \leq \; & O_p(\xi_{3,m_2})n^{-1}\sum_{i=1}^{n} |u_{2,i}| \, (\hat{\phi}_i - \phi_i)^2 = O_p(\xi_{3,m_2} m_1 n^{-1}) = o_p(n^{-1/2}) \end{aligned} \tag{SC.166}$$

where the second equality is by Assumption SC3(iv). Combining the results in (SC.162), (SC.163), (SC.165) and (SC.166), we get

$$n^{-1}\sum_{i=1}^{n} u_{2,i}k_{1,i}\hat{g}_{1,i}(\hat{\beta}_k) = n^{-1}\sum_{i=1}^{n} u_{2,i}k_{1,i}g_1(\nu_{1,i}(\hat{\beta}_k);\hat{\beta}_k) + (\hat{\beta}_k - \beta_{k,0})o_p(1) + o_p(n^{-1/2}). \tag{SC.167}$$

Similarly, we can show that

$$n^{-1}\sum_{i=1}^{n} u_{2,i}k_{1,i}\hat{g}_{1,i}(\beta_{k,0}) = n^{-1}\sum_{i=1}^{n} u_{2,i}k_{1,i}g_1(\nu_{1,i}(\beta_{k,0});\beta_{k,0}) + o_p(n^{-1/2})$$

which together with (SC.167) implies that

$$\begin{aligned} & n^{-1}\sum_{i=1}^{n} u_{2,i}k_{1,i}(\hat{g}_{1,i}(\hat{\beta}_k) - \hat{g}_{1,i}(\beta_{k,0})) \\ = \; & n^{-1}\sum_{i=1}^{n} u_{2,i}k_{1,i}(g_1(\nu_{1,i}(\hat{\beta}_k);\hat{\beta}_k) - g_1(\nu_{1,i}(\beta_{k,0});\beta_{k,0})) \\ & +(\hat{\beta}_k - \beta_{k,0})o_p(1) + o_p(n^{-1/2}). \end{aligned} \quad \text{(SC.168)}$$

Therefore by Assumptions SC1(i), SC2(ii), (SC.108) and the consistency of $\hat{\beta}_k$,

$$n^{-1}\sum_{i=1}^{n} u_{2,i}k_{1,i}(g_1(\nu_{1,i}(\hat{\beta}_k);\hat{\beta}_k) - g_1(\nu_{1,i}(\beta_{k,0});\beta_{k,0})) = (\hat{\beta}_k - \beta_{k,0})o_p(1)$$

which together with (SC.168) proves the claim of the lemma. *Q.E.D.*

Lemma SC20. *Under Assumptions SC1, SC2 and SC3, we have*

$$\hat{\beta}_k - \beta_{k,0} = O_p((m_1^{1/2} + m_2)n^{-1/2}). \quad \text{(SC.169)}$$

Proof of Lemma SC20. Recall that $\hat{J}_i(\beta_k) \equiv \hat{\ell}_i(\beta_k)\left(k_{2,i} - k_{1,i}\hat{g}_1(\hat{\nu}_{1,i}(\beta_k);\beta_k)\right)$ for any $\beta_k \in \Theta_k$. The first order condition of $\hat{\beta}_k$, i.e. (SB.12), can be written as

$$n^{-1}\sum_{i=1}^{n}(\hat{J}_i(\beta_{k,0}) - \hat{J}_i(\hat{\beta}_k)) = n^{-1}\sum_{i=1}^{n}\hat{J}_i(\beta_{k,0}) \quad \text{(SC.170)}$$

where by Lemma SB2 and (SB.41)

$$n^{-1}\sum_{i=1}^{n}\hat{J}_i(\beta_{k,0}) = O_p(n^{-1/2}). \quad \text{(SC.171)}$$

Using Assumption SC1(iii), Lemma SC17, Lemma SC18 and Lemma SC19, we can use the decomposition in (SB.38) to deduce that

$$n^{-1}\sum_{i=1}^{n}(\hat{J}_i(\hat{\beta}_k) - \hat{J}_i(\beta_{k,0})) = -(\hat{\beta}_k - \beta_{k,0})\left(\mathbb{E}[\varsigma_{1,i}^2] + o_p(1)\right) + O_p((m_1^{1/2} + m_2)n^{-1/2}). \quad \text{(SC.172)}$$

The claim of the lemma follows from (SB.14), (SC.170), (SC.171) and (SC.172). *Q.E.D.*

Lemma SC21. *Under Assumptions SC1, SC2 and SC3, we have*

$$\tilde{\beta}_{g,*} - \hat{\beta}_g(\beta_{k,0}) = O_p(\xi_{1,m_2}(m_2 + m_1^{1/2})n^{-1/2}) = o_p(1)$$

where $\tilde{\beta}_{g,*} \equiv (B(\beta_{k,0})')^{-1}B(\hat{\beta}_k)'\hat{\beta}_g(\hat{\beta}_k)$.

PROOF OF LEMMA SC21. By the definition of $\tilde{\beta}_{g,*}$, we can write

$$\tilde{\beta}_{g,*} = (\hat{\mathbf{P}}_2^*(\hat{\beta}_k)'\hat{\mathbf{P}}_2^*(\hat{\beta}_k))^{-1}\hat{\mathbf{P}}_2^*(\hat{\beta}_k)'\hat{\mathbf{Y}}_2^*(\hat{\beta}_k)$$

where $\hat{\mathbf{P}}_2^*(\beta_k) \equiv (\hat{P}_{2,1}^*(\beta_k), \ldots, \hat{P}_{2,n}^*(\beta_k))'$ and $\hat{P}_{2,i}^*(\beta_k) \equiv B(\beta_{k,0})P_2\left(\hat{\nu}_{1,i}(\beta_k)\right)$. Therefore we have the following decomposition

$$\begin{aligned}
\tilde{\beta}_{g,*} - \hat{\beta}_g(\beta_{k,0}) &= \left[(\hat{\mathbf{P}}_2^*(\hat{\beta}_k)'\hat{\mathbf{P}}_2^*(\hat{\beta}_k))^{-1} - (\hat{\mathbf{P}}_2(\beta_{k,0})'\hat{\mathbf{P}}_2(\beta_{k,0}))^{-1}\right]\hat{\mathbf{P}}_2^*(\hat{\beta}_k)'\hat{\mathbf{Y}}_2^*(\hat{\beta}_k) \\
&\quad + (\hat{\mathbf{P}}_2(\beta_{k,0})'\hat{\mathbf{P}}_2(\beta_{k,0}))^{-1}(\hat{\mathbf{P}}_2^*(\hat{\beta}_k) - \hat{\mathbf{P}}_2(\beta_{k,0}))'\hat{\mathbf{Y}}_2^*(\beta_k) \\
&\quad + (\hat{\mathbf{P}}_2(\beta_{k,0})'\hat{\mathbf{P}}_2(\beta_{k,0}))^{-1}\hat{\mathbf{P}}_2(\beta_{k,0})'(\hat{\mathbf{Y}}_2^*(\hat{\beta}_k) - \hat{\mathbf{Y}}_2^*(\beta_{k,0})). \qquad \text{(SC.173)}
\end{aligned}$$

By the Markov inequality, Assumptions SC1(i) and SC2(i), and (SC.66),

$$\left\|(\hat{\mathbf{P}}_2(\beta_{k,0})'\hat{\mathbf{P}}_2(\beta_{k,0}))^{-1}\hat{\mathbf{P}}_2(\beta_{k,0})'\mathbf{K}_2\right\|^2 \leq \frac{n^{-1}\sum_{i=1}^n k_{2,i}^2}{\lambda_{\min}(n^{-1}\hat{\mathbf{P}}_2(\beta_{k,0})'\hat{\mathbf{P}}_2(\beta_{k,0}))} = O_p(1). \qquad \text{(SC.174)}$$

Since $\hat{\mathbf{Y}}_2^*(\hat{\beta}_k) - \hat{\mathbf{Y}}_2^*(\beta_{k,0}) = -(\hat{\beta}_k - \beta_{k,0})\mathbf{K}_2$, by Lemma SC20 and (SC.174) we get

$$(\hat{\mathbf{P}}_2(\beta_{k,0})'\hat{\mathbf{P}}_2(\beta_{k,0}))^{-1}\hat{\mathbf{P}}_2(\beta_{k,0})'(\hat{\mathbf{Y}}_2^*(\hat{\beta}_k) - \hat{\mathbf{Y}}_2^*(\beta_{k,0})) = O_p((m_2 + m_1^{1/2})n^{-1/2}). \qquad \text{(SC.175)}$$

By the mean value expansion, we have for any $b \in \mathbb{R}^{m_2}$,

$$b'(\hat{P}_{2,i}^*(\hat{\beta}_k) - \hat{P}_{2,i}(\beta_{k,0})) = -b'\partial^1\tilde{P}_2(\hat{\nu}_{1,i}(\tilde{\beta}_k);\beta_{k,0})k_{1,i}(\hat{\beta}_k - \beta_{k,0}) \qquad \text{(SC.176)}$$

where $\tilde{\beta}_k$ lies between $\hat{\beta}_k$ and $\beta_{k,0}$. By Assumption SC3(iv), Lemma SC4 and Lemma SC20, $\hat{\nu}_{1,i}(\tilde{\beta}_k) \in \Omega_\varepsilon(\beta_{k,0})$ for any $i = 1, \ldots, n$ wpa1. Therefore by the Cauchy-Schwarz inequality, Assumption SC2(v) and (SC.176)

$$\left|b'(\hat{P}_{2,i}^*(\hat{\beta}_k) - \hat{P}_{2,i}(\beta_{k,0}))\right| \leq \|b\|\,\xi_{1,m_2}\left|k_{1,i}(\hat{\beta}_k - \beta_{k,0})\right|$$

wpa1. Therefore we have wpa1,

$$\begin{aligned}
&b'(\hat{\mathbf{P}}_2^*(\hat{\beta}_k) - \hat{\mathbf{P}}_2(\beta_{k,0}))'(\hat{\mathbf{P}}_2^*(\hat{\beta}_k) - \hat{\mathbf{P}}_2(\beta_{k,0}))b \\
&= \sum_{i=1}^n (b'(\hat{P}_{2,i}^*(\hat{\beta}_k) - \hat{P}_{2,i}(\beta_{k,0}))^2 \leq \|b\|^2\,\xi_{1,m_2}^2(\hat{\beta}_k - \beta_{k,0})^2\sum_{i=1}^n k_{1,i}^2
\end{aligned}$$

which together with Lemma SC20 implies that

$$||\hat{\mathbf{P}}_2^*(\hat{\beta}_k) - \hat{\mathbf{P}}_2(\beta_{k,0})||_S = |\hat{\beta}_k - \beta_{k,0}|O_p(\xi_{1,m_2}n^{1/2}) = O_p(\xi_{1,m_2}(m_2 + m_1^{1/2})n^{-1/2}). \quad \text{(SC.177)}$$

Since $y_{2,i}^*(\beta_k) = y_{2,i}^* - \beta_k k_{2,i}$, by the Cauchy-Schwarz inequality we get

$$n^{-1}\sum_{i=1}^{n}(y_{2,i}^*(\beta_k))^2 \leq 8\left(n^{-1}\sum_{i=1}^{n}(y_{2,i}^*)^2 + \beta_k^2 n^{-1}\sum_{i=1}^{n}k_{2,i}^2\right)$$

which together with the Markov inequality, Assumptions SC1(i) and SC2(i), and the compactness of Θ_k implies that

$$\sup_{\beta_k\in\Theta_k} n^{-1}\sum_{i=1}^{n}(y_{2,i}^*(\beta_k))^2 = O_p(1). \quad \text{(SC.178)}$$

By the Cauchy-Schwarz inequality, (SC.66), (SC.177) and (SC.178),

$$\begin{aligned}
&\left\|(\hat{\mathbf{P}}_2(\beta_{k,0})'\hat{\mathbf{P}}_2(\beta_{k,0}))^{-1}(\hat{\mathbf{P}}_2^*(\hat{\beta}_k) - \hat{\mathbf{P}}_2(\beta_{k,0}))'\hat{\mathbf{Y}}_2^*(\beta_k)\right\| \\
&\leq \frac{n^{-1}||\hat{\mathbf{P}}_2^*(\hat{\beta}_k) - \hat{\mathbf{P}}_2(\beta_{k,0})||_S\left\|\hat{\mathbf{Y}}_2^*(\hat{\beta}_k)\right\|}{\lambda_{\min}(n^{-1}\hat{\mathbf{P}}_2(\beta_{k,0})'\hat{\mathbf{P}}_2(\beta_{k,0}))} = O_p(\xi_{1,m_2}(m_2 + m_1^{1/2})n^{-1/2}). \quad \text{(SC.179)}
\end{aligned}$$

By the definition of $\hat{\beta}_g(\hat{\beta}_k)$, we can write

$$\begin{aligned}
&[(\hat{\mathbf{P}}_2^*(\hat{\beta}_k)'\hat{\mathbf{P}}_2^*(\hat{\beta}_k))^{-1} - (\hat{\mathbf{P}}_2(\beta_{k,0})'\hat{\mathbf{P}}_2(\beta_{k,0}))^{-1}]\hat{\mathbf{P}}_2^*(\hat{\beta}_k)'\hat{\mathbf{Y}}_2^*(\hat{\beta}_k) \\
= \;& (\hat{\mathbf{P}}_2(\beta_{k,0})'\hat{\mathbf{P}}_2(\beta_{k,0}))^{-1}(\hat{\mathbf{P}}_2(\beta_{k,0}) - \hat{\mathbf{P}}_2^*(\hat{\beta}_k))'\hat{\mathbf{P}}_2(\hat{\beta}_k)\hat{\beta}_g(\hat{\beta}_k) \\
&+(\hat{\mathbf{P}}_2(\beta_{k,0})'\hat{\mathbf{P}}_2(\beta_{k,0}))^{-1}\hat{\mathbf{P}}_2(\beta_{k,0})'(\hat{\mathbf{P}}_2(\beta_{k,0}) - \hat{\mathbf{P}}_2^*(\hat{\beta}_k))\tilde{\beta}_{g,*}. \quad \text{(SC.180)}
\end{aligned}$$

By the Cauchy-Schwarz inequality, (SC.66), (SC.78) and (SC.177),

$$\begin{aligned}
&\left\|(\hat{\mathbf{P}}_2(\beta_{k,0})'\hat{\mathbf{P}}_2(\beta_{k,0}))^{-1}(\hat{\mathbf{P}}_2(\beta_{k,0}) - \hat{\mathbf{P}}_2^*(\hat{\beta}_k))'\hat{\mathbf{P}}_2(\hat{\beta}_k)\hat{\beta}_g(\hat{\beta}_k)\right\| \\
\leq \;& \frac{n^{-1}||\hat{\mathbf{P}}_2^*(\hat{\beta}_k) - \hat{\mathbf{P}}_2(\beta_{k,0})||_S\left\|\hat{\mathbf{P}}_2(\hat{\beta}_k)\hat{\beta}_g(\hat{\beta}_k)\right\|}{\lambda_{\min}(n^{-1}\hat{\mathbf{P}}_2(\beta_{k,0})'\hat{\mathbf{P}}_2(\beta_{k,0}))} = O_p(\xi_{1,m_2}(m_2 + m_1^{1/2})n^{-1/2}). \text{(SC.181)}
\end{aligned}$$

By the definition of $\tilde{\beta}_{g,*}$, and the mean value expansion

$$\begin{aligned}
\left\|(\hat{\mathbf{P}}_2(\beta_{k,0}) - \hat{\mathbf{P}}_2^*(\hat{\beta}_k))\tilde{\beta}_{g,*}\right\|^2 &= \sum_{i=1}^{n}(\hat{\beta}_g(\hat{\beta}_k)'(\tilde{P}_2(\hat{\nu}_{1,i}(\beta_{k,0});\hat{\beta}_k) - \tilde{P}_2(\hat{\nu}_{1,i}(\hat{\beta}_k);\hat{\beta}_k)))^2 \\
&= (\hat{\beta}_k - \beta_{k,0})^2\sum_{i=1}^{n}k_{1,i}^2(\hat{\beta}_g(\hat{\beta}_k)'\partial^1\tilde{P}_2(\tilde{\nu}_{1,i}(\tilde{\beta}_k);\hat{\beta}_k))^2 \quad \text{(SC.182)}
\end{aligned}$$

where $\tilde{\beta}_k$ lies between $\hat{\beta}_k$ and $\beta_{k,0}$. By Assumption SC3(iv), Lemma SC4 and Lemma SC20, $\hat{\nu}_{1,i}(\tilde{\beta}_k) \in \Omega_{\varepsilon_n}(\hat{\beta}_k)$ for any $i = 1, \ldots, n$ wpa1. By the Cauchy-Schwarz inequality, Assumption SC2(v), Lemma SC20, (SC.78) and (SC.182)

$$n^{-1/2}(\hat{\mathbf{P}}_2(\beta_{k,0}) - \hat{\mathbf{P}}_2^*(\hat{\beta}_k))\tilde{\beta}_{g,*} = (\hat{\beta}_k - \beta_{k,0})O_p(\xi_{1,m_2}) = O_p(\xi_{1,m_2}(m_2 + m_1^{1/2})n^{-1/2}) \quad \text{(SC.183)}$$

which together with (SC.66) implies that

$$(\hat{\mathbf{P}}_2(\beta_{k,0})'\hat{\mathbf{P}}_2(\beta_{k,0}))^{-1}\hat{\mathbf{P}}_2(\beta_{k,0})'(\hat{\mathbf{P}}_2(\beta_{k,0}) - \hat{\mathbf{P}}_2^*(\hat{\beta}_k))\tilde{\beta}_{g,*} = O_p(\xi_{1,m_2}(m_2 + m_1^{1/2})n^{-1/2}). \quad \text{(SC.184)}$$

Combining the results in (SC.180), (SC.181) and (SC.184) we get

$$[(\hat{\mathbf{P}}_2^*(\hat{\beta}_k)\hat{\mathbf{P}}_2^*(\hat{\beta}_k)')^{-1} - (\hat{\mathbf{P}}_2(\beta_{k,0})'\hat{\mathbf{P}}_2(\beta_{k,0}))^{-1}]\hat{\mathbf{P}}_2^*(\hat{\beta}_k)'\hat{\mathbf{Y}}_2^*(\hat{\beta}_k) = O_p(\xi_{1,m_2}(m_2 + m_1^{1/2})n^{-1/2})$$

which together with Assumption SC3(iv), (SC.173), (SC.175) and (SC.179) proves the lemma. *Q.E.D.*

Lemma SC22. *Let* $\mathbf{U}_2 = (u_{2,1}, \ldots, u_{2,n})'$, $\hat{\mathbf{G}}_n = (\hat{g}(\hat{\nu}_{1,1}(\hat{\beta}_k); \hat{\beta}_k), \ldots, \hat{g}(\hat{\nu}_{1,n}(\hat{\beta}_k); \hat{\beta}_k))'$ *and* $\mathbf{G}_n = (g(\nu_{1,1}), \ldots, g(\nu_{1,n}))'$. *Then under Assumptions SC1, SC2 and SC3, we have*

(i) $n^{-1}\mathbf{U}_2'(\hat{\mathbf{P}}_2^*(\hat{\beta}_k) - \hat{\mathbf{P}}_2(\beta_{k,0}))\hat{\beta}_{\varphi}(\beta_{k,0}) = (\hat{\beta}_k - \beta_{k,0})o_p(1) + o_p(n^{-1/2})$

(ii) $n^{-1}\mathbf{L}_2'(\hat{\mathbf{P}}_2^*(\hat{\beta}_k) - \hat{\mathbf{P}}_2(\beta_{k,0}))\hat{\beta}_{\varphi}(\beta_{k,0}) = o_p(1)$;

(iii) $n^{-1}\mathbf{K}_2'(\hat{\mathbf{P}}_2^*(\hat{\beta}_k) - \hat{\mathbf{P}}_2(\beta_{k,0}))\hat{\beta}_{\varphi}(\beta_{k,0}) = o_p(1)$;

(iv) $n^{-1}(\hat{\mathbf{G}}_n - \mathbf{G}_n)'(\hat{\mathbf{P}}_2^*(\hat{\beta}_k) - \hat{\mathbf{P}}_2(\beta_{k,0}))\hat{\beta}_{\varphi}(\beta_{k,0}) = (\hat{\beta}_k - \beta_{k,0})o_p(1)$.

