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

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

First Version: March 2020; This Version: October 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. 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 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

[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 di culty disappears if the second step is completely parametric, which is not the specification in OP. Cattaneo, Jansson and Ma (2019) adopt such a parametric speci cation 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.

more complicated, than the final step in HR, and requires special attention.[3]

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.

HLR's standard error is calculated using the pre-asymptotic sieve variance which converges to the asymptotic variance as the number of sieve basis functions goes to infinity. Therefore despite the convenience, HLR does not provide the influence function and the asymptotic of the two-step sieve estimator. Moreover, HLR's approach on standard error calculation is not applicable when nonparametric estimation is done by local methods, such as kernel estimation.[4] 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 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.

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

[4] Both kernel and series methods are used in OP.

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.[5] 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.[6] The production function takes the form[7]

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

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

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

where i_t is the log of investment at time t and δ denotes the capital depreciation factor. 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

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

[5] OP themselves did not name the assumption.

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

[7] (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 $J_t \equiv \sigma\{\omega_t, k_t\}$, where $\sigma\{\omega_t, k_t\}$ denotes the sigma field generated by ω_t and 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 (OP 6) as

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

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[8] implies by the first expression on p. 1276 of OP, that $g\left(\underline{\omega}_{t+1}(k_{t+1}), \omega_t\right) = \mathbb{E}\left[\omega_{t+1}|\,\omega_t\right] + \beta_0 \equiv g(\omega_t)$ (substitute $\underline{\omega}_t(k_t) = -\infty$). Therefore their equations (11) and (12) can be rewritten[9]

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

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

where

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

OP's estimator is based on the following multi-step identification strategy:[10]

[8] 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.

[9] 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.

[10] Specific estimators of $\beta_{l,0}$, $\beta_{k,0}$, $\phi_t(\cdot)$ and $g(\cdot)$ constructed using the nonparametric series method can be found in Section SA of the Supplemental Appendix of the paper.

1. In the first step, $\beta_{l,0}$ and $\phi_t(\cdot)$ 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

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

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

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

where we substitute $\beta_{l,0}$ and $\phi_t\left(i_t, k_t\right)$ 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 under OP's Markov assumption on the ω_t process the arguments in the left-hand side (LHS) and the RHS of (OP 11) are not the same in general, 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 and k_t. Therefore, the LHS of (OP 11) must be $\mathbb{E}\left[y_{t+1} - \beta_{l,0} l_{t+1} \middle| J_t\right]$. 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 $\phi_t\left(i_t, k_t\right) = \mathbb{E}\left[y_t \middle| i_t, k_t\right] - \beta_l \mathbb{E}\left[l_t \middle| i_t, k_t\right]$. 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 $\mathbb{E}\left[\eta_{t+1} \middle| J_t\right] = 0$,

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

This suggests that (OP 11) should be read as

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

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

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

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

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

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

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

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 and it is not a state variable in (OP 3), and (ii) $\mathbb{E}[\omega_{t+1}|\,\omega_t, k_t, u_t] = \mathbb{E}[\omega_{t+1}|\,\omega_t, k_t]$.[13]

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

[13] Note that in this subsection, J_t denotes the generic information set at time t and therefore $\sigma\{\omega_t, k_t\} \subseteq J_t$.

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.[14]

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)

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

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.

The problem is easiest to understand under the assumption that $\underline{\omega}_{t+1}(k_{t+1})$ is a strictly decreasing function of k_{t+1} (see OP, p. 1274). If so, $\bar{g}(\omega_t, k_{t+1}) \equiv g\left(\underline{\omega}_{t+1}(k_{t+1}), \omega_t\right)$ is strictly decreasing in k_{t+1} because $g\left(\underline{\omega}_{t+1}(k_{t+1}), \omega_t\right)$ is a strictly increasing function of $\underline{\omega}_{t+1}(k_{t+1})$ for any given ω_t.

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.

[14] 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

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

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.[15]

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:

[15]Throughout the rest of this section, we assume that one can switch the order of expectation and differentiation. This assumption can be justified under the Dominated Convergence Theorem and some regularity conditions (such as bounded Sobolev norm) on the functions to be differentiated.

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

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

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 $\gamma_j\left(\phi_t - \beta_k k_t\right) \equiv \mathbb{E}\left[\left. Y_j \right| \phi_t - \beta_k k_t\right]$ 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 $\mathbb{E}\left[\left. Y_j \right| \phi_t - \beta_k k_t\right]$ 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).[16] Newey shows that if an infinite 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).

