Abstract

We revisit the production function estimators of Olley and Pakes (1996) and Levinsohn and Petrin (2003). They use control functions to address the simultaneous determination of inputs and productivity. Both assume that input demand is a monotonic function of productivity holding capital constant and then invert this function to condition on productivity during estimation. If the observed capital variable is measured with error, input demand will not generally be monotonic in the productivity shock holding observed capital constant. We develop consistent estimators of production function parameters in the face of this measurement error. Our identification and estimation results combine the nonlinear measurement error literature with Wooldridge (2009)'s joint estimation method to construct a proxy for productivity that addresses simultaneity. Our approach directly extends to the case where other inputs like intermediates or labor are observed with error.

Keywords: Production Function, Unobserved Productivity, Measurement Error, Nonparametric Estimation, Control Variate

JEL Classification: C14, C18, D24

*We thank Jinyong Hahn, Bruce Hansen, Yingyao Hu, Arthur Lewbel, Rosa Matzkin, and other seminar participants at UCLA, NSF-NBER CEME conference at Cornell University, 2013 North American Summer Meeting of Econometric Society at USC, IO conference at University of Tokyo, European Meeting of Econometric Society at Gothenburg, Midwest Econometrics Group at Bloomington for many helpful comments. Suyong Song gratefully acknowledges research grant from Yoshio Niho Excellence Fund. All errors are our own. Contacts: Kyoo il Kim at kyookim@msu.edu; Tel. 517-353-9008, Amil Petrin at petrin@umn.edu, and Suyong Song at suyong-song@uiowa.edu; Tel. 319-335-0832.
1 Introduction

Production function estimates are a critical input into understanding the sources of economic growth. A major challenge in estimating production function parameters is that output and inputs are simultaneously determined (Marschak and Andrews 1944). This simultaneity problem results in inputs being correlated with the productivity shock, leading to biased and inconsistent estimates of the parameters.

Historically researchers have used firm-specific fixed effects in panel data to address this problem, relying only on within-firm variation in inputs over time to identify the parameters. In their review of the literature Griliches and Mairesse (1998) remark that fixed effect estimators frequently have led researchers to find point estimates for the capital coefficient that are very low and often not significantly different from zero. They attribute these findings to capital being measured with a significant amount of error coupled with the fact that capital varies very little within firms over the time in a standard panel data set. The fixed effect effectively dispenses with all of the “signal,” leading to a significant amount of attenuation bias in the capital coefficient.

A major contribution of Olley and Pakes (1996; hereafter OP) is their development of an estimation method that is both robust to the simultaneity problem and that does not dispense with all of the between-firm variation in capital. They assume that investment is a function of the capital and the productivity shock, the two state variables. Pakes (1996) provides conditions under which - for any given level of capital - investment is monotonically increasing in the productivity shock. By inverting the investment function they recover the productivity shock, on which they then condition during estimation. Unlike fixed effects the approach has the added advantage that the productivity shock is allowed to vary over time. Since its publication many researchers have confirmed that the OP approach generally leads to higher and more reasonable point estimates for capital than the fixed effect estimator.

One issue that has arisen for the OP estimator is that the inversion is only valid when investment is positive. Levinsohn and Petrin (2003; hereafter LP) note that in panel data many firms report zero investment, which can force researchers to drop a large fraction of their observations from the estimation procedure. In the Chilean data they use investment is missing for between 30% and 70% of their observations depending upon which industry was under consideration. The fact that investment is lumpy in firm-level data also may suggest that investment does not fully respond to the productivity shock even when investment is positive.
In this paper we show how to use recent insights from Hu and Schennach (2008) on nonclassical measurement error to allow for measurement error in capital and other inputs in the LP setting. To achieve identification we recast the conditional moment restrictions used in LP as integrals over conditional density functions of the true inputs. Using lagged input levels and input prices as instruments, Hu and Schennach’s method can be used to estimate these conditional density functions when capital and possibly other inputs are measured with errors. Once we have estimates of the conditional density functions of the true inputs we can integrate over them to calculate the LP conditional moment restrictions for any candidate parameter value. Our two-step estimator first uses sieve maximum likelihood (hereafter ML) estimation to estimate the conditional density functions. In the second step we use Ai and Chen (2003) to develop a sieve minimum distance (hereafter MD) estimator, integrating the conditional moment restrictions over the measurement-error-corrected conditional density estimates.2

The closest work to our paper is Huang and Hu (2011). They extend the LP setting to the case when intermediate inputs are measured with an additive error. We extend their approach to settings where both the intermediate input and capital can be contaminated by additive or non-additive measurement error.