Proof of Lemma SC22. (i) First note that

$$\begin{aligned}
& n^{-1}\mathbf{U}_2'(\hat{\mathbf{P}}_2^*(\hat{\beta}_k) - \hat{\mathbf{P}}_2(\beta_{k,0}))\hat{\beta}_{\varphi}(\beta_{k,0}) \\
= \; & n^{-1}\sum_{i=1}^n u_{2,i}(\tilde{P}_2(\hat{\nu}_{1,i}(\hat{\beta}_k); \beta_{k,0}) - \tilde{P}_2(\nu_{1,i}(\hat{\beta}_k); \beta_{k,0}))'\hat{\beta}_{\varphi}(\beta_{k,0}) \\
& -n^{-1}\sum_{i=1}^n u_{2,i}\left(\tilde{P}_2(\hat{\nu}_{1,i}(\beta_{k,0}); \beta_{k,0}) - \tilde{P}_2(\nu_{1,i}(\beta_{k,0}); \beta_{k,0})\right)'\hat{\beta}_{\varphi}(\beta_{k,0}) \\
& +n^{-1}\sum_{i=1}^n u_{2,i}(\tilde{P}_2(\nu_{1,i}(\hat{\beta}_k); \beta_{k,0}) - \tilde{P}_2(\nu_{1,i}(\beta_{k,0}); \beta_{k,0}))'\hat{\beta}_{\varphi}(\beta_{k,0}). \quad \text{(SC.185)}
\end{aligned}$$

By the second order expansion,

$$
\begin{aligned}
& n^{-1}\sum_{i=1}^{n}u_{2,i}(\tilde{P}_2(\hat{\nu}_{1,i}(\hat{\beta}_k);\beta_{k,0})-\tilde{P}_2(\nu_{1,i}(\hat{\beta}_k);\beta_{k,0}))'\hat{\beta}_{\varphi}(\beta_{k,0}) \\
= \;& n^{-1}\sum_{i=1}^{n}u_{2,i}(\hat{\phi}_i-\phi_i)\partial^1\tilde{P}_2(\nu_{1,i}(\hat{\beta}_k);\beta_{k,0})'\hat{\beta}_{\varphi}(\beta_{k,0}) \\
& +n^{-1}\sum_{i=1}^{n}u_{2,i}(\hat{\phi}_i-\phi_i)^2\partial^2\tilde{P}_2(\tilde{\nu}_{1,i}(\hat{\beta}_k);\beta_{k,0})'\hat{\beta}_{\varphi}(\beta_{k,0}) \qquad \text{(SC.186)}
\end{aligned}
$$

where $\tilde{\nu}_{1,i}(\hat{\beta}_k)$ is between $\hat{\nu}_{1,i}(\hat{\beta}_k)$ and $\nu_{1,i}(\hat{\beta}_k)$. By Assumption SC3(iv), Lemma SC4 and Lemma SC20, both $\hat{\nu}_{1,i}(\hat{\beta}_k)$ and $\nu_{1,i}(\hat{\beta}_k)$ are in $\Omega_{\varepsilon}(\beta_{k,0})$ for any $i=1,\ldots,n$ wpa1. Therefore $\tilde{\nu}_{1,i}(\hat{\beta}_k)\in\Omega_{\varepsilon}(\beta_{k,0})$ for any $i=1,\ldots,n$ wpa1. By the triangle inequality, the Cauchy-Schwarz inequality, Assumptions SC2(v) and SC3(iv), (SC.92) and (SC.127)

$$
\begin{aligned}
& \left|n^{-1}\sum_{i=1}^{n}u_{2,i}(\hat{\phi}_i-\phi_i)^2\partial^2\tilde{P}_2(\tilde{\nu}_{1,i}(\hat{\beta}_k);\beta_{k,0})'\hat{\beta}_{\varphi}(\beta_{k,0})\right| \\
\leq \;& O_p(\xi_{2,m_2})n^{-1}\sum_{i=1}^{n}|u_{2,i}|\,(\hat{\phi}_i-\phi_i)^2=O_p(\xi_{2,m_2}m_1n^{-1})=o_p(n^{-1/2}). \qquad \text{(SC.187)}
\end{aligned}
$$

Using similar arguments for proving (SC.165), we can show that

$$
n^{-1}\sum_{i=1}^{n}u_{2,i}(\hat{\phi}_i-\phi_i)\partial^1\tilde{P}_2(\nu_{1,i}(\hat{\beta}_k);\beta_{k,0})'\hat{\beta}_{\varphi}(\beta_{k,0})=o_p(n^{-1/2})
$$

which together with (SC.186) and (SC.187) implies that

$$
n^{-1}\sum_{i=1}^{n}u_{2,i}(\tilde{P}_2(\hat{\nu}_{1,i}(\hat{\beta}_k);\beta_{k,0})-\tilde{P}_2(\nu_{1,i}(\hat{\beta}_k);\beta_{k,0}))'\hat{\beta}_{\varphi}(\beta_{k,0})=o_p(n^{-1/2}). \qquad \text{(SC.188)}
$$

Similarly, we can show that

$$
n^{-1}\sum_{i=1}^{n}u_{2,i}\left(\tilde{P}_2(\hat{\nu}_{1,i}(\beta_{k,0});\beta_{k,0})-\tilde{P}_2(\nu_{1,i}(\beta_{k,0});\beta_{k,0})\right)'\hat{\beta}_{\varphi}(\beta_{k,0})=o_p(n^{-1/2}). \qquad \text{(SC.189)}
$$

By the third order expansion,

$$\begin{aligned} & n^{-1}\sum_{i=1}^{n}u_{2,i}(\tilde{P}_2(\nu_{1,i}(\hat{\beta}_k);\beta_{k,0})-\tilde{P}_2(\nu_{1,i}(\beta_{k,0});\beta_{k,0}))'\hat{\beta}_{\varphi}(\beta_{k,0}) \\ = \; & (\hat{\beta}_k-\beta_{k,0})n^{-1}\sum_{i=1}^{n}u_{2,i}\partial^1\tilde{P}_{2,i}(\beta_{k,0})'\hat{\beta}_{\varphi}(\beta_{k,0}) \\ & +(\hat{\beta}_k-\beta_{k,0})^2n^{-1}\sum_{i=1}^{n}u_{2,i}\partial^2\tilde{P}_{2,i}(\beta_{k,0})'\hat{\beta}_{\varphi}(\beta_{k,0}) \\ & +(\hat{\beta}_k-\beta_{k,0})^3n^{-1}\sum_{i=1}^{n}u_{2,i}\partial^3\tilde{P}_2(\nu_{1,i}(\tilde{\beta}_k);\beta_{k,0})'\hat{\beta}_{\varphi}(\beta_{k,0}) \end{aligned} \tag{SC.190}$$

where $\tilde{\beta}_k$ is between $\hat{\beta}_k$ and $\beta_{k,0}$. By Lemma SC20 and Assumption SC3(iv), $\nu_{1,i}(\tilde{\beta}_k)\in\Omega_{\varepsilon}(\beta_{k,0})$ for any $i=1,\ldots,n$ wpa1. Therefore by the triangle inequality and the Cauchy-Schwarz inequality, Assumptions SC1(i) and SC2(v), Lemma SC20, (SC.88) and (SC.127)

$$\begin{aligned} & (\hat{\beta}_k-\beta_{k,0})^3n^{-1}\sum_{i=1}^{n}u_{2,i}\partial^3\tilde{P}_2(\nu_{1,i}(\tilde{\beta}_k);\beta_{k,0})'\hat{\beta}_{\varphi}(\beta_{k,0}) \\ = \; & (\hat{\beta}_k-\beta_{k,0})O_p(\xi_{3,m_2}(m_1+m_2^2)n^{-1})=(\hat{\beta}_k-\beta_{k,0})o_p(1) \end{aligned} \tag{SC.191}$$

where the second equality is by Assumption SC3(iv). By Assumptions SC1(i), SC2(v) and SC3(iv), Lemma SC20, (SC.88) and (SC.127), we can show that

$$(\hat{\beta}_k-\beta_{k,0})n^{-1}\sum_{i=1}^{n}u_{2,i}\partial^1\tilde{P}_{2,i}(\beta_{k,0})'\hat{\beta}_{\varphi}(\beta_{k,0})=(\hat{\beta}_k-\beta_{k,0})o_p(1) \tag{SC.192}$$

and

$$(\hat{\beta}_k-\beta_{k,0})^2n^{-1}\sum_{i=1}^{n}u_{2,i}\partial^2\tilde{P}_{2,i}(\beta_{k,0})'\hat{\beta}_{\varphi}(\beta_{k,0})=(\hat{\beta}_k-\beta_{k,0})o_p(1). \tag{SC.193}$$

Collecting the results in (SC.190), (SC.191), (SC.192) and (SC.193), we obtain

$$n^{-1}\sum_{i=1}^{n}u_{2,i}(\tilde{P}_2(\nu_{1,i}(\hat{\beta}_k);\beta_{k,0})-\tilde{P}_2(\nu_{1,i}(\beta_{k,0});\beta_{k,0}))'\hat{\beta}_{\varphi}(\beta_{k,0})=(\hat{\beta}_k-\beta_{k,0})o_p(1)$$

which together with (SC.185), (SC.188) and (SC.189) finishes the proof.

(ii) By the mean value expansion,

$$n^{-1}\mathbf{L}_2'(\hat{\mathbf{P}}_2^*(\hat{\beta}_k)-\hat{\mathbf{P}}_2(\beta_{k,0}))\hat{\beta}_{\varphi}(\beta_{k,0})=(\hat{\beta}_k-\beta_{k,0})n^{-1}\sum_{i=1}^{n}l_{2,i}\partial^1\tilde{P}_2(\hat{\nu}_{1,i}(\tilde{\beta}_k);\beta_{k,0})'\hat{\beta}_{\varphi}(\beta_{k,0}) \tag{SC.194}$$

where $\tilde{\beta}_k$ is between $\hat{\beta}_k$ and $\beta_{k,0}$. By Assumption SC3(iv), Lemma SC4 and Lemma SC20, $\hat{\nu}_{1,i}(\tilde{\beta}_k) \in \Omega_\varepsilon(\beta_{k,0})$ for any $i = 1, \ldots, n$ wpa1. By the triangle inequality, the Cauchy-Schwarz inequality, (SC.72) and (SC.127),

$$n^{-1}\sum_{i=1}^{n} l_{2,i}\partial^1 \tilde{P}_2(\nu_{1,i}(\tilde{\beta}_k);\beta_{k,0})'\hat{\beta}_\varphi(\beta_{k,0}) = O_p(\xi_{1,m_2})$$

which together with Assumption SC3(iv) and Lemma SC20 finishes the proof.

(iii) The third claim of the lemma can be proved in the same way as the second one.

(iv) By the mean value expansion,

$$\begin{aligned} & n^{-1}(\hat{\mathbf{G}}_n - \mathbf{G}_n)'(\hat{\mathbf{P}}_2^*(\hat{\beta}_k) - \hat{\mathbf{P}}_2(\beta_{k,0}))\hat{\beta}_\varphi(\beta_{k,0}) \\ = \; & n^{-1}\sum_{i=1}^{n}(\hat{g}_i(\hat{\beta}_k) - g(\nu_{1,i}))(\tilde{P}_2(\hat{\nu}_{1,i}(\hat{\beta}_k);\beta_{k,0}) - \tilde{P}_2(\hat{\nu}_{1,i}(\beta_{k,0});\beta_{k,0}))'\hat{\beta}_\varphi(\beta_{k,0}) \\ = \; & -(\hat{\beta}_k - \beta_{k,0})n^{-1}\sum_{i=1}^{n}(\hat{g}_i(\hat{\beta}_k) - g(\nu_{1,i}))\partial^1\tilde{P}_2(\hat{\nu}_{1,i}(\tilde{\beta}_k);\beta_{k,0})'\hat{\beta}_\varphi(\beta_{k,0}) \end{aligned} \quad \text{(SC.195)}$$

where $\tilde{\beta}_k$ is between $\hat{\beta}_k$ and $\beta_{k,0}$. By Assumption SC3(iv), Lemma SC4 and Lemma SC20, $\hat{\nu}_{1,i}(\tilde{\beta}_k) \in \Omega_\varepsilon(\beta_{k,0})$ for any $i = 1, \ldots, n$ wpa1. By Assumptions SC2(v) and SC3(iv), Lemma SC15, Lemma SC20 and (SC.127), we get

$$n^{-1}\sum_{i=1}^{n}(\hat{g}_i(\hat{\beta}_k) - g(\nu_{1,i}))\partial^1\tilde{P}_2(\hat{\nu}_{1,i}(\tilde{\beta}_k);\beta_{k,0})'\hat{\beta}_\varphi(\beta_{k,0}) = o_p(1)$$

which together with (SC.195) finishes the proof. *Q.E.D.*

Lemma SC23. *Under Assumptions SC1, SC2 and SC3, we have*

$$\begin{aligned} & n^{-1}\sum_{i=1}^{n}(\hat{g}(\hat{\nu}_{1,i}(\hat{\beta}_k);\hat{\beta}_k) - \hat{g}(\hat{\nu}_{1,i}(\beta_{k,0});\beta_{k,0}))(k_{2,i} - k_{1,i}\hat{g}_{1,i}(\hat{\beta}_k)) \\ & = -(\hat{\beta}_k - \beta_{k,0})\left[\mathbb{E}[k_{1,i}g_{1,i}\varsigma_{1,i}] + \mathbb{E}\left[k_{2,i}(\gamma_{2,i} - \gamma_{1,i}g_{1,i})\right] + o_p(1)\right] + o_p(n^{-1/2}) \end{aligned}$$

where $a_{j,i} \equiv \mathbb{E}[k_{j,i}|\nu_{1,i}]$ *and* $g_{1,i} \equiv g_1(\nu_{1,i})$.

Proof of Lemma SC23. In view of (SC.158), to prove the lemma it is sufficient to show that

$$\begin{aligned} & n^{-1}\sum_{i=1}^{n}(\hat{g}(\hat{\nu}_{1,i}(\hat{\beta}_k);\hat{\beta}_k) - \hat{g}(\hat{\nu}_{1,i}(\beta_{k,0});\beta_{k,0}))(k_{2,i} - k_{1,i}g_{1,i}) \\ = \; & -(\hat{\beta}_k - \beta_{k,0})\left[\mathbb{E}[k_{1,i}g_{1,i}\varsigma_{1,i}] + \mathbb{E}\left[k_{2,i}(\gamma_{2,i} - \gamma_{1,i}g_{1,i})\right] + o_p(1)\right] + o_p(n^{-1/2}). \end{aligned} \quad \text{(SC.196)}$$

By the definition of $\hat{g}(\hat{\nu}_{1,i}(\beta_k);\beta_k)$,

$$\hat{g}(\hat{\nu}_{1,i}(\beta_k);\beta_k) = \hat{P}_{2,i}(\beta_k)'\hat{\beta}_g(\beta_k) = \hat{P}^*_{2,i}(\beta_k)(\hat{\mathbf{P}}^*_2(\beta_k)'\hat{\mathbf{P}}^*_2(\beta_k))^{-1}\hat{\mathbf{P}}^*_2(\beta_k)'\hat{\mathbf{Y}}^*_2(\beta_k)$$

where $\hat{\mathbf{P}}^*_2(\beta_k) \equiv (\hat{P}^*_{2,1}(\beta_k), \ldots, \hat{P}^*_{2,n}(\beta_k))'$ and $\hat{P}^*_{2,i}(\beta_k) \equiv B(\beta_{k,0})P_2(\hat{\nu}_{1,i}(\beta_k))$. Therefore we obtain the following decomposition

$$\begin{aligned}
&\hat{g}(\hat{\nu}_{1,i}(\hat{\beta}_k);\hat{\beta}_k) - \hat{g}(\hat{\nu}_{1,i}(\beta_{k,0});\beta_{k,0}) \\
&= (\hat{P}^*_{2,i}(\hat{\beta}_k) - \hat{P}_{2,i}(\beta_{k,0}))'(\hat{\mathbf{P}}^*_2(\hat{\beta}_k)'\hat{\mathbf{P}}^*_2(\hat{\beta}_k))^{-1}\hat{\mathbf{P}}^*_2(\hat{\beta}_k)'\hat{\mathbf{Y}}^*_2(\hat{\beta}_k) \\
&\quad + \hat{P}_{2,i}(\beta_{k,0})'\left[(\hat{\mathbf{P}}^*_2(\hat{\beta}_k)'\hat{\mathbf{P}}^*_2(\hat{\beta}_k))^{-1} - (\hat{\mathbf{P}}_2(\beta_{k,0})'\hat{\mathbf{P}}_2(\beta_{k,0}))^{-1}\right]\hat{\mathbf{P}}^*_2(\hat{\beta}_k)'\hat{\mathbf{Y}}^*_2(\hat{\beta}_k) \\
&\quad + \hat{P}_{2,i}(\beta_{k,0})'(\hat{\mathbf{P}}_2(\beta_{k,0})'\hat{\mathbf{P}}_2(\beta_{k,0}))^{-1}(\hat{\mathbf{P}}^*_2(\hat{\beta}_k) - \hat{\mathbf{P}}_2(\beta_{k,0}))'\hat{\mathbf{Y}}^*_2(\beta_k) \\
&\quad + \hat{P}_{2,i}(\beta_{k,0})'(\hat{\mathbf{P}}_2(\beta_{k,0})'\hat{\mathbf{P}}_2(\beta_{k,0}))^{-1}\hat{\mathbf{P}}_2(\beta_{k,0})'(\hat{\mathbf{Y}}^*_2(\hat{\beta}_k) - \mathbf{Y}^*_2(\beta_{k,0})). \qquad \text{(SC.197)}
\end{aligned}$$

The proof is divided into 4 steps. The claim in (SC.196) follows from the results in (SC.198), (SC.212), (SC.225) and (SC.227).

Step 1. In this step, we show that

$$\begin{aligned}
& n^{-1}\sum_{i=1}^n \tilde{\beta}'_{g,*}(\hat{P}^*_{2,i}(\hat{\beta}_k) - \hat{P}_{2,i}(\beta_{k,0}))(k_{2,i} - k_{1,i}g_{1,i}) \\
= \ & -(\hat{\beta}_k - \beta_{k,0})\left(\mathbb{E}[k_{1,i}g_{1,i}(k_{2,i} - k_{1,i}g_{1,i})] + o_p(1)\right) + o_p(n^{-1/2}). \qquad \text{(SC.198)}
\end{aligned}$$

where $\tilde{\beta}_{g,*} \equiv (\hat{\mathbf{P}}^*_2(\hat{\beta}_k)'\hat{\mathbf{P}}^*_2(\hat{\beta}_k))^{-1}\hat{\mathbf{P}}^*_2(\hat{\beta}_k)'\hat{\mathbf{Y}}^*_2(\hat{\beta}_k)$.

By Lemma SC21 and (SC.78)

$$\left\|\tilde{\beta}_{g,*}\right\| = O_p(1). \qquad \text{(SC.199)}$$

By the second order expansion,

$$\begin{aligned}
& \tilde{\beta}'_{g,*}\left(\hat{P}^*_{2,i}(\hat{\beta}_k) - \tilde{P}_{2,i}(\beta_{k,0}) - \partial^1\tilde{P}_{2,i}(\beta_{k,0})(\hat{\nu}_{1,i}(\hat{\beta}_k) - \nu_{1,i}(\beta_{k,0}))\right) \\
= \ & \tilde{\beta}'_{g,*}\partial^2\tilde{P}_2(\tilde{\nu}_{1,i};\beta_{k,0})(\hat{\nu}_{1,i}(\hat{\beta}_k) - \nu_{1,i}(\beta_{k,0}))^2 \qquad \text{(SC.200)}
\end{aligned}$$

where $\tilde{\nu}_{1,i}$ lies between $\hat{\nu}_{1,i}(\hat{\beta}_k)$ and $\nu_{1,i}(\beta_{k,0})$. By Assumption SC3(iv), Lemma SC4 and Lemma SC20, $\tilde{\nu}_{1,i} \in \Omega_\varepsilon(\beta_{k,0})$ for any $i = 1, \ldots, n$ wpa1. Since $\hat{\nu}_{1,i}(\hat{\beta}_k) - \nu_{1,i}(\beta_{k,0}) = (\hat{\phi}_i - \phi_i) - k_{1,i}(\hat{\beta}_k - \beta_{k,0})$, by Assumption SC2(v), (SC.199) and (SC.200) we have

$$
\begin{aligned}
&\left| n^{-1}\sum_{i=1}^{n}\tilde{\beta}_{g,*}'\begin{pmatrix} \hat{P}_{2,i}^{*}(\hat{\beta}_k)-\tilde{P}_{2,i}(\beta_{k,0}) \\ -\partial^1\tilde{P}_{2,i}(\beta_{k,0})(\hat{\nu}_{1,i}(\hat{\beta}_k)-\nu_{1,i}(\beta_{k,0})) \end{pmatrix}(k_{2,i}-k_{1,i}g_{1,i})\right| \\
&\leq O_p(\xi_{2,m_2})\left(n^{-1}\sum_{i=1}^{n}(\hat{\phi}_i-\phi_i)^2+(\hat{\beta}_k-\beta_{k,0})^2 n^{-1}\sum_{i=1}^{n}k_{1,i}^2\right) \\
&=(\hat{\beta}_k-\beta_{k,0})O_p(\xi_{2,m_2}(m_1^{1/2}+m_2)n^{-1/2})+O_p(\xi_{2,m_2}m_1 n^{-1}) \\
&=(\hat{\beta}_k-\beta_{k,0})o_p(1)+o_p(n^{-1/2}) \qquad \text{(SC.201)}
\end{aligned}
$$

where the first equality is by Lemma SC4 and Lemma SC20, and the second equality is by Assumption SC3(iv). Similarly, we can show that

$$
n^{-1}\sum_{i=1}^{n}\tilde{\beta}_{g,*}'\begin{pmatrix} \hat{P}_{2,i}(\beta_{k,0})-\tilde{P}_{2,i}(\beta_{k,0}) \\ -\partial^1\tilde{P}_{2,i}(\beta_{k,0})(\hat{\nu}_{1,i}(\beta_{k,0})-\nu_{1,i}(\beta_{k,0}) \end{pmatrix}(k_{2,i}-k_{1,i}g_{1,i})=o_p(n^{-1/2}). \qquad \text{(SC.202)}
$$