[16] 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).

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

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

The conditional expectation $\mathbb{E}[Y_j | \nu_t(\beta_k)]$ 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

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

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)[17] 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.[18] 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*

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

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.

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

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

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

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

The next proposition gives the influence function of OP's estimator of $\beta_{k,0}$.

Proposition 3 *The influence function of OP's estimator of* $\beta_{k,0}$ *is the sum of the main term that is the normalized sum of the moment functions in (24) in the Appendix:*

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

and the adjustment

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

where

$$\begin{aligned}\Lambda_{1i} &\equiv \frac{\partial g(\nu_{i,t})}{\partial \nu_{i,t}}\left((k_{i,t+1} - \mathbb{E}[k_{i,t+1}|\, \nu_{i,t}]) - \frac{\partial g(\nu_{i,t})}{\partial \nu_{i,t}}(k_{i,t} - \mathbb{E}[k_{i,t}|\, \nu_{i,t}])\right), \\ \Lambda_{2i} &\equiv \frac{\Gamma}{\mathbb{E}\left[(l_{i,t} - \mathbb{E}[l_{i,t}|\, i_{i,t}, k_{i,t}])^2\right]}(l_{i,t} - \mathbb{E}[l_{i,t}|\, i_{i,t}, k_{i,t}]),\end{aligned}$$

and

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

Remark 1 *Our decomposition of the influence function is helpful in case* $\beta_{l,0}$ *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*

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

Remark 2 *The influence function provided in Proposition 3 makes it straightforward to estimate the standard error of OP's estimator. First, one can estimate* Υ *and* Γ *by their sample analogs and replacing the unknown functions such as* $\mathbb{E}[k_t|\, \nu_t = v]$, $\mathbb{E}[k_{t+1}|\nu_t = v]$ *and*

$\partial g(v)/\partial \nu_t$ by some nonparametric (kernel or series) estimators. Denote the resulting estimators of Υ and Γ by $\hat{\Upsilon}_n$ and $\hat{\Gamma}_n$, respectively. Second, the moment functions in the summation of the main term (16) can be estimated by replacing $\beta_{l,0}$ and $\beta_{k,0}$ by their consistent estimators, and replacing the unknown functions such as $\mathbb{E}[k_t|\nu_t = v]$, $\mathbb{E}[k_{t+1}|\nu_t = v]$ and $\partial g(v)/\partial \nu_t$ by some nonparametric estimators. Denote this moment function estimator by $\hat{h}_{1,i}$ for any $i = 1, \ldots, n$. The moment functions in the summation of the adjustment term (17) can be estimated similarly and its estimator for any $i = 1, \ldots, n$ is denoted as $\hat{h}_{2,i}$. The asymptotic variance of OP's estimator of $\beta_{k,0}$ is then given by $\hat{\Upsilon}_n^{-1}\hat{\Omega}_n\hat{\Upsilon}_n^{-1}$ where $\hat{\Omega}_n = n^{-1}\sum_{i=1}^{n}(\hat{h}_{1,i} - \hat{h}_{2,i})^2$.[19]

Remark 3 *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, where one can also find a specific form of the estimator of $\beta_{k,0}$ and its asymptotic properties such as consistency and root-n asymptotic normality. Consistent estimation of the asymptotic variance of OP's estimator of $\beta_{k,0}$, and hence its valid inference are also provided in the Supplemental Appendix.*

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.

[19] More details on constructing the asymptotic variance estimator can be found in Subsection SB.2 of the Supplemental Appendix.

[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

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

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)

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

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:[20]

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

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

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

   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

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

   where

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

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

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

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

which is equal to

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

with

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

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

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

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

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

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

   (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 $\beta_{l,0}$:

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

   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

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

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

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

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

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

that is equal to

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

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

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 $\mathbb{E}\left[y_t\middle|\,i_t,k_t\right]$ and $\mathbb{E}\left[l_t\middle|\,i_t,k_t\right]$, 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

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

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

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

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

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

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

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

with the adjustment calculated by Newey (1994)

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

It is straightforward to show that

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

Therefore (26) is equal to

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

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

Note that

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

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

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

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

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

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

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

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

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

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 $\mathbb{E}\left[l_t\middle|\, i_t,k_t\right]$

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

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

so that the adjustment is

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

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

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

we can use

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

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

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