The remainder of this paper is organized as follows. In Section 2 we review the OP and LP estimators and provide a non-technical overview of our estimation approach. In Section 3 we use the framework from Wooldridge (2009) to develop the general framework for our approach, including the case when both capital and intermediates are measured with error. Section 4 discusses identification and estimation and Sections 5 and 6 provide details on consistency, convergence rates, and asymptotic normality. Section 7 concludes and the technical details are gathered in the Appendix.

2 Allowing for Measurement Error

We review the OP and LP methodologies within the Wooldridge (2009) framework and show where measurement error in capital causes problems. We then show how to use recent results from the nonclassical measurement error literature to correct the problem.

2.1 OP/LP Methodology

The production function is written with the log of output as a function of the log of inputs and shocks

\[ y_t = \beta_l l_t + \beta_k k_t + \beta m m_t + \omega_t + \epsilon_t \]

where \( l_t \) denotes labor, \( k_t \) denotes capital, and \( m_t \) denotes the intermediate input (such as materials or energy). \( \omega_t \) is the productivity shock, a state variable observed by the firm but unobserved to the econometrician and assumed to be a first-order Markov. \( \omega_t \) is the source of the simultaneity

\[ ^2 \text{This step extends Ai and Chen (2003) to a setting with unobserved endogenous variables.} \]
problem as freely variable inputs $l_t$ and $m_t$ respond to it. $k_t$ is a state variable and is allowed to be correlated with $E[\omega_t|\omega_{t-1}]$, but it is assumed that $\xi_t = \omega_t - E[\omega_t|\omega_{t-1}]$, the innovation in the productivity shock, is uncorrelated with $k_t$. $\epsilon_t$ denotes an i.i.d. shock that is assumed to be uncorrelated with all of the inputs.

LP write intermediate input demand as a function of the state variables

$$m_t = m_t(\omega_t, k_t)$$

and provide weak conditions under which $m_t(\cdot, \cdot)$ is strictly monotonic in $\omega_t$ holding $k_t$ constant. The intermediate demand function can then be inverted to obtain the control function for $\omega_t$ as a function of observed $m_t$ and $k_t$, written as $\omega_t = h_t(m_t, k_t).$ Wooldridge (2009) uses a single index restriction to approximate unobserved productivity, so in the LP setting one has

$$\omega_t = h_t(m_t, k_t) = c(m_t, k_t)' \beta_{\omega}$$

where $c(m_t, k_t)$ is a known vector function of $(m_t, k_t)$ chosen by researchers. He also writes the nonparametric conditional mean function $E[\omega_t|\omega_{t-1}]$ as

$$E[\omega_t|\omega_{t-1}] = q(c(m_{t-1}, k_{t-1})' \beta_{\omega})$$

for some unknown function $q(\cdot)$.

Rewriting the production function as

$$y_t = \beta_0 l_t + \beta_k k_t + \beta_m m_t + E[\omega_t|\omega_{t-1}] + \xi_t + \epsilon_t$$

yields

$$[\xi_t + \epsilon_t](\theta) = y_t - \beta_0 l_t - \beta_k k_t - \beta_m m_t - q(c(m_{t-1}, k_{t-1})' \beta_{\omega})$$

with $\beta = (\beta_0, \beta_k, \beta_m, \beta_{\omega})$, $\theta = (\beta, q)$. Let the set of conditioning variables be $x_t$ that include the current $k_t$ and other lagged variables, which are uncorrelated with both $\xi_t$ and $\epsilon_t$ and let $\theta_0$ denote the true parameter value. Wooldridge shows that the conditional moment restriction

$$g(x_t; \theta) \equiv E[|\xi_t + \epsilon_t|(\theta)|x_t]$$

and $g(x_t; \theta_0) = 0$

is sufficient for identification of $(\beta_0, \beta_k, \beta_m)$ and $E[\omega_t|\omega_{t-1}]$. It is also robust to the Ackerberg, Caves, and Frazer (2006) criticism of LP (and OP). In equation (1) a function of $m_{t-1}$ and $k_{t-1}$ conditions out $E[\omega_t|\omega_{t-1}]$. $\xi_t$ is not correlated with $k_t$, so $k_t$ can serve as an instrument for itself.