Since $\hat{\nu}_{1,i}(\hat{\beta}_k)-\hat{\nu}_{1,i}(\beta_{k,0})=-k_{1,i}(\hat{\beta}_k-\beta_{k,0})$, using (SC.201) and (SC.202) we get

$$
\begin{aligned}
& n^{-1}\sum_{i=1}^{n}\tilde{\beta}_{g,*}'(\hat{P}_{2,i}^{*}(\hat{\beta}_k)-\hat{P}_{2,i}(\beta_{k,0}))(k_{2,i}-k_{1,i}g_{1,i}) \qquad \text{(SC.203)} \\
= \;& -(\hat{\beta}_k-\beta_{k,0})n^{-1}\sum_{i=1}^{n}\tilde{\beta}_{g,*}'\partial^1\tilde{P}_{2,i}(\beta_{k,0})k_{1,i}(k_{2,i}-k_{1,i}g_{1,i})+(\hat{\beta}_k-\beta_{k,0})o_p(1)+o_p(n^{-1/2})
\end{aligned}
$$

By the definition of $\tilde{\beta}_{g,*}$, we can write $\tilde{\beta}_{g,*}'\partial^1\tilde{P}_{2,i}(\beta_{k,0})=\hat{\beta}_g(\hat{\beta}_k)'\partial^1\tilde{P}_2(\nu_{1,i}(\beta_{k,0});\hat{\beta}_k)$. Therefore

$$
\begin{aligned}
& n^{-1}\sum_{i=1}^{n}\tilde{\beta}_{g,*}'\partial^1\tilde{P}_{2,i}(\beta_{k,0})k_{1,i}(k_{2,i}-k_{1,i}g_{1,i}) \\
= \;& \mathbb{E}[g_{1,i}k_{1,i}(k_{2,i}-k_{1,i}g_{1,i})] \\
& +n^{-1}\sum_{i=1}^{n}(g_{1,i}k_{1,i}(k_{2,i}-k_{1,i}g_{1,i})-\mathbb{E}[g_{1,i}k_{1,i}(k_{2,i}-k_{1,i}g_{1,i})]) \\
& +n^{-1}\sum_{i=1}^{n}(\hat{\beta}_g(\hat{\beta}_k)'\partial^1\tilde{P}_2(\nu_{1,i}(\beta_{k,0});\hat{\beta}_k)-g_{1,i})k_{1,i}(k_{2,i}-k_{1,i}g_{1,i}). \qquad \text{(SC.204)}
\end{aligned}
$$

By Assumption SC3(iv), Lemma SC14 and Lemma SC20

$$
\hat{\beta}_g(\hat{\beta}_k)-\tilde{\beta}_{g,m_2}(\hat{\beta}_k)=O_p((m_2+m_1^{1/2})n^{-1/2}). \qquad \text{(SC.205)}
$$

By the Markov inequality, Assumptions SC1(i) and SC2(i, ii, iii, v)

$$n^{-1}\sum_{i=1}^{n}k_{1,i}^{2}(k_{2,i}-k_{1,i}g_{1,i})^{2}=O_p(1). \tag{SC.206}$$

By the mean value expansion,

$$\begin{aligned}
& n^{-1}\sum_{i=1}^{n}\hat{\beta}_g(\hat{\beta}_k)'(\partial^1\tilde{P}_2(\nu_{1,i}(\beta_{k,0});\hat{\beta}_k)-\partial^1\tilde{P}_2(\nu_{1,i}(\hat{\beta}_k);\hat{\beta}_k))k_{1,i}(k_{2,i}-k_{1,i}g_{1,i}) \\
= \;& (\hat{\beta}_k-\beta_{k,0})n^{-1}\sum_{i=1}^{n}\hat{\beta}_g(\hat{\beta}_k)'\partial^2\tilde{P}_2(\nu_{1,i}(\tilde{\beta}_k);\hat{\beta}_k)k_{1,i}(k_{2,i}-k_{1,i}g_{1,i}) \qquad \text{(SC.207)}
\end{aligned}$$

where $\tilde{\beta}_k$ is between $\hat{\beta}_k$ and $\beta_{k,0}$. By Lemma SC20 and Assumption SC3(iv), $\nu_{1,i}(\tilde{\beta}_k)\in\Omega_\varepsilon(\hat{\beta}_k)$ for any $i=1,\ldots,n$ wpa1. Therefore by the Cauchy-Schwarz inequality, Assumptions SC2(v) and Assumption SC3(iv), Lemma SC20, (SC.78), (SC.206) and (SC.207),

$$n^{-1}\sum_{i=1}^{n}\hat{\beta}_g(\hat{\beta}_k)'(\partial^1\tilde{P}_2(\nu_{1,i}(\beta_{k,0});\hat{\beta}_k)-\partial^1\tilde{P}_2(\nu_{1,i}(\hat{\beta}_k);\hat{\beta}_k))(k_{2,i}-k_{1,i}g_{1,i})=o_p(1). \tag{SC.208}$$

By the triangle inequality and the Cauchy-Schwarz inequality, Assumptions SC2(ii, iii, v) and SC3(iv), Lemma SC20, (SC.205) and (SC.206),

$$\begin{aligned}
& \left|n^{-1}\sum_{i=1}^{n}(\hat{\beta}_g(\hat{\beta}_k)'\partial^1\tilde{P}_2(\nu_{1,i}(\hat{\beta}_k);\hat{\beta}_k)-g_{1,i})k_{1,i}(k_{2,i}-k_{1,i}g_{1,i})\right| \\
\leq \;& n^{-1}\sum_{i=1}^{n}\left|(\hat{\beta}_g(\hat{\beta}_k)-\tilde{\beta}_{g,m_2}(\hat{\beta}_k))'\partial^1\tilde{P}_2(\nu_{1,i}(\hat{\beta}_k);\hat{\beta}_k)k_{1,i}(k_{2,i}-k_{1,i}g_{1,i})\right| \\
& +n^{-1}\sum_{i=1}^{n}\left|(\tilde{\beta}_{g,m_2}(\hat{\beta}_k)'\partial^1\tilde{P}_2(\nu_{1,i}(\hat{\beta}_k);\hat{\beta}_k)-g_1(\nu_{1,i}(\hat{\beta}_k);\hat{\beta}_k))k_{1,i}(k_{2,i}-k_{1,i}g_{1,i})\right| \\
& +n^{-1}\sum_{i=1}^{n}\left|(g_1(\nu_{1,i}(\hat{\beta}_k);\hat{\beta}_k)-g_1(\nu_{1,i}(\beta_{k,0});\beta_{k,0}))k_{1,i}(k_{2,i}-k_{1,i}g_{1,i})\right| \\
= \;& O_p(\xi_{1,m_2}(m_2+m_1^{1/2})n^{-1/2})+O_p(m^{-r_g})+(\hat{\beta}_k-\beta_{k,0})O_p(1)=o_p(1)
\end{aligned}$$

which together with (SC.208) implies that

$$n^{-1}\sum_{i=1}^{n}(\hat{\beta}_g(\hat{\beta}_k)'\partial^1\tilde{P}_2(\nu_{1,i}(\beta_{k,0});\hat{\beta}_k)-g_{1,i})k_{1,i}(k_{2,i}-k_{1,i}g_{1,i})=o_p(1). \tag{SC.209}$$

By Assumptions SC1(i) and SC2(i, ii), and the Markov inequality,

$$n^{-1}\sum_{i=1}^{n} g_{1,i}k_{1,i}(k_{2,i}-k_{1,i}g_{1,i})-\mathbb{E}[k_{1,i}g_{1,i}(k_{2,i}-k_{1,i}g_{1,i})]=O_p(n^{-1/2}) \tag{SC.210}$$

which together with (SC.204), (SC.209) and (SC.210) implies that

$$n^{-1}\sum_{i=1}^{n} \tilde{\beta}_{g,*}'\partial^1\tilde{P}_{2,i}(\beta_{k,0})k_{1,i}(k_{2,i}-k_{1,i}g_{1,i})=\mathbb{E}[k_{1,i}g_{1,i}(k_{2,i}-k_{1,i}g_{1,i})]+o_p(1). \tag{SC.211}$$

The claim in (SC.198) follows from (SC.203) and (SC.211).

Step 2. In this step, we show that

$$\begin{aligned}
&\tilde{\beta}_{g,*}'\left[\hat{\mathbf{P}}_2(\beta_{k,0})'\hat{\mathbf{P}}_2(\beta_{k,0})-\hat{\mathbf{P}}_2^*(\hat{\beta}_k)'\hat{\mathbf{P}}_2^*(\hat{\beta}_k)\right]\hat{\beta}_\varphi(\beta_{k,0})\\
&=(\hat{\beta}_k-\beta_{k,0})\left(\mathbb{E}[g_{1,i}k_{1,i}(\gamma_{2,i}-\gamma_{1,i}g_{1,i})]+o_p(1)\right)\\
&\quad+n^{-1}\tilde{\beta}_{g,*}'\hat{\mathbf{P}}_2^*(\hat{\beta}_k)'(\hat{\mathbf{P}}_2(\beta_{k,0})-\hat{\mathbf{P}}_2^*(\hat{\beta}_k))\hat{\beta}_\varphi(\beta_{k,0})+o_p(n^{-1/2})
\end{aligned} \tag{SC.212}$$

where $\hat{\beta}_\varphi(\beta_{k,0})\equiv(\hat{\mathbf{P}}_2(\beta_{k,0})'\hat{\mathbf{P}}_2(\beta_{k,0}))^{-1}\sum_{i=1}^{n}\hat{P}_{2,i}(\beta_{k,0})(k_{2,i}-k_{1,i}g_{1,i})$.

First note that

$$\begin{aligned}
&\tilde{\beta}_{g,*}'\left[\hat{\mathbf{P}}_2(\beta_{k,0})'\hat{\mathbf{P}}_2(\beta_{k,0})-\hat{\mathbf{P}}_2^*(\hat{\beta}_k)'\hat{\mathbf{P}}_2^*(\hat{\beta}_k)\right]\hat{\beta}_\varphi(\beta_{k,0})\\
&=\tilde{\beta}_{g,*}'(\hat{\mathbf{P}}_2(\beta_{k,0})-\hat{\mathbf{P}}_2^*(\hat{\beta}_k))'\hat{\mathbf{P}}_2(\beta_{k,0})\hat{\beta}_\varphi(\beta_{k,0})+\tilde{\beta}_{g,*}'\hat{\mathbf{P}}_2^*(\hat{\beta}_k)'(\hat{\mathbf{P}}_2(\beta_{k,0})-\hat{\mathbf{P}}_2^*(\hat{\beta}_k))\hat{\beta}_\varphi(\beta_{k,0}).
\end{aligned}$$

Therefore to prove (SC.212), it is sufficient to show that

$$n^{-1}\tilde{\beta}_{g,*}'(\hat{\mathbf{P}}_2(\beta_{k,0})-\hat{\mathbf{P}}_2^*(\hat{\beta}_k))'\hat{\mathbf{P}}_2(\beta_{k,0})\hat{\beta}_\varphi(\beta_{k,0})=(\hat{\beta}_k-\beta_{k,0})\left(\mathbb{E}\left[g_{1,i}k_{1,i}\varphi_i\right]+o_p(1)\right)+o_p(n^{-1/2}). \tag{SC.213}$$

where $\varphi_i\equiv\gamma_{2,i}-\gamma_{1,i}g_{1,i}$.

By the Cauchy-Schwarz inequality, Assumption SC3(iv), (SC.66), (SC.126) and (SC.183),

$$\begin{aligned}
&n^{-1}\left|\tilde{\beta}_{g,*}'(\hat{\mathbf{P}}_2(\beta_{k,0})-\hat{\mathbf{P}}_2^*(\hat{\beta}_k))'\hat{\mathbf{P}}_2(\beta_{k,0})(\hat{\beta}_\varphi(\beta_{k,0})-\tilde{\beta}_{\varphi,m_2}(\beta_{k,0}))\right|\\
\leq\;&n^{-1}\left\|\tilde{\beta}_{g,*}'(\hat{\mathbf{P}}_2(\beta_{k,0})-\hat{\mathbf{P}}_2^*(\hat{\beta}_k))\right\|\left\|\hat{\mathbf{P}}_2(\beta_{k,0})(\hat{\beta}_\varphi(\beta_{k,0})-\tilde{\beta}_{\varphi,m_2}(\beta_{k,0}))\right\|\\
\leq\;&\frac{\left\|n^{-1/2}\tilde{\beta}_{g,*}'(\hat{\mathbf{P}}_2(\beta_{k,0})-\hat{\mathbf{P}}_2^*(\hat{\beta}_k))\right\|\left\|\hat{\beta}_\varphi(\beta_{k,0})-\tilde{\beta}_{\varphi,m_2}(\beta_{k,0})\right\|}{(\lambda_{\max}(n^{-1}\hat{\mathbf{P}}_2(\beta_{k,0})'\hat{\mathbf{P}}_2(\beta_{k,0})))^{-1/2}}\\
=\;&|\hat{\beta}_k-\beta_{k,0}|O_p(\xi_{1,m_2}(m_1^{1/2}+m_2^{1/2})n^{-1/2})=|\hat{\beta}_k-\beta_{k,0}|o_p(1).
\end{aligned} \tag{SC.214}$$

By the Cauchy-Schwarz inequality, Assumptions SC2(vi) and SC3(iv), (SC.61), (SC.127) and

(SC.183)

$$
\begin{aligned}
& n^{-1}\left|\tilde{\beta}_{g,*}'(\hat{\mathbf{P}}_2(\beta_{k,0})-\hat{\mathbf{P}}_2^*(\hat{\beta}_k))'(\hat{\mathbf{P}}_2(\beta_{k,0})-\tilde{\mathbf{P}}_2(\beta_{k,0}))'\tilde{\beta}_{\varphi,m_2}(\beta_{k,0})\right| \\
& \leq n^{-1}\left\|(\hat{\mathbf{P}}_2(\beta_{k,0})-\hat{\mathbf{P}}_2^*(\hat{\beta}_k))\tilde{\beta}_{g,*}\right\|\left\|(\hat{\mathbf{P}}_2(\beta_{k,0})-\tilde{\mathbf{P}}_2(\beta_{k,0}))\tilde{\beta}_{\varphi,m_2}(\beta_{k,0})\right\| \\
& = |\hat{\beta}_k-\beta_{k,0}|O_p(\xi_{1,m_2}^2 m_1^{1/2}n^{-1/2}) = |\hat{\beta}_k-\beta_{k,0}|o_p(1). \qquad \text{(SC.215)}
\end{aligned}
$$

By Assumption Assumptions SC2(vi) and SC3(ii, iv), and (SC.183), where $\boldsymbol{\varphi}_n \equiv (\varphi_1, \ldots, \varphi_n)'$, which together with (SC.214) and (SC.215) implies that

$$
\begin{aligned}
& n^{-1}\tilde{\beta}_{g,*}'(\hat{\mathbf{P}}_2(\beta_{k,0})-\hat{\mathbf{P}}_2^*(\hat{\beta}_k))'\hat{\mathbf{P}}_2(\beta_{k,0})\hat{\beta}_{\varphi}(\beta_{k,0}) \\
= \; & n^{-1}\tilde{\beta}_{g,*}'(\hat{\mathbf{P}}_2(\beta_{k,0})-\hat{\mathbf{P}}_2^*(\hat{\beta}_k))'\boldsymbol{\varphi}_n + (\hat{\beta}_k-\beta_{k,0})o_p(1). \qquad \text{(SC.216)}
\end{aligned}
$$

Since $\tilde{\beta}_{g,*} = (B(\beta_{k,0})')^{-1}B(\hat{\beta}_k)'\hat{\beta}_g(\hat{\beta}_k)$, we can write

$$
\tilde{\beta}_{g,*}'(\hat{\mathbf{P}}_2(\beta_{k,0})-\hat{\mathbf{P}}_2^*(\hat{\beta}_k))'\boldsymbol{\varphi}_n = \sum_{i=1}^{n}\hat{\beta}_g(\hat{\beta}_k)'(\tilde{P}_2(\hat{\nu}_{1,i}(\beta_{k,0});\hat{\beta}_k)-\tilde{P}_2(\hat{\nu}_{1,i}(\hat{\beta}_k);\hat{\beta}_k))\varphi_i. \qquad \text{(SC.217)}
$$

By the first-order expansion, the triangle inequality and the Cauchy-Schwarz inequality, Assumptions SC1(i) and SC3(i, iv), and (SC.205), we have

$$
\begin{aligned}
& n^{-1}\sum_{i=1}^{n}(\hat{\beta}_g(\hat{\beta}_k)-\tilde{\beta}_{g,m_2}(\hat{\beta}_k))'(\tilde{P}_2(\hat{\nu}_{1,i}(\beta_{k,0});\hat{\beta}_k)-\tilde{P}_2(\hat{\nu}_{1,i}(\hat{\beta}_k);\hat{\beta}_k))\varphi_i \\
= \; & (\hat{\beta}_k-\beta_{k,0})n^{-1}\sum_{i=1}^{n}(\hat{\beta}_g(\hat{\beta}_k)-\tilde{\beta}_{g,m_2}(\hat{\beta}_k))'\partial^1\tilde{P}_2(\tilde{\nu}_{1,i};\hat{\beta}_k)k_{1,i}\varphi_i \\
= \; & (\hat{\beta}_k-\beta_{k,0})O_p((m_1^{1/2}+m_2)n^{-1/2})O_p(\xi_{1,m_2}) = (\hat{\beta}_k-\beta_{k,0})o_p(1) \qquad \text{(SC.218)}
\end{aligned}
$$

where $\tilde{\nu}_{1,i}$ is between $\hat{\nu}_{1,i}(\beta_{k,0})$ and $\nu_{1,i}(\hat{\beta}_k)$ and it is in $\Omega_\varepsilon(\hat{\beta}_k)$ for any $i = 1, \ldots, n$ wpa1 by Assumption SC3(iv), Lemma SC4 and Lemma SC20. From (SC.217) and (SC.218),

$$
\begin{aligned}
& \tilde{\beta}_{g,*}'(\hat{\mathbf{P}}_2(\beta_{k,0})-\hat{\mathbf{P}}_2^*(\hat{\beta}_k))'\boldsymbol{\varphi}_n \\
= \; & \sum_{i=1}^{n}\tilde{\beta}_{g,m_2}(\hat{\beta}_k)'(\tilde{P}_2(\hat{\nu}_{1,i}(\beta_{k,0});\hat{\beta}_k)-\tilde{P}_2(\hat{\nu}_{1,i}(\hat{\beta}_k);\hat{\beta}_k))\varphi_i + (\hat{\beta}_k-\beta_{k,0})o_p(1). \qquad \text{(SC.219)}
\end{aligned}
$$

By the second order expansion,

$$
\begin{aligned}
& n^{-1}\sum_{i=1}^{n}\tilde{\beta}_{g,m_2}(\hat{\beta}_k)'(\tilde{P}_2(\hat{\nu}_{1,i}(\beta_{k,0});\hat{\beta}_k)-\tilde{P}_2(\nu_{1,i}(\hat{\beta}_k);\hat{\beta}_k))\varphi_i \\
= \; & n^{-1}\sum_{i=1}^{n}\tilde{\beta}_{g,m_2}(\hat{\beta}_k)'\partial^1\tilde{P}_2(\nu_{1,i}(\hat{\beta}_k);\hat{\beta}_k)(\hat{\nu}_{1,i}(\beta_{k,0})-\nu_{1,i}(\hat{\beta}_k))\varphi_i \\
& +n^{-1}\sum_{i=1}^{n}\tilde{\beta}_{g,m_2}(\hat{\beta}_k)'\partial^2\tilde{P}_2(\tilde{\nu}_{1,i};\hat{\beta}_k)(\hat{\nu}_{1,i}(\beta_{k,0})-\nu_{1,i}(\hat{\beta}_k))^2\varphi_i \qquad \text{(SC.220)}
\end{aligned}
$$

where $\tilde{\nu}_{1,i}$ lies between $\hat{\nu}_{1,i}(\beta_{k,0})$ and $\nu_{1,i}(\hat{\beta}_k)$. By Assumption SC3(iv), Lemma SC4 and Lemma SC20, $\tilde{\nu}_{1,i}\in\Omega_\varepsilon(\hat{\beta}_k)$ for any $i=1,\ldots,n$ wpa1. By Assumptions SC2(iii, vi) and SC3(i, iv), Lemma SC4 and Lemma SC20

$$
\begin{aligned}
& n^{-1}\sum_{i=1}^{n}\tilde{\beta}_{g,m_2}(\hat{\beta}_k)'\partial^1\tilde{P}_2(\nu_{1,i}(\hat{\beta}_k);\hat{\beta}_k)(\hat{\nu}_{1,i}(\beta_{k,0})-\nu_{1,i}(\hat{\beta}_k))\varphi_i \\
= \; & n^{-1}\sum_{i=1}^{n}g_1(\nu_{1,i}(\hat{\beta}_k);\hat{\beta}_k)(\hat{\nu}_{1,i}(\beta_{k,0})-\nu_{1,i}(\hat{\beta}_k))\varphi_i+o_p(n^{-1/2}). \qquad \text{(SC.221)}
\end{aligned}
$$

By Assumptions SC2(v, vi) and SC3(i, iv), Lemma SC4 and Lemma SC20, and (SC.69)

$$
n^{-1}\sum_{i=1}^{n}\tilde{\beta}_{g,m_2}(\hat{\beta}_k)'\partial^2\tilde{P}_2(\tilde{\nu}_{1,i};\hat{\beta}_k)(\hat{\nu}_{1,i}(\beta_{k,0})-\nu_{1,i}(\hat{\beta}_k))^2\varphi_i=(\hat{\beta}_k-\beta_{k,0})o_p(1)+o_p(n^{-1/2})
$$

which together with (SC.220) and (SC.221) implies that

$$
\begin{aligned}
& n^{-1}\sum_{i=1}^{n}\tilde{\beta}_{g,m_2}(\hat{\beta}_k)'(\tilde{P}_2(\hat{\nu}_{1,i}(\beta_{k,0});\hat{\beta}_k)-\tilde{P}_2(\nu_{1,i}(\hat{\beta}_k);\hat{\beta}_k))\varphi_i \qquad \text{(SC.222)} \\
= \; & n^{-1}\sum_{i=1}^{n}g_1(\nu_{1,i}(\hat{\beta}_k);\hat{\beta}_k)(\hat{\nu}_{1,i}(\beta_{k,0})-\nu_{1,i}(\hat{\beta}_k))\varphi_i+(\hat{\beta}_k-\beta_{k,0})o_p(1)+o_p(n^{-1/2}).
\end{aligned}
$$

Similarly, we can show that

$$
\begin{aligned}
& n^{-1}\sum_{i=1}^{n}\tilde{\beta}_{g,m_2}(\hat{\beta}_k)'(\tilde{P}_2(\hat{\nu}_{1,i}(\hat{\beta}_k);\hat{\beta}_k)-\tilde{P}_2(\nu_{1,i}(\hat{\beta}_k);\hat{\beta}_k))\varphi_i \\
= \; & n^{-1}\sum_{i=1}^{n}g_1(\nu_{1,i}(\hat{\beta}_k);\hat{\beta}_k)(\hat{\nu}_{1,i}(\hat{\beta}_k)-\nu_{1,i}(\hat{\beta}_k))\varphi_i+o_p(n^{-1/2})
\end{aligned}
$$

which together with (SC.222) implies that

$$
\begin{aligned}
& n^{-1}\sum_{i=1}^{n}\tilde{\beta}_{g,m_2}(\hat{\beta}_k)'(\tilde{P}_2(\hat{\nu}_{1,i}(\beta_{k,0});\hat{\beta}_k)-\tilde{P}_2(\nu_{1,i}(\hat{\beta}_k);\hat{\beta}_k))\varphi_i \\
= \;& (\hat{\beta}_k-\beta_{k,0})n^{-1}\sum_{i=1}^{n}g_1(\nu_{1,i}(\hat{\beta}_k);\hat{\beta}_k)k_{1,i}\varphi_i+(\hat{\beta}_k-\beta_{k,0})o_p(1)+o_p(n^{-1/2}) \quad \text{(SC.223)}
\end{aligned}
$$

By Assumptions SC2(ii, vi) and SC3(i, iv), and Lemma SC20,

$$
n^{-1}\sum_{i=1}^{n}g_1(\nu_{1,i}(\hat{\beta}_k);\hat{\beta}_k)k_{1,i}\varphi_i=n^{-1}\sum_{i=1}^{n}g_{1,i}k_{1,i}\varphi_i+o_p(1)
$$

which combined with (SC.223) implies that

$$
\begin{aligned}
& n^{-1}\sum_{i=1}^{n}\tilde{\beta}_{g,m_2}(\hat{\beta}_k)'(\tilde{P}_2(\hat{\nu}_{1,i}(\beta_{k,0});\hat{\beta}_k)-\tilde{P}_2(\hat{\nu}_{1,i}(\hat{\beta}_k);\hat{\beta}_k))\varphi_i \\
= \;& (\hat{\beta}_k-\beta_{k,0})n^{-1}\sum_{i=1}^{n}g_{1,i}k_{1,i}\varphi_i+(\hat{\beta}_k-\beta_{k,0})o_p(1)+o_p(n^{-1/2}) \\
= \;& (\hat{\beta}_k-\beta_{k,0})\mathbb{E}\left[g_{1,i}k_{1,i}\varphi_i\right]+(\hat{\beta}_k-\beta_{k,0})o_p(1)+o_p(n^{-1/2}) \qquad \text{(SC.224)}
\end{aligned}
$$

where the second equality is by the Markov inequality. The claim in (SC.213) now follows from (SC.219) and (SC.224).

Step 3. In this step, we show that

$$
n^{-1}(\hat{\mathbf{Y}}_2^*(\hat{\beta}_k)-\hat{\mathbf{P}}_2(\hat{\beta}_k)\hat{\beta}_g(\hat{\beta}_k))'(\hat{\mathbf{P}}_2^*(\hat{\beta}_k)-\hat{\mathbf{P}}_2(\beta_{k,0}))\hat{\beta}_{\varphi}(\beta_{k,0})=(\hat{\beta}_k-\beta_{k,0})o_p(1)+o_p(n^{-1/2}). \quad \text{(SC.225)}
$$

Since $\hat{y}_2^*(\hat{\beta}_k)=y_2^*-l_{2,i}\hat{\beta}_l-k_{2,i}\hat{\beta}_k$, we can write

$$
\begin{aligned}
\hat{y}_2^*(\hat{\beta}_k)-\hat{P}_2(\hat{\beta}_k)'\hat{\beta}_g(\hat{\beta}_k) &= y_2^*-l_{2,i}\hat{\beta}_l-k_{2,i}\hat{\beta}_k-\hat{g}(\hat{\nu}_{1,i}(\hat{\beta}_k);\hat{\beta}_k) \\
&= u_{2,i}-l_{2,i}(\hat{\beta}_l-\beta_{l,o})-k_{2,i}(\hat{\beta}_k-\beta_{k,o})-(\hat{g}(\hat{\nu}_{1,i}(\hat{\beta}_k);\hat{\beta}_k)-g(\nu_{1,i})).
\end{aligned}
$$

Therefore,

$$
\begin{aligned}
& n^{-1}(\hat{\mathbf{Y}}_2^*(\hat{\beta}_k)-\hat{\mathbf{P}}_2(\hat{\beta}_k)\hat{\beta}_g(\hat{\beta}_k))'(\hat{\mathbf{P}}_2^*(\hat{\beta}_k)-\hat{\mathbf{P}}_2(\beta_{k,0}))\hat{\beta}_{\varphi}(\beta_{k,0}) \\
= \;& n^{-1}\mathbf{U}_2'(\hat{\mathbf{P}}_2^*(\hat{\beta}_k)-\hat{\mathbf{P}}_2(\beta_{k,0}))\hat{\beta}_{\varphi}(\beta_{k,0}) \\
& -(\hat{\beta}_l-\beta_{l,o})n^{-1}\mathbf{L}_2'(\hat{\mathbf{P}}_2^*(\hat{\beta}_k)-\hat{\mathbf{P}}_2(\beta_{k,0}))\hat{\beta}_{\varphi}(\beta_{k,0}) \\
& -n^{-1}(\hat{\beta}_k-\beta_{l,o})\mathbf{K}_2'(\hat{\mathbf{P}}_2^*(\hat{\beta}_k)-\hat{\mathbf{P}}_2(\beta_{k,0}))\hat{\beta}_{\varphi}(\beta_{k,0}) \\
& -n^{-1}(\hat{\mathbf{G}}_2-\mathbf{G}_2)'(\hat{\mathbf{P}}_2^*(\hat{\beta}_k)-\hat{\mathbf{P}}_2(\beta_{k,0}))\hat{\beta}_{\varphi}(\beta_{k,0}) \qquad \text{(SC.226)}
\end{aligned}
$$

which combined with Lemma SC22 proves (SC.225).

Step 4. In this step, we show that

$$n^{-1}(\hat{\mathbf{Y}}_2^*(\hat{\beta}_k)-\hat{\mathbf{Y}}_2^*(\beta_{k,0}))'\hat{\mathbf{P}}_2(\beta_{k,0})\hat{\beta}_\varphi(\beta_{k,0}) = -(\hat{\beta}_k-\beta_{k,0})\left(\mathbb{E}[k_{2,i}(\gamma_{2,i}-\gamma_{1,i}g_{1,i})]+o_p(1)\right). \quad \text{(SC.227)}$$

Since $\hat{y}_2^*(\hat{\beta}_k)-\hat{y}_2^*(\beta_{k,0}) = -k_{2,i}(\hat{\beta}_k-\beta_{k,0})$, we can write

$$n^{-1}(\hat{\mathbf{Y}}_2^*(\hat{\beta}_k)-\hat{\mathbf{Y}}_2^*(\beta_{k,0}))'\hat{\mathbf{P}}_2(\beta_{k,0})\hat{\beta}_\varphi(\beta_{k,0}) = -n^{-1}(\hat{\beta}_k-\beta_{k,0})'\mathbf{K}_2'\hat{\mathbf{P}}_2(\beta_{k,0})\hat{\beta}_\varphi(\beta_{k,0}) \quad \text{(SC.228)}$$

and

$$\begin{aligned} n^{-1}\mathbf{K}_2'\hat{\mathbf{P}}_2(\beta_{k,0})\hat{\beta}_\varphi(\beta_{k,0}) &= \mathbb{E}[k_{2,i}\varphi(\nu_{1,i})] + n^{-1}\sum_{i=1}^n(k_{2,i}\varphi(\nu_{1,i}) - \mathbb{E}[k_{2,i}\varphi(\nu_{1,i})]) \\ &\quad + n^{-1}\sum_{i=1}^n k_{2,i}(\varphi(\nu_{1,i}) - \hat{P}_{2,i}(\beta_{k,0})'\tilde{\beta}_{\varphi,m_2}(\beta_{k,0})) \\ &\quad + n^{-1}\sum_{i=1}^n k_{2,i}\hat{P}_{2,i}(\beta_{k,0})'(\hat{\beta}_\varphi(\beta_{k,0}) - \tilde{\beta}_{\varphi,m_2}(\beta_{k,0})). \quad \text{(SC.229)} \end{aligned}$$

By the Markov inequality, Assumptions SC1(i) and SC3(i)

$$n^{-1}\sum_{i=1}^n(k_{2,i}\varphi(\nu_{1,i}) - \mathbb{E}[k_{2,i}\varphi(\nu_{1,i})]) = o_p(1) \quad \text{(SC.230)}$$

By the mean value expansion, Assumptions SC1(i), SC2(i) and SC3(ii, iv)

$$\begin{aligned} & n^{-1}\sum_{i=1}^n k_{2,i}(\varphi(\nu_{1,i}) - \hat{P}_{2,i}(\beta_{k,0})'\tilde{\beta}_{\varphi,m_2}(\beta_{k,0})) \\ =\ & n^{-1}\sum_{i=1}^n k_{2,i}(\tilde{P}_2(\nu_{1,i}(\beta_{k,0});\beta_{k,0}) - \tilde{P}_2(\hat{\nu}_{1,i}(\beta_{k,0});\beta_{k,0}))'\tilde{\beta}_{\varphi,m_2}(\beta_{k,0}) \\ & + n^{-1}\sum_{i=1}^n k_{2,i}(\varphi(\nu_{1,i}) - \tilde{P}_2(\nu_{1,i}(\beta_{k,0});\beta_{k,0})'\tilde{\beta}_{\varphi,m_2}(\beta_{k,0})) \\ =\ & -n^{-1}\sum_{i=1}^n k_{2,i}(\hat{\phi}_i - \phi_i)\partial^1\tilde{P}_2(\tilde{\nu}_{1,i}(\beta_{k,0});\beta_{k,0})'\tilde{\beta}_{\varphi,m_2}(\beta_{k,0}) + o_p(1) \quad \text{(SC.231)} \end{aligned}$$

where $\tilde{\nu}_{1,i}(\beta_{k,0})$ lies between $\hat{\nu}_{1,i}(\beta_{k,0})$ and $\nu_{1,i}(\beta_{k,0})$. By (SC.59), $\tilde{\nu}_{1,i}(\beta_{k,0}) \in \Omega_\varepsilon(\beta_{k,0})$ for any $i = 1, \ldots, n$ wpa1. By the triangle inequality and the Cauchy-Schwarz inequality, Assumptions SC1(i), SC2(i, v) and SC3(iv), and (SC.127)

$$n^{-1}\sum_{i=1}^{n}k_{2,i}(\hat{\phi}_i-\phi_i)\partial^1\tilde{P}_2(\tilde{\nu}_{1,i}(\beta_{k,0});\beta_{k,0})'\tilde{\beta}_{\varphi,m_2}(\beta_{k,0})=O_p(\xi_{1,m_2}m_1^{1/2}n^{-1/2})=o_p(1)$$

which together with (SC.231) implies that

$$n^{-1}\sum_{i=1}^{n}k_{2,i}(\varphi(\nu_{1,i})-\hat{P}_{2,i}(\beta_{k,0})'\tilde{\beta}_{\varphi,m_2}(\beta_{k,0}))=o_p(1). \tag{SC.232}$$

By the Cauchy-Schwarz inequality, Assumptions SC1(i) and SC2(i) (SC.66) and (SC.126)

$$\begin{aligned}&\left|n^{-1}\sum_{i=1}^{n}k_{2,i}\hat{P}_{2,i}(\beta_{k,0})'(\hat{\beta}_{\varphi}(\beta_{k,0})-\tilde{\beta}_{\varphi,m_2}(\beta_{k,0}))\right| \qquad \text{(SC.233)}\\ \leq\ &(\lambda_{\max}(n^{-1}\hat{\mathbf{P}}_2(\beta_{k,0})'\hat{\mathbf{P}}_2(\beta_{k,0})))^{1/2}\left(n^{-1}\sum_{i=1}^{n}k_{2,i}^2\right)^{1/2}\left\|\hat{\beta}_{\varphi}(\beta_{k,0})-\tilde{\beta}_{\varphi,m_2}(\beta_{k,0})\right\|=o_p(1).\end{aligned}$$

The claim in (SC.227) follows from (SC.228), (SC.229), (SC.230), (SC.232) and (SC.233). *Q.E.D.*

SC.4 Auxiliary results for the standard error estimation

Assumption SC4. *(i) There exist $\hat{\varepsilon}_{1,i}$ for $i=1,\ldots,n$ such that $n^{-1}\sum_{i=1}^{n}(\hat{\varepsilon}_{1,i}-\varepsilon_{1,i})^4=o_p(1)$; (ii) there exist $r_a>0$ and $\beta_{a_2,m}\in\mathbb{R}^m$ such that $\sup_{x\in\mathcal{X}}|a_{2,m}(x)-a_2(x)|=O(m^{-r_a})$ where $a_{2,m}(x)\equiv P_1(x)'\beta_{a_2,m}$ and $\xi_{0,m_1}m^{-r_a}=o(1)$; (iii) $\Omega>0$; (iv) $\xi_{0,m_1}m_1^{1/2}m_2^3n^{-1/2}=o(1)$.*

Assumption SC4(i) assumes the existence of estimators of the random variables $\varepsilon_{1,i}$ in the linear representation of the estimation error in $\hat{\beta}_l$. Specific estimator $\hat{\varepsilon}_{1,i}$ can be constructed using the form of $\varepsilon_{1,i}$.[5] Assumption SC4(ii) requires that the unknown function $a_2(x_{1,i})\equiv\mathbb{E}[k_{2,i}|x_{1,i}]$ can be well approximated by the approximating functions $P_1(x_{1,i})$. Assumption SC4(iii) requires that the asymptotic variance Ω is bounded away from zero. Assumption SC4(iv) restricts the numbers of the approximation functions used in the multi-step estimation procedure.

The following lemma is useful to show the consistency of the standard error estimator.

Lemma SC24. *Under Assumptions SC1, SC2, SC3 and SC4, we have*

(i) $n^{-1}\hat{\mathbf{P}}_2(\hat{\beta}_k)'\hat{\mathbf{P}}_2(\hat{\beta}_k)-n^{-1}\hat{\mathbf{P}}_2(\beta_{k,0})'\hat{\mathbf{P}}_2(\beta_{k,0})=O_p(\xi_{1,m_2}n^{-1/2})$;

(ii) $\max_{i\leq n}|\hat{g}_{1,i}-g_{1,i}|^4=o_p(1)$;

(iii) $n^{-1}\sum_{i=1}^{n}(\hat{\varsigma}_{1,i}-\varsigma_{1,i})^4=o_p(1)$;

(iv) $n^{-1}\sum_{i=1}^{n}(\hat{u}_{2,i}-u_{2,i})^4=o_p(1)$;

(v) $\max_{i\leq n}|\hat{\varsigma}_{2,i}-\varsigma_{2,i}|=o_p(1)$.

[5] See (SC.271) in Subsection SC.5 for the form of $\hat{\varepsilon}_{1,i}$ when $\beta_{l,0}$ is estimated by the partially linear regression proposed in Olley and Pakes (1996).

Proof of Lemma SC24. (i) For any $b \in \mathbb{R}^{m_2}$, by the mean value expansion, Assumption SC2(v) and (SB.16)

$$
\begin{aligned}
& b'(\hat{\mathbf{P}}_2(\hat{\beta}_k) - \hat{\mathbf{P}}_2(\beta_{k,0}))'(\hat{\mathbf{P}}_2(\hat{\beta}_k) - \hat{\mathbf{P}}_2(\beta_{k,0}))b \\
= & \sum_{i=1}^{n} (b'((\hat{P}_{2,i}(\hat{\beta}_k) - \hat{P}_{2,i}(\beta_{k,0})))^2 \\
= & (\hat{\beta}_k - \beta_{k,0})^2 \sum_{i=1}^{n} (b' \partial \tilde{P}_2(\hat{\nu}_{1,i}(\tilde{\beta}_k); \tilde{\beta}_k)/\partial \beta_k)^2 = \|b\|^2 \, O_p(\xi_{1,m_2}^2)
\end{aligned}
$$

where $\tilde{\beta}_k$ is between $\hat{\beta}_k$ and $\beta_{k,0}$, and by (SC.58) and Assumption SC2(vi), $\hat{\nu}_{1,i}(\tilde{\beta}_k) \in \Omega_\varepsilon(\tilde{\beta}_k)$ for any $i = 1, \ldots, n$ wpa1. Therefore

$$
\left\| \hat{\mathbf{P}}_2(\hat{\beta}_k) - \hat{\mathbf{P}}_2(\beta_{k,0}) \right\|_S = O_p(\xi_{1,m_2}). \tag{SC.234}
$$

By the triangle inequality and the Cauchy-Schwarz inequality, Assumptions SC2(vi) and SC3(iv), (SC.66) and (SC.234)

$$
\begin{aligned}
& n^{-1} \left\| \hat{\mathbf{P}}_2(\hat{\beta}_k)'\hat{\mathbf{P}}_2(\hat{\beta}_k) - \hat{\mathbf{P}}_2(\beta_{k,0})'\hat{\mathbf{P}}_2(\beta_{k,0}) \right\|_S \\
\leq & \; n^{-1} \left\| (\hat{\mathbf{P}}_2(\hat{\beta}_k) - \hat{\mathbf{P}}_2(\beta_{k,0}))'\hat{\mathbf{P}}_2(\beta_{k,0}) \right\|_S \\
& + n^{-1} \left\| \hat{\mathbf{P}}_2(\beta_{k,0})'(\hat{\mathbf{P}}_2(\hat{\beta}_k) - \hat{\mathbf{P}}_2(\beta_{k,0})) \right\|_S \\
& + n^{-1} \left\| (\hat{\mathbf{P}}_2(\hat{\beta}_k) - \hat{\mathbf{P}}_2(\beta_{k,0}))'(\hat{\mathbf{P}}_2(\hat{\beta}_k) - \hat{\mathbf{P}}_2(\beta_{k,0})) \right\|_S = O_p(\xi_{1,m_2} n^{-1/2})
\end{aligned} \tag{SC.235}
$$

which finishes the proof.