---

3OP write investment as a function of the two state variables $i_t = i_t(\omega_t, k_t)$ and Pakes (1996) provides conditions under which investment is strictly monotonic in $\omega_t$ holding $k_t$ constant. OP then invert this function to get the control function with arguments $i_t$ and $k_t$.  
4OP use $i_t$ and $i_{t-1}$ instead of $m_t$ and $m_{t-1}$, respectively, for $\omega_t$ and $E[\omega_t|\omega_{t-1}]$.  
5In Section 3 we show how to identify $q$ and $\beta_{\omega}$ using an additional moment condition from LP.
Other lagged variables serve as instruments for \( l_t \) and \( m_t \). It is also clear from this moment why measurement error in \( k_t \) is problematic for the LP (and OP) estimators as both current and lagged values of capital are used in the estimation, and the lagged values enter the estimation problem in a nonlinear way.

2.2 Solution Overview

When we move to the setting with measurement error we will require additional instruments that are excluded from the production function but enter the input demands to satisfy the order condition. Given that some form of input prices are often observed we add them to the list of instruments. In the specifics of our setting they shift input demand conditional on productivity and capital. We write \( p_t \) as the vector of input prices at time \( t \) and include it in the intermediate input demand function

\[ m_t = \mathbf{m}_t(\omega_t, k_t, p_t), \]

and similarly for labor denoted by \( l_t = l_t(\omega_t, k_t, p_t) \). The original proof of monotonicity from LP is written with input prices as arguments in the input demand equation so monotonicity follows directly with \( m_t(\cdot, \cdot, \cdot) \) monotonic in \( \omega_t \) holding \( k_t \) and \( p_t \) constant. The single index restriction becomes

\[ \omega_t = h_t(m_t, k_t, p_t) = \mathbf{c}(m_t, k_t, p_t)'\beta_{\omega} \]

where \( \mathbf{c}(m_t, k_t, p_t) \) is a known vector function of \( (m_t, k_t, p_t) \) chosen by researchers. We then write \( E[\omega_t|\omega_{t-1}] \) as

\[ E[\omega_t|\omega_{t-1}] = q(\mathbf{c}(m_{t-1}, k_{t-1}, p_{t-1})'\beta_{\omega}) \]

for some unknown function \( q(\cdot) \). The new residual is

\[ [\xi_t + \epsilon_t](\theta) = y_t - \beta_l l_t - \beta_k k_t - \beta_m m_t - q(\mathbf{c}(m_{t-1}, k_{t-1}, p_{t-1})'\beta_{\omega}). \]

The new set of conditioning variables is given as \( x_t = (k_t, k_{t-1}, p_t, p_{t-1}, m_{t-1}) \). The conditional moment restriction is then

\[ g(x_t; \theta) \equiv E[[\xi_t + \epsilon_t](\theta)|x_t] \text{ and } g(x_t; \theta_0) = 0, \tag{2} \]

where now a function of \( m_{t-1}, k_{t-1}, \) and \( p_{t-1} \) conditions out \( E[\omega_t|\omega_{t-1}] \). \( \xi_t \) is not correlated with \( k_t \), so \( k_t \) can serve as an instrument for itself. Current input prices instrument for \( l_t \) and \( m_t \) and the order condition is satisfied as long as any two input prices are observed.

We rewrite the estimation problem in terms of the conditional densities upon which it is based.

\[ ^{6}\text{If either current or lagged output prices are exogenous they will also affect input demand and can therefore also act as instruments. See the discussion in Doraszelski and Jaumandreu (2013) who suggest the use of past input and output prices as instruments.} \]

\[ ^{7}\text{In this setup labor also qualifies as a possible proxy variable.} \]
For $g(x_t; \theta)$ we have:

$$g(x_t; \theta) \equiv \int y_t f_{y_t|x_t} dy_t - \beta_t \int l_t f_{l_t|x_t} dl_t - \beta_t k_t$$

$$- \beta_m \int m_t f_{m_t|x_t} dm_t - q(c(m_{t-1}, k_{t-1}, p_{t-1})')$$

and for the population objective function we have e.g.,

$$Q(\theta) = E \left[ g(x_t; \theta)^2 \right] = \int g(x_t; \theta)^2 f_{k_t, k_{t-1}|p_t, p_{t-1}, m_{t-1}} f_{p_t, p_{t-1}, m_{t-1}} dx_t$$

(3)

(using Bayes Rule). Written in this way the OP/LP estimation problem can be viewed as one that requires estimates of the conditional densities $f_{y_t|x_t}$, $f_{l_t|x_t}$, and $f_{m_t|x_t}$ to estimate $g(x_t; \theta)$ and then estimates of $f_{k_t, k_{t-1}|p_t, p_{t-1}, m_{t-1}}$ and $f_{p_t, p_{t-1}, m_{t-1}}$ to integrate to get the sample analog of $Q(\theta)$.