(ii) By triangle inequality and the Cauchy-Schwarz inequality, Assumption SC2(iii, v), and (SC.205)

$$
\begin{aligned}
& \max_{i \leq n} \left| \partial^1 \hat{P}_{2,i}(\hat{\beta}_k)'\hat{\beta}_g(\hat{\beta}_k) - g_1(\nu_{1,i}(\hat{\beta}_k); \hat{\beta}_k) \right| \\
\leq & \max_{i \leq n} \left| (\partial^1 \hat{P}_{2,i}(\hat{\beta}_k) - \partial^1 \tilde{P}_{2,i}(\hat{\beta}_k))'\hat{\beta}_g(\hat{\beta}_k) \right| \\
& + \max_{i \leq n} \left| \partial^1 \tilde{P}_{2,i}(\hat{\beta}_k)'(\hat{\beta}_g(\hat{\beta}_k) - \tilde{\beta}_{g,m_2}(\hat{\beta}_k)) \right| \\
& + \max_{i \leq n} \left| \partial^1 \tilde{P}_{2,i}(\hat{\beta}_k)'\tilde{\beta}_{g,m_2}(\hat{\beta}_k) - g_1(\nu_{1,i}(\hat{\beta}_k); \hat{\beta}_k) \right| \\
\leq & \max_{i \leq n} \left| (\partial^1 \hat{P}_{2,i}(\hat{\beta}_k) - \partial^1 \tilde{P}_{2,i}(\hat{\beta}_k))'\hat{\beta}_g(\hat{\beta}_k) \right| + O_p(\xi_{1,m_2}(m_2 + m_1^{1/2})n^{-1/2}).
\end{aligned} \tag{SC.236}
$$

By the mean value expansion,

$$
(\partial^1 \hat{P}_{2,i}(\hat{\beta}_k) - \partial^1 \tilde{P}_{2,i}(\hat{\beta}_k))'\hat{\beta}_g(\hat{\beta}_k) = (\hat{\phi}_i - \phi_i)\partial^2 \tilde{P}_2(\tilde{\nu}_{1,i}(\hat{\beta}_k); \hat{\beta}_k)'\hat{\beta}_g(\hat{\beta}_k) \tag{SC.237}
$$

where $\tilde{\nu}_{1,i}(\hat{\beta}_k)$ is between $\nu_{1,i}(\hat{\beta}_k)$ and $\hat{\nu}_{1,i}(\hat{\beta}_k)$. By (SC.59), $\tilde{\nu}_{1,i}(\hat{\beta}_k) \in \Omega_\varepsilon(\hat{\beta}_k)$ for any $i = 1, \ldots, n$ wpa1. Therefore by the Cauchy-Schwarz inequality, Assumptions SC2(v, vi), SC3(iv) and SC4(iv), Lemma SC4, (SC.78), (SC.236) and (SC.237),

$$\begin{aligned}
& \max_{i \le n} \left| \partial^1 \hat{P}_{2,i}(\hat{\beta}_k)' \hat{\beta}_g(\hat{\beta}_k) - g_1(\nu_{1,i}(\hat{\beta}_k); \hat{\beta}_k) \right| \\
\le \;& O_p(\xi_{2,m_2}) \max_{i \le n} |\hat{\phi}_i - \phi_i| + O_p(\xi_{1,m_2}(m_2 + m_1^{1/2}) n^{-1/2}) \\
= \;& O_p(\xi_{0,m_1}\xi_{2,m_2} m_1^{1/2} n^{-1/2}) + O_p(\xi_{1,m_2}(m_2 + m_1^{1/2}) n^{-1/2}) = o_p(1). \qquad \text{(SC.238)}
\end{aligned}$$

By Assumption SC2(ii) and (SB.16), we have

$$\max_{i \le n} \left| g_1(\nu_{1,i}(\hat{\beta}_k); \hat{\beta}_k) - g_1(\nu_{1,i}(\beta_{k,0}); \beta_{k,0}) \right| = O_p(n^{-1/2}) \qquad \text{(SC.239)}$$

which together with (SC.238) proves the second claim of the lemma.

(iii) Define $\hat{\varphi}_i \equiv \hat{P}_{2,i}(\hat{\beta}_k)' \hat{\beta}_\varphi(\hat{\beta}_k)$ for $i \le n$, where

$$\hat{\beta}_\varphi(\hat{\beta}_k) \equiv (\hat{\mathbf{P}}_2(\hat{\beta}_k)' \hat{\mathbf{P}}_2(\hat{\beta}_k))^{-1} \sum_{i=1}^n \hat{P}_{2,i}(\hat{\beta}_k)(k_{2,i} - k_{1,i}\hat{g}_{1,i}(\hat{\beta}_k)).$$

Recall that $\Delta k_{2,i} \equiv k_{2,i} - k_{1,i} g_{1,i}$ and $\Delta \hat{k}_{2,i} \equiv k_{2,i} - k_{1,i}\hat{g}_{1,i}(\hat{\beta}_k)$. Since $\varsigma_{1,i} = \Delta k_{2,i} - \varphi_i$ and $\hat{\varsigma}_{1,i} = \Delta \hat{k}_{2,i} - \hat{\varphi}_i$, we have

$$n^{-1} \sum_{i=1}^n (\hat{\varsigma}_{1,i} - \varsigma_{1,i})^4 \le C n^{-1} \sum_{i=1}^n (\Delta \hat{k}_{2,i} - \Delta k_{2,i})^4 + C n^{-1} \sum_{i=1}^n (\hat{\varphi}_i - \varphi_i)^4. \qquad \text{(SC.240)}$$

By Lemma SC24(ii),

$$n^{-1} \sum_{i=1}^n (\Delta \hat{k}_{2,i} - \Delta k_{2,i})^4 = n^{-1} \sum_{i=1}^n k_{1,i}^4 (\hat{g}_{1,i}(\hat{\beta}_k) - g_1(\nu_{1,i}(\beta_{k,0}); \beta_{k,0}))^4 = o_p(1). \qquad \text{(SC.241)}$$

By Assumptions SC2(vi) and SC3(iv), (SC.152), (SC.153) and (SC.205)

$$n^{-1} \sum_{i=1}^n \left| \hat{g}_{1,i}(\hat{\beta}_k) - g_1(\nu_{1,i}(\hat{\beta}_k); \hat{\beta}_k) \right|^2 = O_p(m_1 m_2^6 n^{-1})$$

which together with (SC.239) implies that

$$n^{-1} \sum_{i=1}^n (\hat{g}_{1,i}(\hat{\beta}_k) - g_1(\nu_{1,i}(\beta_{k,0}); \beta_{k,0}))^2 = O_p(m_1 m_2^6 n^{-1}). \qquad \text{(SC.242)}$$

Therefore,

$$n^{-1}\sum_{i=1}^{n}(\Delta\hat{k}_{2,i}-\Delta k_{2,i})^2 \leq Cn^{-1}\sum_{i=1}^{n}(\hat{g}_{1,i}(\hat{\beta}_k)-g_1(\nu_{1,i}(\beta_{k,0});\beta_{k,0}))^2 = O_p(m_1 m_2^6 n^{-1}). \quad \text{(SC.243)}$$

By the definition of $\hat{\varphi}_i$, we can write

$$\begin{aligned}
\hat{\varphi}_i-\varphi_i \;=\;& \hat{P}_{2,i}(\hat{\beta}_k)'(\hat{\mathbf{P}}_2(\hat{\beta}_k)'\hat{\mathbf{P}}_2(\hat{\beta}_k))^{-1}\hat{\mathbf{P}}_2(\hat{\beta}_k)(\Delta\hat{\mathbf{K}}_2-\Delta\mathbf{K}_2)\\
&+\hat{P}_{2,i}(\hat{\beta}_k)'(\hat{\mathbf{P}}_2(\hat{\beta}_k)'\hat{\mathbf{P}}_2(\hat{\beta}_k))^{-1}(\hat{\mathbf{P}}_2(\hat{\beta}_k)-\hat{\mathbf{P}}_2(\beta_{k,0}))\Delta\mathbf{K}_2\\
&+\hat{P}_{2,i}(\hat{\beta}_k)'[(\hat{\mathbf{P}}_2(\hat{\beta}_k)'\hat{\mathbf{P}}_2(\hat{\beta}_k))^{-1}-(\hat{\mathbf{P}}_2(\beta_{k,0})'\hat{\mathbf{P}}_2(\beta_{k,0}))^{-1}]\hat{\mathbf{P}}_2(\beta_{k,0})\Delta\mathbf{K}_2\\
&+(\hat{P}_{2,i}(\hat{\beta}_k)-\hat{P}_{2,i}(\beta_{k,0}))'(\hat{\mathbf{P}}_2(\beta_{k,0})'\hat{\mathbf{P}}_2(\beta_{k,0}))^{-1}\hat{\mathbf{P}}_2(\beta_{k,0})\Delta\mathbf{K}_2\\
&+\hat{P}_{2,i}(\beta_{k,0})'(\hat{\mathbf{P}}_2(\beta_{k,0})'\hat{\mathbf{P}}_2(\beta_{k,0}))^{-1}\hat{\mathbf{P}}_2(\beta_{k,0})\Delta\mathbf{K}_2-\varphi_i. \qquad \text{(SC.244)}
\end{aligned}$$

where $\Delta\hat{\mathbf{K}}_2 \equiv (\Delta\hat{k}_{2,1},\ldots,\Delta\hat{k}_{2,n})'$ and $\Delta\mathbf{K}_2 \equiv (\Delta k_{2,1},\ldots,\Delta k_{2,n})'$. By Assumption SC2(v) and (SC.59),

$$\max_{i\leq n}\left\|\hat{P}_{2,i}(\hat{\beta}_k)\right\| = O_p(\xi_{0,m_2}). \quad \text{(SC.245)}$$

By Assumption SC2(v, vi), (SC.66), (SC.243) and (SC.245),

$$\begin{aligned}
& n^{-1}\sum_{i=1}^{n}(\hat{P}_{2,i}(\hat{\beta}_k)'(\hat{\mathbf{P}}_2(\hat{\beta}_k)'\hat{\mathbf{P}}_2(\hat{\beta}_k))^{-1}\hat{\mathbf{P}}_2(\hat{\beta}_k)(\Delta\hat{\mathbf{K}}_2-\Delta\mathbf{K}_2))^4\\
\leq\;& \xi_{0,m_2}^2\left\|(\hat{\mathbf{P}}_2(\hat{\beta}_k)'\hat{\mathbf{P}}_2(\hat{\beta}_k))^{-1}\hat{\mathbf{P}}_2(\hat{\beta}_k)(\Delta\hat{\mathbf{K}}_2-\Delta\mathbf{K}_2)\right\|^2\\
&\times n^{-1}\sum_{i=1}^{n}(\hat{P}_{2,i}(\hat{\beta}_k)'(\hat{\mathbf{P}}_2(\hat{\beta}_k)'\hat{\mathbf{P}}_2(\hat{\beta}_k))^{-1}\hat{\mathbf{P}}_2(\hat{\beta}_k)(\Delta\hat{\mathbf{K}}_2-\Delta\mathbf{K}_2))^2\\
\leq\;& (\lambda_{\min}(n^{-1}\hat{\mathbf{P}}_2(\hat{\beta}_k)'\hat{\mathbf{P}}_2(\hat{\beta}_k)))^{-1}\xi_{0,m_2}^2\left(n^{-1}\sum_{i=1}^{n}(\Delta\hat{k}_{2,i}-\Delta k_{2,i})^2\right)^2\\
=\;& O_p\left(m_1\xi_{0,m_2}^2 m_2^6 n^{-1}\right) = o_p(1). \qquad \text{(SC.246)}
\end{aligned}$$

By the the Cauchy-Schwarz inequality and Assumptions SC1(i), SC2(i, v, vi) and SC3(iv), (SC.66), (SC.234) and (SC.245),

$$
\begin{aligned}
& n^{-1}\sum_{i=1}^{n}\left|\hat{P}_{2,i}(\hat{\beta}_k)'(\hat{\mathbf{P}}_2(\hat{\beta}_k)'\hat{\mathbf{P}}_2(\hat{\beta}_k))^{-1}(\hat{\mathbf{P}}_2(\hat{\beta}_k)-\hat{\mathbf{P}}_2(\beta_{k,0}))\Delta\mathbf{K}_2\right|^4 \\
\leq\ & n^{-1}\xi_{0,m_2}^2\left\|(\hat{\mathbf{P}}_2(\hat{\beta}_k)'\hat{\mathbf{P}}_2(\hat{\beta}_k))^{-1}(\hat{\mathbf{P}}_2(\hat{\beta}_k)-\hat{\mathbf{P}}_2(\beta_{k,0}))\Delta\mathbf{K}_2\right\|^2 \\
& \times\left\|(\hat{\mathbf{P}}_2(\hat{\beta}_k)'\hat{\mathbf{P}}_2(\hat{\beta}_k))^{-1/2}(\hat{\mathbf{P}}_2(\hat{\beta}_k)-\hat{\mathbf{P}}_2(\beta_{k,0}))\Delta\mathbf{K}_2\right\|^2 \qquad \text{(SC.247)} \\
\leq\ & \frac{\xi_{0,m_2}^2\left\|\hat{\mathbf{P}}_2(\hat{\beta}_k)-\hat{\mathbf{P}}_2(\beta_{k,0})\right\|_S^4}{n^2(\lambda_{\min}(n^{-1}\hat{\mathbf{P}}_2(\hat{\beta}_k)'\hat{\mathbf{P}}_2(\hat{\beta}_k)))^3}\left|n^{-1}\sum_{i=1}^{n}(\Delta k_{2,i})^2\right|^2 = O_p(\xi_{0,m_2}^2\xi_{1,m_2}^4 n^{-2}) = o_p(1).
\end{aligned}
$$

By the Cauchy-Schwarz inequality, Lemma SC24(i), (SC.78), (SC.127) and (SC.245),

$$
\begin{aligned}
& n^{-1}\sum_{i=1}^{n}\left|\hat{P}_{2,i}(\hat{\beta}_k)'[(\hat{\mathbf{P}}_2(\hat{\beta}_k)'\hat{\mathbf{P}}_2(\hat{\beta}_k))^{-1}-(\hat{\mathbf{P}}_2(\beta_{k,0})'\hat{\mathbf{P}}_2(\beta_{k,0}))^{-1}]\hat{\mathbf{P}}_2(\beta_{k,0})\Delta\mathbf{K}_2\right|^4 \\
=\ & n^{-1}\sum_{i=1}^{n}\left|\hat{P}_{2,i}(\hat{\beta}_k)'(\hat{\mathbf{P}}_2(\hat{\beta}_k)'\hat{\mathbf{P}}_2(\hat{\beta}_k))^{-1}[\hat{\mathbf{P}}_2(\hat{\beta}_k)'\hat{\mathbf{P}}_2(\hat{\beta}_k)-\hat{\mathbf{P}}_2(\beta_{k,0})'\hat{\mathbf{P}}_2(\beta_{k,0})]\hat{\beta}_{\varphi}(\beta_{k,0})\right|^4 \\
\leq\ & \xi_{0,m_2}^2\left\|(\hat{\mathbf{P}}_2(\hat{\beta}_k)'\hat{\mathbf{P}}_2(\hat{\beta}_k))^{-1}[\hat{\mathbf{P}}_2(\hat{\beta}_k)'\hat{\mathbf{P}}_2(\hat{\beta}_k)-\hat{\mathbf{P}}_2(\beta_{k,0})'\hat{\mathbf{P}}_2(\beta_{k,0})]\hat{\beta}_{\varphi}(\beta_{k,0})\right\|^2 \\
& \times n^{-1}\sum_{i=1}^{n}\left|\hat{P}_{2,i}(\hat{\beta}_k)'(\hat{\mathbf{P}}_2(\hat{\beta}_k)'\hat{\mathbf{P}}_2(\hat{\beta}_k))^{-1}[\hat{\mathbf{P}}_2(\hat{\beta}_k)'\hat{\mathbf{P}}_2(\hat{\beta}_k)-\hat{\mathbf{P}}_2(\beta_{k,0})'\hat{\mathbf{P}}_2(\beta_{k,0})]\hat{\beta}_{\varphi}(\beta_{k,0})\right|^2 \\
\leq\ & \frac{\xi_{0,m_2}^2\left\|\hat{\beta}_{\varphi}(\beta_{k,0})\right\|^4}{(\lambda_{\min}(n^{-1}\hat{\mathbf{P}}_2(\hat{\beta}_k)'\hat{\mathbf{P}}_2(\hat{\beta}_k)))^3}\left\|n^{-1}\hat{\mathbf{P}}_2(\hat{\beta}_k)'\hat{\mathbf{P}}_2(\hat{\beta}_k)-n^{-1}\hat{\mathbf{P}}_2(\beta_{k,0})'\hat{\mathbf{P}}_2(\beta_{k,0})\right\|_S^4 \\
=\ & O_p(\xi_{0,m_2}^2\xi_{1,m_2}^4 n^{-2}) = o_p(1) \qquad \text{(SC.248)}
\end{aligned}
$$

where the second equality is by Assumptions SC2(vi) and SC3(iv). By the first order expansion, (SB.16) in Theorem SB1, Assumption SC3(ii) and (SC.127),

$$
\begin{aligned}
& n^{-1}\sum_{i=1}^{n}((\hat{P}_{2,i}(\hat{\beta}_k)-\hat{P}_{2,i}(\beta_{k,0}))'\hat{\beta}_{\varphi}(\beta_{k,0}))^4 \\
=\ & (\hat{\beta}_k-\beta_{k,0})^4 n^{-1}\sum_{i=1}^{n}(\partial\tilde{P}_2(\hat{\nu}_{1,i}(\tilde{\beta}_k);\tilde{\beta}_k)/\partial\beta_k)'\hat{\beta}_{\varphi}(\beta_{k,0}))^4 \\
\leq\ & (\hat{\beta}_k-\beta_{k,0})^4\xi_{1,m_2}^4\left\|\hat{\beta}_{\varphi}(\beta_{k,0})\right\|^4 = O_p(\xi_{1,m_2}^4 n^{-2}) = o_p(1) \qquad \text{(SC.249)}
\end{aligned}
$$

where the second equality is by Assumptions SC2(vi) and SC3(iv). By Assumptions SC2(v) and

SC3(i, iv), Lemma SC4, (SC.62), (SC.126), and (SC.127)

$$
\begin{aligned}
& n^{-1}\sum_{i=1}^{n}(\hat{P}_{2,i}(\beta_{k,0})'(\hat{\mathbf{P}}_2(\beta_{k,0})'\hat{\mathbf{P}}_2(\beta_{k,0}))^{-1}\hat{\mathbf{P}}_2(\beta_{k,0})\Delta\mathbf{K}_2-\varphi_i)^4 \\
\leq\; & Cn^{-1}\sum_{i=1}^{n}((\hat{P}_{2,i}(\beta_{k,0})-\tilde{P}_{2,i}(\beta_{k,0}))'\hat{\beta}_{\varphi}(\beta_{k,0}))^4 \\
& +Cn^{-1}\sum_{i=1}^{n}(\tilde{P}_{2,i}(\beta_{k,0})'(\hat{\beta}_{\varphi}(\beta_{k,0})-\tilde{\beta}_{\varphi,m_2}(\beta_{k,0})))^4 \\
& +Cn^{-1}\sum_{i=1}^{n}(\tilde{P}_{2,i}(\beta_{k,0})'\tilde{\beta}_{\varphi,m_2}(\beta_{k,0})-\varphi_i)^4 \\
\leq\; & C\left\|\hat{\beta}_{\varphi}(\beta_{k,0})\right\|^4\xi_{1,m_2}^4 n^{-1}\sum_{i=1}^{n}(\hat{\phi}_i-\phi_i)^4 \\
& +\xi_{0,m_2}^2\lambda_{\max}(n^{-1}\tilde{\mathbf{P}}_2(\beta_{k,0})'\tilde{\mathbf{P}}_2(\beta_{k,0}))\left\|\hat{\beta}_{\varphi}(\beta_{k,0})-\tilde{\beta}_{\varphi,m_2}(\beta_{k,0})\right\|^4+O(m_2^{-4r_{\varphi}}) \\
=\; & O_p((\xi_{1,m_2}^4 m_1^2+m_2^2)\xi_{0,m_2}^2 n^{-2})=o_p(1) \qquad \text{(SC.250)}
\end{aligned}
$$

where the second equality is by Assumptions SC2(vi) and SC3(iv). Collecting the results in (SC.244), (SC.246), (SC.247), (SC.248), (SC.249) and (SC.250), we get

$$
n^{-1}\sum_{i=1}^{n}(\hat{\varphi}_i-\varphi_i)^4=o_p\left(1\right)
$$

which together with Assumption SC3(iv), (SC.240) and (SC.241) proves the third claim of the lemma.

(iv) By the definition of $\hat{u}_{2,i}$, we can write

$$
\hat{u}_{2,i}-u_{2,i}=-l_{2,i}(\hat{\beta}_l-\beta_{l,0})-k_{2,i}(\hat{\beta}_k-\beta_{k,0})-(\hat{g}_i(\hat{\beta}_k)-g(\nu_{1,i}(\beta_{k,0});\beta_{k,0}))
$$

which implies that

$$
\begin{aligned}
n^{-1}\sum_{i=1}^{n}(\hat{u}_{2,i}-u_{2,i})^4 \;\leq\; & C(\hat{\beta}_l-\beta_{l,0})^4 n^{-1}\sum_{i=1}^{n}l_{2,i}^4+C(\hat{\beta}_k-\beta_{k,0})^4 n^{-1}\sum_{i=1}^{n}k_{2,i}^4 \\
& +Cn^{-1}\sum_{i=1}^{n}(\hat{g}_i(\hat{\beta}_k)-g(\nu_{1,i}(\beta_{k,0});\beta_{k,0}))^4 \\
=\; & Cn^{-1}\sum_{i=1}^{n}(\hat{g}_i(\hat{\beta}_k)-g(\nu_{1,i}(\beta_{k,0});\beta_{k,0}))^4+O_p(n^{-2}) \qquad \text{(SC.251)}
\end{aligned}
$$

where the equality is by Assumptions SC1(i, iii) and SC2(i, ii), and (SB.16). Using similar arguments for proving (SC.238), we can show that

$$\begin{aligned}
& \max_{i\leq n}\left|\hat{g}_i(\hat{\beta}_k)-g(\nu_{1,i}(\beta_{k,0});\beta_{k,0})\right| \\
= \;& O_p(\xi_{0,m_1}\xi_{1,m_2}m_1^{1/2}n^{-1/2})+O_p(\xi_{0,m_2}(m_2+m_1^{1/2})n^{-1/2})=o_p(1) \qquad \text{(SC.252)}
\end{aligned}$$

where the second equality is by Assumption SC2(vi). By Assumption SC2(ii) and (SB.16), we have

$$\max_{i\leq n}\left|g(\nu_{1,i}(\hat{\beta}_k);\hat{\beta}_k)-g(\nu_{1,i}(\beta_{k,0});\beta_{k,0})\right|=O_p(n^{-1/2})$$

which together with (SC.252) shows that

$$n^{-1}\sum_{i=1}^{n}(\hat{g}_i(\hat{\beta}_k)-g(\nu_{1,i}(\beta_{k,0});\beta_{k,0}))^4=o_p(1). \qquad \text{(SC.253)}$$

The claim of the lemma follows from (SC.251) and (SC.253).