Letting capital that is measured with error be denoted as $k_t^*$, and $x_t^* = (k_t^*, k_{t-1}^*, p_t, p_{t-1}, m_{t-1})$, our main challenge is recovering the true densities given that the observed densities are $f_{y_t|x_t^*}$, $f_{l_t|x_t^*}$, $f_{m_t|x_t^*}$, and $f_{k_t^*, k_{t-1}^*|p_t, p_{t-1}, m_{t-1}}$ ($f_{p_t, p_{t-1}, m_{t-1}}$ is directly observable from the data).

If we use $(k_{t-2}^*, k_{t-3}^*)$ as instruments for the current and lagged mismeasured capital measurements $(k_t^*, k_{t-1}^*)$, the true conditional densities can be recovered using recent developments in the nonclassical measurement error literature. In the data we observe the conditional density functions $f_{r_t, k_t^*, k_{t-1}^*|k_{t-2}^*, k_{t-3}^*, s_t}$ for either (i) $r_t = y_t$ and $s_t = (m_t, l_t, p_t, p_{t-1}, m_{t-1})$ or (ii) $r_t \in \{l_t, m_t\}$ and $s_t = (p_t, p_{t-1}, m_{t-1})$. We can express these observed densities $f_{r_t, k_t^*, k_{t-1}^*|k_{t-2}^*, k_{t-3}^*, s_t}$ without loss of generality as

$$f_{r_t, k_t^*, k_{t-1}^*|k_{t-2}^*, k_{t-3}^*, s_t} = \int f_{r_t|k_t, k_{t-1}, k_{t-2}^*, k_{t-3}^*, s_t} f_{k_t|x_t} f_{k_{t-1}|x_t} f_{k_{t-2}^*|x_t} f_{k_{t-3}^*|x_t} f_{s_t|k_t, k_{t-1}, k_{t-2}^*, k_{t-3}^*, s_t} dk_t dk_{t-1}$$

(4)

where the first equality follows from the properties of densities and the second and third equalities follow from Bayes Rule.

Our key identification assumption is that current and lagged mismeasured capital inputs do not provide additional information on output and other inputs beyond what the true capital inputs do. This implies that the first term under the integral in the last line of equation (4) can be written as $f_{r_t|k_t, k_{t-1}, s_t}$ and similarly the second term can be written as $f_{k_t^*, k_{t-1}^*|k_t, k_{t-1}, s_t}$. Thus the exclusion restriction allows us to re-express the last line of equation (4) as

$$f_{r_t, k_t^*, k_{t-1}^*|k_{t-2}^*, k_{t-3}^*, s_t} = \int f_{r_t|k_t, k_{t-1}, s_t} f_{k_t^*, k_{t-1}^*|k_t, k_{t-1}, s_t} f_{k_{t-2}^*|k_t, k_{t-1}, s_t} f_{k_{t-3}^*|k_t, k_{t-1}, s_t} dk_t dk_{t-1}.$$
Under weak regularity conditions and a standard rank condition for nonparametric instrumental variables we show in Section 4 that this equation has unique solutions for $f_{x_1|k_t,k_{t-1},s_t}$, $f_{x_1^*|k_t^*,k_{t-1}^*,s_t}$, and $f_{k_t,k_{t-1}|k_t^*,k_{t-1}^*,s_t}$ following similar arguments to Hu and Schennach (2008). Given uniqueness a Maximum Likelihood estimator can be used in a first stage to recover the conditional densities. A second stage uses these estimated conditional densities to construct the sample analog of the objective function (3), which is then used to recover estimates of $\theta_0$. The estimator extends directly to cases with measurement error in both capital and other inputs.

3 A Generalized Approach in the Wooldridge (2009) Setting

We follow the development of OP/LP given in Wooldridge (2009). We first extend this framework to the case when only capital is measured with error. We then generalize the setup to allow for measurement error in other inputs and capital.