(v) Let $\hat{\beta}_{a_2}\equiv(\mathbf{P}_1'\mathbf{P}_1)^{-1}\sum_{i=1}^{n}P_1(x_{1,i})k_{2,i}$. By Assumptions SC1 and SC4(ii), we can use similar arguments for proving (SC.55) to show

$$\hat{\beta}_{a_2}-\beta_{a_2,m_1}=O_p(m_1^{1/2}n_1^{-1/2}+m_1^{-r_a}). \qquad \text{(SC.254)}$$

Therefore by the triangle inequality, Assumption SC1(vi) and (SC.254),

$$\begin{aligned}
\max_{i\leq n}|\hat{\varsigma}_{2,i}-\varsigma_{2,i}| &\leq \xi_{0,m_1}\left\|\hat{\beta}_{a_2}-\beta_{a_2,m_1}\right\|+\max_{i\leq n}|a_{2,m_1}(x_{1,i})-a_2(x_{1,i})| \\
&= O_p(\xi_{0,m_1}m_1^{1/2}n_1^{-1/2}+\xi_{0,m_1}m_1^{-r_a})=o_p(1)
\end{aligned}$$

where the second equality is by Assumptions SC1(vi) and SC4(ii). *Q.E.D.*

Lemma SC25. *Under Assumptions SC1, SC2, SC3 and SC4, we have*

(i) $\hat{\Upsilon}_n-\Upsilon=o_p(1)$;

(ii) $\hat{\Gamma}_n-\Gamma=o_p(1)$;

(iii) $\hat{\Omega}_n-\Omega=o_p(1)$.

Proof of Lemma SC32. (i) By Assumptions SC1(i) and SC2(i, ii), and the Markov inequality

$$n^{-1}\sum_{i=1}^{n}\varsigma_{1,i}^2=\Upsilon+O_p(n^{-1/2})=O_p(1) \qquad \text{(SC.255)}$$

which together with Lemma SC24(iii) proves the first claim of the lemma.

(ii) Let $\tilde{\Gamma}_n=\sum_{i=1}^{n}[(l_{2,i}-l_{1,i}g_1(\nu_{1,i}))\,\varsigma_{1,i}+l_{1,i}g_1(\nu_{1,i})\varsigma_{2,i}]$. Then by Assumptions SC1(ii) and

SC2(i, ii), and the Markov inequality, we have

$$\mathbb{E}\left[l_{1,i}^4 + l_{2,i}^4 + \varsigma_{1,i}^4 + \varsigma_{2,i}^4 + g_1(\nu_{1,i})^4\right] \leq C \tag{SC.256}$$

which together with Assumption SC1(i) and the Markov inequality implies that

$$\begin{aligned} & n^{-1}\sum_{i=1}^{n}\left[(l_{2,i} - l_{1,i}g_1(\nu_{1,i}))\,\varsigma_{1,i} + l_{1,i}g_1(\nu_{1,i})\varsigma_{2,i}\right] \\ = \;& \mathbb{E}\left[(l_{2,i} - l_{1,i}g_1(\nu_{1,i}))\,\varsigma_{1,i} + l_{1,i}g_1(\nu_{1,i})\varsigma_{2,i}\right] + O_p(n^{-1/2}) \end{aligned} \tag{SC.257}$$

Therefore

$$\tilde{\Gamma}_n = \Gamma + o_p(1). \tag{SC.258}$$

By the definition of $\hat{\Gamma}_n$, we can write

$$\begin{aligned} \hat{\Gamma}_n - \tilde{\Gamma}_n \;=\; & -n^{-1}\sum_{i=1}^{n} l_{1,i}(\hat{g}_{1,i} - g_{1,i})(\hat{\varsigma}_{1,i} - \varsigma_{1,i}) - n^{-1}\sum_{i=1}^{n} l_{1,i}(\hat{g}_{1,i} - g_{1,i})\varsigma_{1,i} \\ & + n^{-1}\sum_{i=1}^{n}(l_{2,i} - l_{1,i}g_{1,i})(\hat{\varsigma}_{1,i} - \varsigma_{1,i}) + n^{-1}\sum_{i=1}^{n} l_{1,i}(g_{1,i}\varsigma_{2,i} - \hat{g}_{1,i}\hat{\varsigma}_{2,i}). \end{aligned} \tag{SC.259}$$

The second claim of the lemma follows from Assumption SC1(i), Lemma SC24(ii, iii, v), (SC.256), (SC.258) and (SC.259).

(iii) Since $\hat{\eta}_{1,i} = \eta_{1,i} - l_{1,i}(\hat{\beta}_l - \beta_{l,0}) - (\hat{\phi}_i - \phi_i)$, by the Markov inequality, Assumptions SC1(i, iii) and SC2(vi), Lemma SC4 and (SC.256),

$$\begin{aligned} n^{-1}\sum_{i=1}^{n}(\hat{\eta}_{1,i} - \eta_{1,i})^4 \;\leq\; & C(\hat{\beta}_l - \beta_{l,0})^4 n^{-1}\sum_{i=1}^{n} l_{1,i}^4 + \max_{i\leq n}(\hat{\phi}_i - \phi_i)^2 n^{-1}\sum_{i=1}^{n}(\hat{\phi}_i - \phi_i)^2 \\ =\; & O_p(n^{-2}) + O_p(\xi_{0,m_1}^2 m_1^2 n^{-2}) = O_p(\xi_{0,m_1}^2 m_1^2 n^{-2}) = o_p(1). \end{aligned} \tag{SC.260}$$

By Assumption SC2(ii) and Lemma SC24(ii)

$$\max_{i\leq n}\hat{g}_{1,i}^4 \leq C\max_{i\leq n}(\hat{g}_{1,i} - g_{1,i})^4 + C\max_{i\leq n} g_{1,i}^4 = O_p(1). \tag{SC.261}$$

By Assumption SC1(i, ii), Lemma SC24(ii, iv), (SC.260) and (SC.261), we get

$$
\begin{aligned}
& n^{-1}\sum_{i=1}^{n}(\hat{u}_{2,i}-\hat{\eta}_{1,i}\hat{g}_{1,i}-u_{2,i}+\eta_{1,i}g_{1,i})^{4} \\
\leq\ & Cn^{-1}\sum_{i=1}^{n}(\hat{u}_{2,i}-u_{2,i})^{4}+C\max_{i\leq n}\hat{g}_{1,i}^{4}n^{-1}\sum_{i=1}^{n}(\hat{\eta}_{1,i}-\eta_{1,i})^{4} \\
& +C\max_{i\leq n}(\hat{g}_{1,i}-g_{1,i})^{4}n^{-1}\sum_{i=1}^{n}\eta_{1,i}^{4}=o_p(1) \qquad \text{(SC.262)}
\end{aligned}
$$

which together with Assumption SC1(i, ii) and SC2(i, ii), Lemma SC24(iii), (SC.88) and (SC.256) implies that

$$
n^{-1}\sum_{i=1}^{n}\left((\hat{u}_{2,i}-\hat{\eta}_{1,i}\hat{g}_{1,i})\hat{\varsigma}_{1,i}-(u_{2,i}-\eta_{1,i}g_{1,i})\varsigma_{1,i}\right)^{2}=o_p(1). \qquad \text{(SC.263)}
$$

By Assumptions SC1(i, ii, iii) and SC4(i), and (SC.260), we have

$$
n^{-1}\sum_{i=1}^{n}\hat{\varepsilon}_{1,i}^{4}+n^{-1}\sum_{i=1}^{n}\hat{\eta}_{1,i}^{4}=O_p(1) \qquad \text{(SC.264)}
$$

which combined with Lemma SC25(ii), (SC.260) and Assumption SC4(i) implies that

$$
\begin{aligned}
n^{-1}\sum_{i=1}^{n}(\hat{\Gamma}_n\hat{\varepsilon}_{1,i}\hat{\eta}_{1,i}-\Gamma\varepsilon_{1,i}\eta_{1,i})^{2} \leq\ & C(\hat{\Gamma}_n-\Gamma)^{2}n^{-1}\sum_{i=1}^{n}\hat{\varepsilon}_{1,i}^{2}\hat{\eta}_{1,i}^{2} \\
& +C\Gamma^{2}n^{-1}\sum_{i=1}^{n}(\hat{\varepsilon}_{1,i}-\varepsilon_{1,i})^{2}\hat{\eta}_{1,i}^{2} \\
& +C\Gamma^{2}n^{-1}\sum_{i=1}^{n}\varepsilon_{1,i}^{2}(\hat{\eta}_{1,i}-\eta_{1,i})^{2}=o_p(1). \qquad \text{(SC.265)}
\end{aligned}
$$

By Assumptions SC1(i, ii) and SC2(ii), Lemma SC24(ii, v) and (SC.256), we have

$$
\begin{aligned}
& n^{-1}\sum_{i=1}^{n}(\hat{\eta}_{1,i}\hat{g}_{1,i}\hat{\varsigma}_{2,i}-\eta_{1,i}g_{1,i}\varsigma_{2,i})^{2} \\
\leq\ & \max_{i\leq n}\hat{g}_{1,i}^{2}n^{-1}\sum_{i=1}^{n}(\hat{\eta}_{1,i}-\eta_{1,i})^{2}\hat{\varsigma}_{2,i}^{2}+\max_{i\leq n}|\hat{g}_{1,i}-g_{1,i}|\,n^{-1}\sum_{i=1}^{n}\eta_{1,i}^{2}\hat{\varsigma}_{2,i}^{2} \\
& +\max_{i\leq n}|\hat{\varsigma}_{2,i}-\varsigma_{2,i}|\,n^{-1}\sum_{i=1}^{n}\eta_{1,i}^{2}g_{1,i}^{2}=o_p(1). \qquad \text{(SC.266)}
\end{aligned}
$$

Let $\tilde{\Omega}_n=n^{-1}\sum_{i=1}^{n}\left((u_{2,i}-\eta_{1,i}g_{1,i})\varsigma_{1,i}-\Gamma\varepsilon_{1,i}+\eta_{1,i}g_{1,i}\varsigma_{2,i}\right)^{2}$. Then by Assumptions SC1(i) and

SC2(ii), and the Markov inequality

$$\tilde{\Omega}_n = \Omega + O_p(n^{-1/2}). \tag{SC.267}$$

By the definition of $\tilde{\Omega}_n$ and $\hat{\Omega}_n$, the triangle inequality and the Cauchy-Schwarz inequality, (SC.263), (SC.265) and (SC.266), we get

$$\begin{aligned}\left|\hat{\Omega}_n - \tilde{\Omega}_n\right| \leq\; & Cn^{-1}\sum_{i=1}^{n}\left((\hat{u}_{2,i} - \hat{\eta}_{1,i}\hat{g}_{1,i})\hat{\varsigma}_{1,i} - (u_{2,i} - \eta_{1,i}g_{1,i})\varsigma_{1,i}\right)^2 \\ & + Cn^{-1}\sum_{i=1}^{n}\left(\hat{\Gamma}_n\hat{\varepsilon}_{1,i}\hat{\eta}_{1,i} - \Gamma\varepsilon_{1,i}\eta_{1,i}\right)^2 + Cn^{-1}\sum_{i=1}^{n}\left(\hat{\eta}_{1,i}\hat{g}_{1,i}\hat{\varsigma}_{2,i} - \eta_{1,i}g_{1,i}\varsigma_{2,i}\right)^2 \\ & + C\tilde{\Omega}_n^{1/2}\left(n^{-1}\sum_{i=1}^{n}\left((\hat{u}_{2,i} - \hat{\eta}_{1,i}\hat{g}_{1,i})\hat{\varsigma}_{1,i} - (u_{2,i} - \eta_{1,i}g_{1,i})\varsigma_{1,i}\right)^2\right)^{1/2} \\ & + C\tilde{\Omega}_n^{1/2}\left(n^{-1}\sum_{i=1}^{n}(\hat{\Gamma}_n\hat{\varepsilon}_{1,i}\hat{\eta}_{1,i} - \Gamma\varepsilon_{1,i}\eta_{1,i})^2\right)^{1/2} \\ & + C\tilde{\Omega}_n^{1/2}\left(n^{-1}\sum_{i=1}^{n}(\hat{\eta}_{1,i}\hat{g}_{1,i}\hat{\varsigma}_{2,i} - \eta_{1,i}g_{1,i}\varsigma_{2,i})^2\right)^{1/2} = o_p(1)\end{aligned}$$

which together with (SC.267) proves the third claim of the Lemma. *Q.E.D.*

SC.5 Partially linear regression

In this subsection, we provide the preliminary estimator of $\hat{\beta}_l$ when $\beta_{l,0}$ is estimated together with $\phi(\cdot)$ in the partially linear regression proposed in Olley and Pakes (1996). Define $\tilde{x}_{1,i} \equiv (l_{1,i}, i_{1,i}, k_{1,i})'$ and $\overline{P}_1(\tilde{x}_{1,i}) \equiv (l_{1,i}, P_1(x_{1,i})')'$. Let $\hat{\beta}_l$ and $\hat{\beta}_{\phi_{pl}}$ be the first element and the last m_1 elements of $\hat{\beta}_1$ respectively, where

$$\hat{\beta}_1 \equiv (\overline{\mathbf{P}}_1'\overline{\mathbf{P}}_1)^{-1}(\overline{\mathbf{P}}_1'\mathbf{Y}_1)$$

where $\overline{\mathbf{P}}_1 \equiv \left(\overline{P}_1(\tilde{x}_{1,1}), \ldots, \overline{P}_1(\tilde{x}_{1,n})\right)'$ and $\mathbf{Y}_1 \equiv (y_{1,1}, \ldots, y_{1,n})'$. The unknown function $\phi(\cdot)$ is estimated by $\hat{\phi}_{pl}(\cdot) \equiv P_1(\cdot)'\hat{\beta}_{\phi_{pl}}$.

Let $\overline{Q}_{m_1} \equiv \mathbb{E}\left[\overline{P}_1(\tilde{x}_{1,1})\overline{P}_1(\tilde{x}_{1,1})'\right]$ and $h_1(x_{1,i}) \equiv \mathbb{E}[l_{1,i}|\, x_{1,i}]$. The following assumptions are needed.

Assumption SC5. *(i) there exist $r_h > 0$ and $\beta_{h_1,m} \in \mathbb{R}^m$ such that $\sup_{x\in\mathcal{X}}|h_{1,m}(x) - h_1(x)| = O(m^{-r_h})$ where $h_{1,m}(\cdot) \equiv P_1(\cdot)'\beta_{h_1,m}$ and $n^{1/2}m_1^{-r_h} = O(1)$; (ii) $C^{-1} \leq \lambda_{\min}(\overline{Q}_{m_1})$ uniformly over m_1.*

Assumption SC5(i) the unknown function $h_1(x_{1,i})$ can be well approximated by the approximating functions $P_1(x_{1,i})$. Assumption SC5(ii) imposes a uniform lower bound on the eigenvalues of $\overline{Q}_{m_1}$. This condition implicitly imposes a identification condition on the unknown parameter $\beta_{l,0}$. That is in Lemma SC28 below, we show that

$$\|l_{1,i} - h_1(x_{1,i})\|_2 \geq C^{-1} \tag{SC.268}$$

which together with (SA.1) implies that

$$\beta_{l,0} = \frac{\mathbb{E}[(l_{1,i} - h_1(x_{1,i}))(y_{1,i} - \mathbb{E}[y_{1,i}| x_{1,i}])]}{\mathbb{E}[|l_{1,i} - h_1(x_{1,i})|^2]}. \tag{SC.269}$$

We shall show below that Assumption SC1(iii) holds

$$\varepsilon_{1,i} = \frac{l_{1,i} - h_1(x_{1,i})}{\mathbb{E}[|l_{1,i} - h_1(x_{1,i})|^2]}\eta_{1,i}. \tag{SC.270}$$

Let $\hat{h}_{1,i} \equiv P_1(x_{1,i})'(\mathbf{P}_1'\mathbf{P}_1)\mathbf{P}_1'\mathbf{L}_1$ where $\mathbf{L}_1 \equiv (l_{1,1}, \ldots, l_{1,n})'$. Then $\varepsilon_{1,i}$ can be estimated by

$$\hat{\varepsilon}_{1,i} \equiv \frac{l_{1,i} - \hat{h}_{1,i}}{n^{-1}\sum_{i=1}^n (l_{1,i} - \hat{h}_{1,i})^2}\hat{\eta}_{1,i} \tag{SC.271}$$

where $\hat{\eta}_{1,i} \equiv y_{1,i} - l_{1,i}\hat{\beta}_l - \hat{\phi}(x_{1,i})$ is defined in Subsection SB.2.

Lemma SC26. *Under Assumptions SC1(i, ii, iv, v, vi) and SC5, we have*

$$\hat{\beta}_l - \beta_{l,0} = n^{-1}\sum_{i=1}^n \frac{l_{1,i} - h_1(x_{1,i})}{\mathbb{E}[|l_{1,i} - h_1(x_{1,i})|^2]}\eta_{1,i} + o_p(n^{-1/2}). \tag{SC.272}$$

Proof of Lemma SC26. First note that we can write $\hat{\beta}_l = (\mathbf{L}_1'\mathbf{M}_1\mathbf{L}_1)^{-1}(\mathbf{L}_1'\mathbf{M}_1\mathbf{Y}_1)$. Therefore

$$\begin{aligned}\hat{\beta}_l - \beta_{l,0} &= (\mathbf{L}_1'\mathbf{M}_1\mathbf{L}_1)^{-1}(\mathbf{L}_1'\mathbf{M}_1\mathbf{Y}_1 - \beta_{l,0}\mathbf{L}_1) \\ &= n^{-1}\sum_{i=1}^n \frac{l_{1,i} - h_1(x_{1,i})}{\mathbb{E}[|l_{1,i} - h_1(x_{1,i})|^2]}\eta_{1,i} + O_p(m_1 n^{-1})\end{aligned} \tag{SC.273}$$

where the second equality is by Lemma SC30(i, ii). The claim in (SC.272) follows by Assumption SC1(vi) and (SC.273).

Lemma SC27. *Under Assumption SC1(ii, v), we have* $\lambda_{\max}(\overline{Q}_{m_1}) \leq C$.

Proof of Lemma SC27. Consider any $b = (b_1, b_2')' \in \mathbb{R}^{m_1+1}$ with $b'b = 1$ where $b_2 \in \mathbb{R}^{m_1}$. Then

$$b'\overline{Q}_{m_1}b = b_1^2\mathbb{E}[l_{1,i}^2] + 2b_1 b'\mathbb{E}[P_1(x_{1,i})l_{1,i}] + b'Q_{m_1}b \leq C + 2b_1 b'\mathbb{E}[P_1(x_{1,i})l_{1,i}] \tag{SC.274}$$

where the second inequality is by Assumption SC1(ii, v). Moreover by Assumption SC1(ii, v)

$$\|\mathbb{E}[P_1(x_{1,i})l_{1,i}]\| \leq \mathbb{E}[l_{1,i}^2] \leq C$$

which together with the Cauchy-Schwarz inequality and (SC.274) implies that $b'\overline{Q}_{m_1}b \leq C$.*Q.E.D.*

Lemma SC28. *Under Assumptions SC1(ii, v) and SC5, we have* $\|l_{1,i} - h_1(x_{1,i})\|_2 \geq C^{-1}$.