3.1 Measurement Error in Capital Only

Let $\beta = (\beta_l, \beta_k, \beta_m, \beta_\omega)$, $\theta = (\beta, q)$ and $(\beta_0, \theta_0)$ denote the true values. Define $\tilde{y}_t \equiv (y_t, l_t, k_t, m_t)$ and $\tilde{y}_{t,t-1} \equiv (\tilde{y}_t, k_{t-1}, m_{t-1}, p_{t-1})$. Similarly define $y_t \equiv (y_t, p_t)$ and $y_{t,t-1} \equiv (y_t, k_{t-1}, m_{t-1}, p_{t-1})$.

Following Wooldridge we define the residual functions as

$$\epsilon_t(\beta) \equiv \rho_1(y_t; \beta) = y_t - \beta l_t - \beta_k k_t - \beta_m m_t - c(m_t, k_t, p_t)' \beta_\omega$$

$$[\xi_t + \epsilon_t](\theta) \equiv \rho_2(\tilde{y}_{t,t-1}; \theta) = y_t - \beta l_t - \beta_k k_t - \beta_m m_t - q(c(m_{t-1}, k_{t-1}, p_{t-1})' \beta_\omega).$$

The first residual function corresponds to the i.i.d. shock and the second function corresponds to the composite error.

In our general setting, following Wooldridge (2009), we include lagged variables in the instruments set. Denoting the entire history of input prices we observe by $\bar{p}_t = (p_t, p_{t-1}, \ldots, p_1)$, let $\tilde{x}_t = (k_t, k_{t-1}, \bar{p}_t, m_{t-1}, l_{t-1})$ and let $x_t = (\tilde{x}_t, m_t, l_t)$.\(^8\) Then the conditional moment restrictions take the form

$$g_1(x_t; \beta) \equiv E[\rho_1(y_t; \beta)|x_t] \text{ and } g_1(x_t; \beta_0) = 0,$$

$$g_2(\tilde{x}_t; \theta) \equiv E[\rho_2(\tilde{y}_{t,t-1}; \theta)|\tilde{x}_t] \text{ and } g_2(\tilde{x}_t; \theta_0) = 0.$$  

As noted by Wooldridge the second moment condition alone is sufficient for identification of production function parameters, $(\beta_l, \beta_k, \beta_m)$. Since $(\beta_l, \beta_k, \beta_m)$ are identified from the second moment alone, $\beta_\omega$ is also identified from the first moment. Then $q(\cdot)$ is also identified from the second moment because $\beta_\omega$ is now known from the first moment.

Letting

$$g(x_t; \theta) = (g_1(x_t; \beta) \ g_2(\tilde{x}_t; \theta))^\prime$$

---

\(^8\)Under the assumptions maintained in Wooldridge’s setup higher-order lagged values of inputs are also available as instruments.
we define the population objective function as

\[
Q(\theta) = E [g(x_i; \theta)'A g(x_i; \theta)] = \int g(x_i; \theta)'A g(x_i; \theta) f_{x_t} dx_t
\]

\[
= \int g(x_i; \theta)'A g(x_i; \theta) f_{k_t,k_{t-1}} m_{t,i,t}, \pi_{t-1}, m_{t-1,i,t-1} f_{m_{t,i}, \pi_{t-1}, m_{t-1,i,t-1}} dx_t
\]

where \(A = A(x_t)\) denotes some positive-definite weight matrix.

Our estimator is then obtained as a minimizer of the sample objective function based on (5). We use a sieve approach to estimate the unknown function \(q(\cdot)\). For this purpose let \(H_n = \{q : q = u^{k_2n}(c(m_{t-1}, k_{t-1}, p_{t-1})' \beta_q)'p_q\}\) denote a sieve space to approximate the space of functions \(q\) where \(u^{k_2n}(\cdot)\) denotes approximating basis functions with length equal to \(k_2n\) (e.g. polynomial approximations or spline approximations). In the first stage we estimate \(f_{y_t|k_t, k_{t-1}, m_t, l_t, \pi_t, m_{t-1}, l_{t-1}}\), \(f_{l_t|k_t, k_{t-1}, \pi_t, m_{t-1}, l_{t-1}}\), and \(f_{m_t|k_t, k_{t-1}, \pi_t, m_{t-1}, l_{t-1}}\) as we describe in Section 4. We also estimate \(f_{k_t, k_{t-1}} m_{t, l_t}, \pi_{t, t-1}, l_{t-1}\) in the first stage, which we need for integrating out the unobserved variables \((k_t, k_{t-1})\) in the sample objective function analog to (5). For any parameter value \(\theta\) we use the first four densities to evaluate the conditional moment functions \(\hat{g}(x_t; \theta) = (\hat{g}_1(x_t; \beta) \hat{g}_2(x_t; \theta))'\) as