PROOF OF LEMMA SC28. By Assumption SC5(ii), there exists a fixed m_c such that

$$\sup_{x\in\mathcal{X}} |h_{1,m}(x) - h_1(x)| \leq (2C)^{-1} \tag{SC.275}$$

for any $m \geq m_c$. Consider any $m \geq m_c$. By the triangle inequality and (SC.275)

$$\|l_{1,i} - h_1(x_{1,i})\|_2 \geq \|l_{1,i} - h_{1,m}(\cdot)\|_2 - (2C^{1/2})^{-1}. \tag{SC.276}$$

Let $\beta^*_{h_1,m} \equiv Q^{-1}_{m_1}\mathbb{E}[P_1(x_{1,i})l_{1,i}]$. Then $P_1(x_{1,i})'\beta^*_{h_1,m}$ is the projection of $l_{1,i}$ on $P_1(x_{1,i})$ under the L_2-norm. Therefore

$$\|l_{1,i} - h_{1,m}(x_{1,i})\|_2 \geq \left\|l_{1,i} - P_1(x_{1,i})'\beta^*_{h_1,m}\right\|_2 \geq (\lambda_{\min}(\overline{Q}_{m_1}))^{1/2} \geq C^{1/2}$$

which together with (SC.276) finishes the proof. *Q.E.D.*

Lemma SC29. *Under Assumption SC1(i, ii, v, vi), we have*

$$\left\|n^{-1}\overline{\mathbf{P}}_1'\overline{\mathbf{P}}_1 - \overline{Q}_{m_1}\right\|_S = O_p((\log m_1)^{1/2}\xi_{0,m_1}n^{-1/2}) = o_p(1) \tag{SC.277}$$

and

$$C^{-1} \leq \lambda_{\min}(n^{-1}\overline{\mathbf{P}}_1'\overline{\mathbf{P}}_1) \leq \lambda_{\max}(n^{-1}\overline{\mathbf{P}}_1'\overline{\mathbf{P}}_1) \leq C \text{ wpa1}. \tag{SC.278}$$

PROOF OF LEMMA SC29. By Assumption SC1(i, ii, v) and the Markov inequality, we have

$$n^{-1}\sum_{i=1} l_{1,i}^2 - \mathbb{E}\left[l_{1,i}^2\right] = O_p(n^{-1/2}) \tag{SC.279}$$

and

$$n^{-1}\sum_{i=1} P_{1,i}l_{1,i} - \mathbb{E}\left[P_{1,i}l_{1,i}\right] = O_p(m_1^{1/2}n^{-1/2}). \tag{SC.280}$$

Let $A_{11,n} = n^{-1}\sum_{i=1} l_{1,i}^2$, $A_{12,n} = A_{21,n}' = n^{-1}\sum_{i=1} l_{1,i}P_{1,i}$ and $A_{22,n} = n^{-1}\mathbf{P}_1'\mathbf{P}_1$. Consider any

$b = (b_1, b_2')' \in \mathbb{R}^{m_1+1}$ with $b'b = 1$ where $b_2 \in \mathbb{R}^{m_1}$. By the Cauchy-Schwarz inequality,

$$\begin{aligned}
&\begin{pmatrix} b_1 \\ b_2 \end{pmatrix}' \begin{pmatrix} A_{11,n} - \mathbb{E}[A_{11,n}] & A_{12,n} - \mathbb{E}[A_{12,n}] \\ A_{21,n} - \mathbb{E}[A_{21,n}] & A_{22,n} - \mathbb{E}[A_{22,n}] \end{pmatrix}^2 \begin{pmatrix} b_1 \\ b_2 \end{pmatrix} \\
&\leq C \left[\|A_{11,n} - \mathbb{E}[A_{11,n}]\|^2 + \|A_{12,n} - \mathbb{E}[A_{12,n}]\|^2 + \|A_{22,n} - \mathbb{E}[A_{22,n}]\|_S^2 \right]
\end{aligned}$$

which combined with (SC.46), (SC.279) and (SC.280) implies that

$$\left\| n^{-1}\overline{\mathbf{P}}_1'\overline{\mathbf{P}}_1 - \overline{Q}_{m_1} \right\|_S = O_p((\log m_1)^{1/2}(\xi_{0,m_1} + m_1^{1/2})n^{-1/2}). \tag{SC.281}$$

By (SC.281) and Assumption SC1(vi), we have

$$\left\| n^{-1}\overline{\mathbf{P}}_1'\overline{\mathbf{P}}_1 - \overline{Q}_{m_1} \right\|_S = o_p(1) \tag{SC.282}$$

which together with Assumption SC1(v) proves (SC.278). *Q.E.D.*

Lemma SC30. *Let $\mathbf{M}_1 \equiv \mathbf{I}_n - \mathbf{P}_1(\mathbf{P}_1'\mathbf{P}_1)^{-1}\mathbf{P}_1'$. Under Assumptions SC1(i, ii, iv, v, vi) and SC5, we have*

(i) $n^{-1}\mathbf{L}_1'\mathbf{M}_1\mathbf{L}_1 = \mathbb{E}[|l_{1,i} - h_1(x_{1,i})|^2] + O_p(m_1^{1/2}n^{-1/2})$;

(ii) $n^{-1}\mathbf{L}_1'\mathbf{M}_1(\mathbf{Y}_1 - \mathbf{L}_1\beta_{l,0}) = n^{-1}\sum_{i=1}^n (l_{1,i} - h_1(x_{1,i}))\eta_{1,i} + o_p(n^{-1/2})$.

Proof of Lemma SC30. (i) By Assumption SC1(ii) and Hölder's inequality,

$$h_1^2(x_{1,i}) = (\mathbb{E}[l_{1,i}|\, x_{1,i}])^2 \leq \mathbb{E}[l_{1,i}^2 \big|\, x_{1,i}] \leq C \tag{SC.283}$$

which together with Assumption SC1(ii) implies that

$$\mathbb{E}[\epsilon_{1,i}^2 \big|\, x_{1,i}] \leq 2\mathbb{E}[l_{1,i}^2 \big|\, x_{1,i}] + 2h_1^2(x_{1,i}) \leq C \tag{SC.284}$$

where $\epsilon_{1,i} \equiv l_{1,i} - h_1(x_{1,i})$. Let $\hat{\beta}_{h_1} \equiv (\mathbf{P}_1'\mathbf{P}_1)^{-1}\mathbf{P}_1'\mathbf{L}_1$. Then

$$\hat{\beta}_{h_1} - \beta_{h_1,m_1} = (\mathbf{P}_1'\mathbf{P}_1)^{-1}\sum_{i=1}^n P_{1,i}\epsilon_{1,i} + (\mathbf{P}_1'\mathbf{P}_1)^{-1}\sum_{i=1}^n P_{1,i}(h_1(x_{1,i}) - h_{1,m_1}(x_{1,i})).$$

Therefore by Assumptions SC1(i, v) and SC5(i), (SC.47) and (SC.284), we obtain

$$
\begin{aligned}
&\left\|\hat{\beta}_{h_1}-\beta_{h_1,m_1}\right\|^2 \\
&\leq 2\left(\sum_{i=1}^{n}\epsilon_{1,i}P_1(x_{1,i})'\right)(\mathbf{P}_1'\mathbf{P}_1)^{-2}\left(\sum_{i=1}^{n}P_1(x_{1,i})\epsilon_{1,i}\right) \\
&+2\left(\sum_{i=1}^{n}(h_{1,m_1}(x_{1,i})-h_1(x_{1,i}))P_1(x_{1,i})'\right)(\mathbf{P}_1'\mathbf{P}_1)^{-2}\left(\sum_{i=1}^{n}P_1(x_{1,i})(h_{1,m_1}(x_{1,i})-h_1(x_{1,i}))\right) \\
&\leq \frac{2\left\|n^{-1}\sum_{i=1}^{n}P_{1,i}\epsilon_{1,i}\right\|^2}{(\lambda_{\min}(n^{-1}\mathbf{P}_1'\mathbf{P}_1))^2}+\frac{2\sum_{i=1}^{n}(h_{1,m_1}(x_{1,i})-h_1(x_{1,i}))^2}{n\lambda_{\min}(n^{-1}\mathbf{P}_1'\mathbf{P}_1)}=O_p(m_1n^{-1}) \qquad \text{(SC.285)}
\end{aligned}
$$

which together with Assumption SC5(i) and (SC.47) further implies that

$$
n^{-1}\sum_{i=1}^{n}(h_1(x_{1,i})-\hat{h}_{1,i})^2=O_p(m_1n^{-1}). \qquad \text{(SC.286)}
$$

By (SC.283) and (SC.286)

$$
\begin{aligned}
&\left|n^{-1}\sum_{i=1}^{n}h_1^2(x_{1,i})-n^{-1}\sum_{i=1}^{n}\hat{h}_{1,i}^2\right| \\
&\leq n^{-1}\sum_{i=1}^{n}(h_1(x_{1,i})-\hat{h}_{1,i})^2 \\
&\quad+\left(n^{-1}\sum_{i=1}^{n}h_1^2(x_{1,i})\right)^{1/2}\left(n^{-1}\sum_{i=1}^{n}(h_1(x_{1,i})-\hat{h}_{1,i})^2\right)^{1/2}=O_p(m_1^{1/2}n^{-1/2}), \qquad \text{(SC.287)}
\end{aligned}
$$

Therefore by the Markov inequality, Assumption SC1(i, ii), (SC.283) and (SC.287)

$$
\begin{aligned}
& n^{-1}\mathbf{L}_1'\mathbf{M}_1\mathbf{L}_1-(\mathbb{E}[l_{1,i}^2]-\mathbb{E}[h_1(x_{1,i})^2]) \\
= \;& n^{-1}\sum_{i=1}^{n}(l_{1,i}^2-\mathbb{E}[l_{1,i}^2])+n^{-1}\sum_{i=1}^{n}(h_1^2(x_{1,i})-\hat{h}_{1,i}^2) \\
& -n^{-1}\sum_{i=1}^{n}(h_1(x_{1,i})^2-\mathbb{E}[h_1(x_{1,i})^2])=O_p(m_1^{1/2}n^{-1/2}). \qquad \text{(SC.288)}
\end{aligned}
$$

Since $\mathbb{E}[|l_{1,i}-h_1(x_{1,i})|^2]=\mathbb{E}[h_1(x_{1,i})^2]$, the first claim of the lemma follows from (SC.288).

(ii) Since $\mathbf{Y}_1-\mathbf{L}_1\beta_{l,0}=\boldsymbol{\phi}+\boldsymbol{\eta}_1$ where $\boldsymbol{\phi}\equiv(\phi_1,\ldots,\phi_n)'$ and $\boldsymbol{\eta}_1\equiv(\eta_{1,1},\ldots,\eta_{1,n})'$, we can write

$$
n^{-1}\mathbf{L}_1'\mathbf{M}_1(\mathbf{Y}_1-\mathbf{L}_1\beta_{l,0})=n^{-1}\mathbf{L}_1'\mathbf{M}_1\boldsymbol{\phi}+n^{-1}\mathbf{L}_1'\mathbf{M}_1\boldsymbol{\eta}_1, \qquad \text{(SC.289)}
$$

Let $\boldsymbol{\phi}_{m_1}\equiv(\phi_{m_1}(x_{1,1}),\ldots,\phi_{m_1}(x_{1,n}))'$. Then $\boldsymbol{\phi}_{m_1}=\mathbf{P}_1\beta_{\phi,m_1}$ and $\mathbf{M}_1\mathbf{P}_1=0$. Therefore by

Assumption SC1(iv, vi) and (SC.47)

$$n^{-1}\boldsymbol{\phi}'\mathbf{M}_1\boldsymbol{\phi} = n^{-1}(\boldsymbol{\phi}-\boldsymbol{\phi}_{m_1})'\mathbf{M}_1(\boldsymbol{\phi}-\boldsymbol{\phi}_{m_1}) \leq n^{-1}\sum_{i=1}^{n}(\phi(x_{1,i})-\phi_{m_1}(x_{1,i}))^2 = O(n^{-1}). \quad \text{(SC.290)}$$

Let $\mathbf{h}_1 \equiv (h_1(x_{1,1}),\ldots,h_1(x_{1,n}))'$. Then by the similar arguments of showing (SC.290), we get

$$n^{-1}\mathbf{h}_1'\mathbf{M}_1\mathbf{h}_1 = O(n^{-1}). \quad \text{(SC.291)}$$

By Assumption SC1(i), (SC.284) and (SC.290)

$$\mathbb{E}\left[\left\|n^{-1}\boldsymbol{\epsilon}_1'\mathbf{M}_1\boldsymbol{\phi}\right\|^2\middle|\{x_{1,i}\}_{i=1}^n\right] = n^{-2}\boldsymbol{\phi}'\mathbf{M}_1\mathbb{E}\left[\boldsymbol{\epsilon}_1\boldsymbol{\epsilon}_1'\middle|\{x_{1,i}\}_{i=1}^n\right]\mathbf{M}_1\boldsymbol{\phi} \leq Cn^{-2}\boldsymbol{\phi}'\mathbf{M}_1\boldsymbol{\phi} = O(n^{-2})$$

which together with the Markov inequality implies that

$$n^{-1}\boldsymbol{\epsilon}_1'\mathbf{M}_1\boldsymbol{\phi} = O_p(n^{-1}). \quad \text{(SC.292)}$$

Similarly, we can show that

$$n^{-1}\mathbf{h}_1'\mathbf{M}_1\boldsymbol{\eta}_1 = O_p(n^{-1}). \quad \text{(SC.293)}$$

Collecting the results in (SC.289), (SC.290), (SC.291), (SC.292) and (SC.293), we obtain

$$n^{-1}\mathbf{L}_1'\mathbf{M}_1(\mathbf{Y}_1-\mathbf{L}_1\beta_{l,0}) = n^{-1}\boldsymbol{\epsilon}_1'\mathbf{M}_1\boldsymbol{\eta}_1 + O_p(n^{-1}). \quad \text{(SC.294)}$$

Since $n^{-1}\sum_{i=1}^n P_{1,i}\epsilon_{1,i} = O_p(m_1^{1/2}n^{-1/2})$ and $n^{-1}\sum_{i=1}^n P_{1,i}\eta_{1,i} = O_p(m_1^{1/2}n^{-1/2})$ by Assumption SC1(i, ii), (SC.284) and the Markov inequality, we can use (SC.47) to deduce that

$$n^{-1}\boldsymbol{\epsilon}_1'\mathbf{P}_1(\mathbf{P}_1'\mathbf{P}_1)^{-1}\mathbf{P}_1'\boldsymbol{\eta}_1 = O_p(m_1n^{-1})$$

which together with (SC.294) proves the second claim of the lemma. *Q.E.D.*

SC.6 Preliminary results

Lemma SC31 (Matrix Bernstein)**.** *Consider a finite sequence $\{d_i\}$ of independent, random matrices with dimension $m_1 \times m_2$. Assume that*

$$\mathbb{E}[d_i] = 0 \text{ and } \|d_i\|_S \leq \xi$$

where ξ is a finite constant. Introduce the random matrix $D_n = \sum_{i=1}^n d_i$. Compute the variance parameter

$$\sigma^2 = \max\left\{\left\|\sum_{i=1}^n \mathbb{E}\left[d_i d_i'\right]\right\|_S, \left\|\sum_{i=1}^n \mathbb{E}\left[d_i' d_i\right]\right\|_S\right\}.$$

Then for any $t \geq 0$

$$\mathbb{P}\left(\|D_n\|_S \geq t\right) \leq (m_1 + m_2)\exp\left(-\frac{t^2/2}{\sigma^2 + \xi t/3}\right).$$

The proof of the above lemma can be found in Tropp (2012).

Lemma SC32. *Let $S_{2,i}(\beta_k) = \tilde{P}_{2,i}(\beta_k)\tilde{P}_{2,i}(\beta_k)'$ where $\tilde{P}_{2,i}(\beta_k) = \tilde{P}_2(\nu_{1,i}(\beta_k), \beta_k)$ for any $\beta_k \in \Theta_k$. Then under Assumptions SC1(i) and SC2(iv, v, vi), we have*

$$\sup_{\beta_k \in \Theta_k}\left\|n^{-1}\sum_{i=1}^n S_{2,i}(\beta_k) - \mathbb{E}\left[S_{2,i}(\beta_k)\right]\right\|_S = O_p((\log(n))^{1/2}\xi_{0,m_2}n^{-1/2}).$$

Proof of Lemma SC32. For any $\beta_k \in \Theta_k$, by the triangle inequality and Assumptions SC2(iv, v),

$$\|S_{2,i}(\beta_k) - \mathbb{E}\left[S_{2,i}(\beta_k)\right]\|_S \leq \|S_{2,i}(\beta_k)\|_S + \|\mathbb{E}\left[S_{2,i}(\beta_k)\right]\|_S \leq C\xi_{0,m_2}^2. \qquad \text{(SC.295)}$$

By Assumptions SC1(i) and SC2(iv, v),

$$\left\|\sum_{i=1}^n \mathbb{E}\left[(S_{2,i}(\beta_k) - \mathbb{E}\left[S_{2,i}(\beta_k)\right])^2\right]\right\|_S \leq n\left(\left\|\mathbb{E}\left[(S_{2,i}(\beta_k))^2\right]\right\|_S + \left\|(\mathbb{E}\left[S_{2,i}(\beta_k)\right])^2\right\|_S\right) \leq Cn\xi_{0,m_2}^2. \qquad \text{(SC.296)}$$

Therefore we can use Lemma SC31 to deduce that

$$\mathbb{P}\left(\left\|n^{-1}\sum_{i=1}^n S_{2,i}(\beta_k) - \mathbb{E}\left[S_{2,i}(\beta_k)\right]\right\|_S \geq t\right) \leq 2m_2\exp\left(-\frac{1}{C}\frac{nt^2/2}{\xi_{0,m_2}^2(1+t/3)}\right) \qquad \text{(SC.297)}$$

for any $\beta_k \in \Theta_k$ and any $t \geq 0$.

Since $k_{1,i}$ has bounded support, there exists a finite constant C_k such that $|k_{1,i}| \leq C_k$ for any i. Consider any $\beta_{k,1}, \beta_{k,2} \in \Theta_k$ and any $b \in \mathbb{R}^{m_2}$ with $\|b\| = 1$. By the triangle inequality,

$$\begin{aligned}\|S_{2,i}(\beta_{k,1}) - S_{2,i}(\beta_{k,2})\|_S \leq& \left\|S_{2,i}(\beta_{k,1}) - \tilde{P}_{2,i}(\beta_{k,2})\tilde{P}_{2,i}(\beta_{k,1})'\right\|_S \\ &+ \left\|\tilde{P}_{2,i}(\beta_{k,2})\tilde{P}_{2,i}(\beta_{k,1})' - S_{2,i}(\beta_{k,2})\right\|_S. \qquad \text{(SC.298)}\end{aligned}$$

By the mean value expansion and the Cauchy-Schwarz inequality, and Assumption SC2(v)

$$\begin{aligned} & \left|b'(\tilde{P}_{2,i}(\beta_{k,1})\tilde{P}_{2,i}(\beta_{k,1})' - \tilde{P}_{2,i}(\beta_{k,2})\tilde{P}_{2,i}(\beta_{k,1})')\right|^2 \\ = & \left\|\tilde{P}_{2,i}(\beta_{k,1})\right\|^2 \left|b'(\tilde{P}_{2,i}(\beta_{k,1}) - \tilde{P}_{2,i}(\beta_{k,2}))\right|^2 \\ = & \left\|\tilde{P}_{2,i}(\beta_{k,1})\right\|^2 \left|b'\partial\tilde{P}_2\left(\nu_{1,i}(\tilde{\beta}_{k,12});\tilde{\beta}_{k,12}\right)/\partial\beta_k\right|^2 (\beta_{k,1} - \beta_{k,2})^2 \\ \leq & \|b\|^2 \xi_{0,m_2}^2 \xi_{1,m_2}^2 (\beta_{k,1} - \beta_{k,2})^2 \end{aligned}$$

where $\tilde{\beta}_{k,12}$ lies between $\beta_{k,1}$ and $\beta_{k,2}$, which together with Assumption SC2(vi) implies that

$$\left\|S_{2,i}(\beta_{k,1}) - \tilde{P}_{2,i}(\beta_{k,2})\tilde{P}_{2,i}(\beta_{k,1})'\right\|_S \leq Cm_2^3 \left|\beta_{k,2} - \beta_{k,1}\right|. \quad \text{(SC.299)}$$

The same upper bound can be established for the second term in the right hand side of the inequality of (SC.298). Therefore,

$$\left\|S_{2,i}(\beta_{k,1}) - S_{2,i}(\beta_{k,2})\right\|_S \leq Cm_2^3 \left|\beta_{k,2} - \beta_{k,1}\right|. \quad \text{(SC.300)}$$

Similarly, we can show that

$$\left\|\mathbb{E}\left[S_{2,i}(\beta_{k,1})\right] - \mathbb{E}\left[S_{2,i}(\beta_{k,2})\right]\right\|_S \leq Cm_2^3 \left|\beta_{k,2} - \beta_{k,1}\right|. \quad \text{(SC.301)}$$

Combining the results in (SC.300) and (SC.301), and applying the triangle inequality, we get

$$\left\| \begin{array}{c} n^{-1}\sum_{i=1}^n \left(S_{2,i}(\beta_{k,1}) - \mathbb{E}\left[S_{2,i}(\beta_{k,1})\right]\right) \\ -n^{-1}\sum_{i=1}^n \left(S_{2,i}(\beta_{k,2}) - \mathbb{E}\left[S_{2,i}(\beta_{k,2})\right]\right) \end{array} \right\|_S \leq C_S m_2^3 \left|\beta_{k,2} - \beta_{k,1}\right| \quad \text{(SC.302)}$$

where C_S is a finite fixed constant. Since the parameter space Θ_k is compact, there exist $\{\beta_k(l)\}_{l=1,\ldots,K_n}$ such that for any $\beta_k \in \Theta_k$

$$\min_{l=1,\ldots,K_n} \left|\beta_k - \beta_k(l)\right| \leq (C_S m_2^3 n^{1/2})^{-1} \quad \text{(SC.303)}$$

where $K_n \leq 2C_S m_2^3 n^{1/2}$. For any $\beta_k \in \Theta_k$, by (SC.302) and (SC.303)

$$\left\|n^{-1}\sum_{i=1}^n S_{2,i}(\beta_k) - \mathbb{E}\left[S_{2,i}(\beta_k)\right]\right\|_S \leq \max_{l=1,\ldots,K_n} \left\|n^{-1}\sum_{i=1}^n S_{2,i}(\beta_k(l)) - \mathbb{E}\left[S_{2,i}(\beta_k(l))\right]\right\|_S + n^{-1/2}. \quad \text{(SC.304)}$$

Therefore for any $B > 1$,

$$\begin{aligned}
&\mathbb{P}\left(\sup_{\beta_k\in\Theta_k}\left\|n^{-1}\sum_{i=1}^{n}S_{2,i}(\beta_k)-\mathbb{E}\left[S_{2,i}(\beta_k)\right]\right\|_S \geq B(\xi_{0,m_2}^2\log(n)n^{-1})^{1/2}\right)\\
&\leq \mathbb{P}\left(\max_{l=1,\ldots,K_n}\left\|n^{-1}\sum_{i=1}^{n}S_{2,i}(\beta_k(l))-\mathbb{E}\left[S_{2,i}(\beta_k(l))\right]\right\|_S \geq (B-1)(\xi_{0,m_2}^2\log(n)n^{-1})^{1/2}\right)\\
&\leq \sum_{l=1}^{K_n}\mathbb{P}\left(\left\|n^{-1}\sum_{i=1}^{n}S_{2,i}(\beta_k(l))-\mathbb{E}\left[S_{2,i}(\beta_k(l))\right]\right\|_S \geq (B-1)(\xi_{0,m_2}^2\log(n)n^{-1})^{1/2}\right)\\
&\leq 2K_n m_2\exp\left(-\frac{B}{C}\frac{\log(n)}{1+(\xi_{0,m_2}^2\log(n)n^{-1})^{1/2}}\right) \qquad \text{(SC.305)}
\end{aligned}$$

where the last inequality is by (SC.297). The claim of the theorem follows from (SC.305) and Assumption SC2(vi). *Q.E.D.*

Lemma SC33. *Let $u_{2,i}(\beta_k) = y_{2,i}^* - k_{2,i}\beta_k - g(\nu_{1,i}(\beta_k),\beta_k)$. Then under Assumptions SC1 and SC2(ii, iii, v, vi), we have*

$$\sup_{\beta_k\in\Theta_k}\left\|n^{-1}\sum_{i=1}^{n}\tilde{P}_{2,i}(\beta_k)u_{2,i}(\beta_k)\right\| = O_p(m_2^{5/4}n^{-1/2}).$$

Proof of Lemma SC33. Define $\pi_n(\beta_k) = n^{-1/2}\sum_{i=1}^{n}\tilde{P}_{2,i}(\beta_k)u_{2,i}(\beta_k)$. For any $\beta_k \in \Theta_k$, by Assumption SC2(i) and (SC.68),

$$\mathbb{E}\left[(u_{2,i}(\beta_k))^4\middle|\,\nu_{1,i}(\beta_k)\right] \leq C\mathbb{E}\left[(y_{2,i}^*)^4 + k_{2,i}^4\middle|\,\nu_{1,i}(\beta_k)\right] + C\left|g(\nu_{1,i}(\beta_k);\beta_k)\right|^4 \leq C. \qquad \text{(SC.306)}$$

For any $\beta_{k,1},\beta_{k,2}\in\Theta_k$, by the i.i.d. assumption and the Cauchy-Schwarz inequality

$$\begin{aligned}
&\mathbb{E}\left[\|\pi_n(\beta_{k,1})-\pi_n(\beta_{k,2})\|^2\right]\\
&= \mathbb{E}\left[\left\|\tilde{P}_{2,i}(\beta_{k,1})u_{2,i}(\beta_{k,1})-\tilde{P}_{2,i}(\beta_{k,2})u_{2,i}(\beta_{k,2})\right\|^2\right]\\
&\leq 2\mathbb{E}\left[(u_{2,i}(\beta_{k,2}))^2\left\|\tilde{P}_{2,i}(\beta_{k,1})-\tilde{P}_{2,i}(\beta_{k,2})\right\|^2\right]\\
&\quad + 2\mathbb{E}\left[\left\|\tilde{P}_{2,i}(\beta_{k,1})\right\|^2(u_{2,i}(\beta_{k,2})-u_{2,i}(\beta_{k,1}))^2\right]. \qquad \text{(SC.307)}
\end{aligned}$$

Consider any $b\in\mathbb{R}^{m_2}$. By the mean value expansion and Assumption SC2(v)

$$\left|b'(\tilde{P}_{2,i}(\beta_{k,1})-\tilde{P}_{2,i}(\beta_{k,2}))\right|^2 = \left|b'\partial\tilde{P}_{2,i}\left(\tilde{\beta}_{k,12}\right)/\partial\beta_k\right|^2(\beta_{k,1}-\beta_{k,2})^2 \leq \|b\|^2\,\xi_{1,m_2}^2(\beta_{k,1}-\beta_{k,2})^2$$

where $\tilde{\beta}_{k,12}$ lies between $\beta_{k,1}$ and $\beta_{k,2}$, which implies that

$$\left\|\tilde{P}_{2,i}(\beta_{k,1}) - \tilde{P}_{2,i}(\beta_{k,2})\right\|^2 \leq \xi_{1,m_2}^2(\beta_{k,1} - \beta_{k,2})^2. \tag{SC.308}$$

Therefore, by (SC.306) and (SC.308),

$$\mathbb{E}\left[(u_{2,i}(\beta_{k,2}))^2 \left\|\tilde{P}_{2,i}(\beta_{k,1}) - \tilde{P}_{2,i}(\beta_{k,2})\right\|^2\right] \leq C\xi_{1,m_2}^2(\beta_{k,2} - \beta_{k,1})^2. \tag{SC.309}$$

By the definition of $u_{2,i}(\beta_k)$, we can write

$$u_{1,i}(\beta_{k,2}) - u_{1,i}(\beta_{k,1}) = g(\nu_{1,i}(\beta_{k,1}), \beta_{k,1}) - g(\nu_{1,i}(\beta_{k,2}), \beta_{k,2}) + k_{2,i}(\beta_{k,2} - \beta_{k,1}).$$

Therefore by Assumption SC2(i, ii, iv), we have

$$\mathbb{E}\left[\left\|\tilde{P}_{2,i}(\beta_{k,1})\right\|^2 (u_{1,i}(\beta_{k,2}) - u_{1,i}(\beta_{k,1}))^2\right] \leq Cm_2(\beta_{k,2} - \beta_{k,1})^2 \tag{SC.310}$$

which together with Assumption SC2(vi), (SC.307) and (SC.309) implies that

$$\left\|\|\pi_n(\beta_{k,1}) - \pi_n(\beta_{k,2})\|\right\|_2 \leq Cm_2^2 \left|\beta_{k,2} - \beta_{k,1}\right| \tag{SC.311}$$

for any $\beta_{k,1}, \beta_{k,2} \in \Theta_k$.

We next use the chaining technique to prove the theorem. The proof follows similar arguments for proving Theorem 2.2.4 in van der Vaart and Wellner (1996). Construct nested sets $\Theta_{k,1} \subset \Theta_{k,2} \subset \cdots \subset \Theta_k$ such that $\Theta_{k,j}$ is a maximal set of points in the sense that for every $\beta_{k,j}$, $\beta'_{k,j} \in \Theta_{k,j}$ there is $\left|\beta_{k,j} - \beta'_{k,j}\right| > 2^{-j}$. Since Θ_k is a compact set, the number of the points in $\Theta_{k,j}$ is less than $C2^j$. Link every point $\beta_{k,j+1} \in \Theta_{k,j+1}$ to a unique $\beta_{k,j} \in \Theta_{k,j}$ such that $|\beta_{k,j+1} - \beta_{k,j}| \leq 2^{-j}$. Let $J_n = \min\{j : 2^{-j} \leq Cm_2^{-3/2}\}$. Consider any positive integer $J > J_n$. Obtain for every $\beta_{k,J+1}$ a chain $\beta_{k,J+1}, \ldots, \beta_{k,J_n}$ that connects it to a point β_{k,J_n} in Θ_{k,J_n}. For arbitrary points $\beta_{k,J+1}$, $\beta'_{k,J+1}$ in $\Theta_{k,J+1}$, by the triangle inequality

$$\begin{aligned}
&\left\|\pi_n(\beta_{k,J+1}) - \pi_n(\beta_{k,J_n}) - \left[\pi_n(\beta'_{k,J+1}) - \pi_n(\beta'_{k,J_n})\right]\right\| \\
&= \left\|\sum_{j=J_n}^{J} \left[\pi_n(\beta_{k,j+1}) - \pi_n(\beta_{k,j})\right] - \sum_{j=J_n}^{J} \left[\pi_n(\beta'_{k,j+1}) - \pi_n(\beta'_{k,j})\right]\right\| \\
&\leq 2\sum_{j=J_n}^{J} \max\left\|\pi_n(\beta_{k,j+1}) - \pi_n(\beta_{k,j})\right\|
\end{aligned} \tag{SC.312}$$

where for fixed j the maximum is taken over all links $(\beta_{k,j+1}, \beta_{k,j})$ from $\Theta_{k,j+1}$ to $\Theta_{k,j}$. Thus the jth maximum is taken over at most $C2^{j+1}$ many links. By Assumption SC2(vi), (SC.311), (SC.312), the triangle inequality and the finite maximum inequality,

$$\begin{aligned}&\left\|\max\left\|\pi_n(\beta_{k,J+1})-\pi_n(\beta_{k,J_n})-\left[\pi_n(\beta'_{k,J+1})-\pi_n(\beta'_{k,J_n})\right]\right\|\right\|_2\\&\leq 2\sum_{j=J_n}^{J}\|\max\|\pi_n(\beta_{k,j+1})-\pi_n(\beta_{k,j})\|\|_2\\&\leq C\sum_{j=J_n}^{J}2^{j/2}\max\|\|\pi_n(\beta_{k,j+1})-\pi_n(\beta_{k,j})\|\|_2\leq Cm_2^2\sum_{j=J_n}^{\infty}2^{-j/2}\leq Cm_2^{5/4}\end{aligned} \tag{SC.313}$$

where β_{k,J_n} and β'_{k,J_n} are the endpoints of the chains starting at $\beta_{k,J+1}$ and $\beta'_{k,J+1}$ respectively. Since the set Θ_{k,J_n} has at most $Cm_2^{3/2}$ many elements, by the finite maximum inequality, the triangle inequality, (SC.306) and Assumption SC2(iv)

$$\left\|\max\left\|\pi_n(\beta_{k,J_n})-\pi_n(\beta'_{k,J_n})\right\|\right\|_2\leq Cm_2^{3/4}\max\|\|\pi_n(\beta_{k,J_n})\|\|_2\leq Cm_2^{5/4}. \tag{SC.314}$$

Therefore, by the triangle inequality, (SC.313) and (SC.314),

$$\begin{aligned}&\left\|\max\left\|\pi_n(\beta_{k,J+1})-\pi_n(\beta'_{k,J+1})\right\|\right\|_2\\&\leq\left\|\max\left\|\pi_n(\beta_{k,J+1})-\pi_n(\beta_{k,J_n})-\left[\pi_n(\beta'_{k,J+1})-\pi_n(\beta'_{k,J_n})\right]\right\|\right\|_2\\&\qquad+\left\|\max\left\|\pi_n(\beta_{k,J_n})-\pi_n(\beta'_{k,J_n})\right\|\right\|_2\leq Cm_2^{5/4}.\end{aligned} \tag{SC.315}$$

Let J go to infinity, by (SC.315) we deduce that

$$\left\|\sup_{\beta_k,\beta'_k\in\Theta_k}\left\|\pi_n(\beta_k)-\pi_n(\beta'_k)\right\|\right\|_2\leq Cm_2^{5/4}. \tag{SC.316}$$

By (SC.314), (SC.316) and the triangle inequality,

$$\left\|\sup_{\beta_k\in\Theta_k}\|\pi_n(\beta_k)\|\right\|_2\leq\left\|\sup_{\beta_k\in\Theta_k}\|\pi_n(\beta_k)-\pi_n(\beta_{k,0})\|\right\|_2+\|\|\pi_n(\beta_{k,0})\|\|_2\leq Cm_2^{5/4} \tag{SC.317}$$

which finishes the proof. *Q.E.D.*

Lemma SC34. *Under Assumptions SC1 and SC2(ii, iii, v, vi), we have*

$$\sup_{\beta_k\in\Theta_k}\left\|n^{-1}\sum_{i=1}^{n}u_{2,i}k_{1,i}\partial^1\tilde{P}_{2,i}(\beta_k)\right\|=O_p(m_2^{5/2}n^{-1/2}).$$

Proof of Lemma SC34. Define $\pi_n(\beta_k)=n^{-1/2}\sum_{i=1}^{n}u_{2,i}k_{1,i}\partial^1\tilde{P}_{2,i}(\beta_k)$ for any $\beta_k\in\Theta_k$. By

Assumptions SC1(i) and Assumption SC2(v, vi), and (SC.88)

$$\sup_{\beta_k \in \Theta_k} \| \| \|\pi_n(\beta_k)\| \| \|_2 \leq C\xi_{1,m_2} \leq Cm_2^2 \tag{SC.318}$$

Moreover for any $\beta_{k,1}$ and $\beta_{k,2}$, we can use similar arguments in showing (SC.309) to obtain

$$\| \| \|\pi_n(\beta_{k,1}) - \pi_n(\beta_{k,2})\| \| \|_2 \leq C\xi_{1,m_2} |\beta_{k,1} - \beta_{k,2}| \leq Cm_2^3 |\beta_{k,1} - \beta_{k,2}| \tag{SC.319}$$

Consider the same nested sets $\Theta_{k,j}$ $(j = 1, 2, \ldots)$ constructed in the proof of lemma SC34. Let $J_n = \min\{j : 2^{-j} \leq Cm_2^{-1}\}$. Then for any positive integer $J > J_n$ using the similar arguments in the proof of Lemma SC33, we obtain

$$\left\| \max \left\| \pi_n(\beta_{k,J+1}) - \pi_n(\beta_{k,J_n}) - \left[\pi_n(\beta'_{k,J+1}) - \pi_n(\beta'_{k,J_n})\right] \right\| \right\|_2 \leq m_2^3 \sum_{j=J_n}^{\infty} 2^{-j/2} \leq Cm_2^{5/2} \tag{SC.320}$$

where β_{k,J_n} and β'_{k,J_n} are the endpoints of the chains starting at $\beta_{k,J+1}$ and $\beta'_{k,J+1}$ respectively. Since the set Θ_{k,J_n} has at most Cm_2 many elements, by the finite maximum inequality, the triangle inequality and (SC.318)

$$\left\| \max \left\| \pi_n(\beta_{k,J_n}) - \pi_n(\beta'_{k,J_n}) \right\| \right\|_2 \leq Cm_2^{1/2} \sup_{\beta_k \in \Theta_k} \| \| \|\pi_n(\beta_k)\| \| \|_2 \leq Cm_2^{5/2}. \tag{SC.321}$$

Then the claim of the lemma follows by applying the chaining arguments in the proof of Lemma SC33. *Q.E.D.*

Lemma SC35. *Under Assumptions SC1 and SC2(ii, iii, v, vi), we have*

$$\sup_{\beta_k \in \Theta_k} \left\| n^{-1} \sum_{i=1}^{n} u_{2,i} \partial^1 \tilde{P}_{2,i}(\beta_k) P_1(x_{1,i})' \right\| = O_p(m_2^{5/2} m_1^{1/2} n^{-1/2}).$$

Proof of Lemma SC35. Define $\pi_n(\beta_k) = n^{-1/2} \sum_{i=1}^{n} u_{2,i} \partial^1 \tilde{P}_{2,i}(\beta_k) P_1(x_{1,i})'$ for any $\beta_k \in \Theta_k$. By Assumptions SC1(i) and Assumption SC2(v, vi), and (SC.88)

$$\sup_{\beta_k \in \Theta_k} \| \| \|\pi_n(\beta_k)\| \| \|_2 \leq C\xi_{1,m_2} m_1^{1/2} \leq Cm_1^{1/2} m_2^2. \tag{SC.322}$$

Moreover for any $\beta_{k,1}$ and $\beta_{k,2}$, we can use similar arguments in showing (SC.309) to obtain

$$\| \| \|\pi_n(\beta_{k,1}) - \pi_n(\beta_{k,2})\| \| \|_2 \leq Cm_1^{1/2} m_2^3 |\beta_{k,1} - \beta_{k,2}|. \tag{SC.323}$$

Consider the same nested sets $\Theta_{k,j}$ $(j = 1, 2, \ldots)$ constructed in the proof of lemma SC34. Let

$J_n = \min\{j : 2^{-j} \leq Cm_2^{-1}\}$. Then for any positive integer $J > J_n$ using the similar arguments in the proof of Lemma SC33, we obtain

$$\left\|\max\left\|\pi_n(\beta_{k,J+1}) - \pi_n(\beta_{k,J_n}) - \left[\pi_n(\beta'_{k,J+1}) - \pi_n(\beta'_{k,J_n})\right]\right\|\right\|_2 \leq Cm_1^{1/2}m_2^{5/2} \quad \text{(SC.324)}$$

where β_{k,J_n} and β'_{k,J_n} are the endpoints of the chains starting at $\beta_{k,J+1}$ and $\beta'_{k,J+1}$ respectively. Since the set Θ_{k,J_n} has at most Cm_2 many elements, by the finite maximum inequality, the triangle inequality and (SC.322)

$$\left\|\max\left\|\pi_n(\beta_{k,J_n}) - \pi_n(\beta'_{k,J_n})\right\|\right\|_2 \leq Cm_2^{1/2} \sup_{\beta_k \in \Theta_k} \|\|\pi_n(\beta_k)\|\|_2 \leq Cm_1^{1/2}m_2^{5/2}. \quad \text{(SC.325)}$$

Then the claim of the lemma follows by applying the chaining arguments in the proof of Lemma SC33. *Q.E.D.*

Lemma SC36. *Under Assumptions SC1 and SC2(ii, iii, v, vi), we have*

$$\sup_{\beta_k \in \Theta_k} \left\| n^{-1} \sum_{i=1}^{n} u_{2,i}(\phi_{m_2}(x_{1,i}) - \phi(x_{1,i}))\partial^1 \tilde{P}_{2,i}(\beta_k) \right\| = O_p(m_2^{5/2}n^{-1}).$$

Proof of Lemma SC36. Define $\pi_n(\beta_k) = n^{-1/2} \sum_{i=1}^n u_{2,i}(\phi_{m_2}(x_{1,i}) - \phi(x_{1,i}))\partial^1 \tilde{P}_{2,i}(\beta_k)$ for any $\beta_k \in \Theta_k$. By Assumptions SC1(i) and Assumption SC2(iii, v, vi), and (SC.88)

$$\sup_{\beta_k \in \Theta_k} \|\|\pi_n(\beta_k)\|\|_2 \leq C\xi_{1,m_2} n^{-1/2} \leq Cm_2^2 n^{-1/2}. \quad \text{(SC.326)}$$

Moreover for any $\beta_{k,1}$ and $\beta_{k,2}$, we can use similar arguments in showing (SC.309) to obtain

$$\|\|\pi_n(\beta_{k,1}) - \pi_n(\beta_{k,2})\|\|_2 \leq Cm_2^3 n^{-1/2} \left|\beta_{k,1} - \beta_{k,2}\right|. \quad \text{(SC.327)}$$

Consider the same nested sets $\Theta_{k,j}$ $(j = 1, 2, \ldots)$ constructed in the proof of lemma SC34. Let $J_n = \min\{j : 2^{-j} \leq Cm_2^{-1}\}$. Then for any positive integer $J > J_n$ using the similar arguments in the proof of Lemma SC33, we obtain

$$\left\|\max\left\|\pi_n(\beta_{k,J+1}) - \pi_n(\beta_{k,J_n}) - \left[\pi_n(\beta'_{k,J+1}) - \pi_n(\beta'_{k,J_n})\right]\right\|\right\|_2 \leq Cm_2^{5/2}n^{-1/2} \quad \text{(SC.328)}$$

where β_{k,J_n} and β'_{k,J_n} are the endpoints of the chains starting at $\beta_{k,J+1}$ and $\beta'_{k,J+1}$ respectively. Since the set Θ_{k,J_n} has at most Cm_2 many elements, by the finite maximum inequality, the triangle inequality and (SC.326)

$$\left\|\max\left\|\pi_n(\beta_{k,J_n}) - \pi_n(\beta'_{k,J_n})\right\|\right\|_2 \leq Cm_2^{1/2} \sup_{\beta_k \in \Theta_k} \|\|\pi_n(\beta_k)\|\|_2 \leq Cm_2^{5/2}n^{-1/2}. \quad \text{(SC.329)}$$

Then the claim of the lemma follows by applying the chaining arguments in the proof of Lemma SC33. *Q.E.D.*

References

Ackerberg, D. A., K. Caves, and G. Frazer (2015): "Identification properties of recent production function estimators," *Econometrica*, 83(6), 2411–2451.

Andrews, D. W. K. (1991): "Asymptotic Normality of Series Estimators for Nonparametric and Semiparametric Regression Models," *Econometrica*, 59(2), 307–345.

Belloni, A., V. Chernozhukov, D. Chetverikov, and K. Kato (2015): "Some new asymptotic theory for least squares series: Pointwise and uniform results," *Journal of Econometrics*, 186(2), 345 – 366, High Dimensional Problems in Econometrics.

Chen, X. (2007): "Large Sample Sieve Estimation of Semi-Nonparametric Models," in *Handbook of Econometrics*, ed. by J. Heckman, and E. Leamer, vol. 6B, chap. 76. Elsevier, 1 edn.

Hahn, J., Z. Liao, and G. Ridder (2021): "Identification and the Influence Function of Olley and Pakes' (1996) Production Function Estimator," *Working Paper, UCLA and USC*.

Newey, W. K. (1994): "The asymptotic variance of semiparametric estimators," *Econometrica: Journal of the Econometric Society*, pp. 1349–1382.

Newey, W. K. (1997): "Convergence rates and asymptotic normality for series estimators," *Journal of Econometrics*, 79(1), 147 – 168.

Olley, G. S., and A. Pakes (1996): "The Dynamics of Productivity in the Telecommunications Equipment Industry," *Econometrica: Journal of the Econometric Society*, pp. 1263–1297.

Tropp, J. A. (2012): "User-Friendly Tail Bounds for Sums of Random Matrices," *Foundations of Computational Mathematics*, 12(4), 389–434.

van der Vaart, A., and J. Wellner (1996): *Weak Convergence and Empirical Processes*. Springer-Verlag.