\[
\hat{g}_1(x_t; \beta) = \int y_t \hat{f}_{y_t|k_t, k_{t-1}, m_t, l_t, \pi_t, m_{t-1}, l_{t-1}} dy_t - \beta_1 l_t - \beta_k k_t - \beta_m m_t - c(m_t, k_t, p_t)' \beta_q
\]

\[
\hat{g}_2(x_t; \theta) = \int y_t \hat{f}_{y_t|k_t, k_{t-1}, \pi_t, m_{t-1}, l_{t-1}} dy_t - \beta_t l_t \int l_t \hat{f}_{l_t|k_t, k_{t-1}, \pi_t, m_{t-1}, l_{t-1}} dl_t - \beta_k k_t
\]

\[
- \beta_m \int m_t \hat{f}_{m_t|k_t, k_{t-1}, \pi_t, m_{t-1}, l_{t-1}} dm_t - q(c(m_{t-1}, k_{t-1}, p_{t-1})' \beta_q).
\]

The sieve MD estimator is given by

\[
\hat{\theta}_n = \text{argmin}_{\theta \in (\beta, q), q \in H_n} \hat{Q}_n(\theta)
\]

where the sample objective function is

\[
\hat{Q}_n(\theta) = \frac{1}{n} \sum_{i=1}^{n} \int \hat{g}(x_{ti}; \theta)' \hat{A} \hat{g}(x_{ti}; \theta) \hat{f}_{k_t,k_{t-1}|m_{t,i}, l_{ti}, \pi_{ti}, m_{t-1,i}, l_{t-1,i}} dk_t dk_{t-1}
\]

with \(\hat{A}\) denoting a consistent estimator of \(A\) (e.g. Ai and Chen 2003). We study the asymptotic properties of the proposed sieve MD estimator in Sections 5 and 6.

### 3.2 Measurement Error in Capital and Intermediate Inputs

Our method can be extended to allow both capital and intermediate inputs to be measured with errors. Consider the production function

\[
y_t = \beta_1 l_t + \beta_k k_t + \beta_m m_t + \omega_t + \epsilon_t
\]
where \((k_t, m_t)\) are not directly observable, but instead mismeasured inputs \((k_t^*, m_t^*)\) are observed. The population objective function is rewritten as

\[
Q(\theta) = \int g(x_t; \theta)' A g(x_t; \theta) f_{k_t,k_{t-1},m_t,m_{t-1}|l_t,\tilde{p}_t,l_{t-1}} f_{l_t,\bar{p}_t,l_{t-1}} d\mathbf{x}_t
\]

where \(f_{l_t,\bar{p}_t,l_{t-1}}\) is observable from the data. Then the sieve MD estimator is defined as

\[
\hat{\theta}_n = \arg\min_{\theta = (\beta,q) : q \in \mathcal{H}_n} \hat{Q}_n(\theta)
\]

where the sample objective function is given by

\[
\hat{Q}_n(\theta) = \frac{1}{n} \sum_{i=1}^n \int \hat{g}(x_{ti}; \theta)' \hat{A} \hat{g}(x_{ti}; \theta) \hat{f}_{k_t,k_{t-1},m_t,m_{t-1}|l_t,\tilde{p}_t,l_{t-1}} \hat{f}_{l_t,\bar{p}_t,l_{t-1}} d\mathbf{k}_td\mathbf{m}_td\mathbf{m}_{t-1}.
\]

By similar arguments to the results in Section 4 below we can identify the conditional density functions associated with latent true inputs \((k_t, k_{t-1}, m_t, m_{t-1})\) to construct the moment functions, given mismeasured inputs \((k_t^*, k_{t-1}^*, m_t^*, m_{t-1}^*)\). The identification problem in this case is only about adding more instruments because of the unobserved true intermediate inputs. The set of additional instruments required other than \((k_t^*_{t-2}, k_t^*_{t-3})\) would be lags of mismeasured intermediate input \((m_t^*_{t-2}, m_t^*_{t-3})\) or lags of other (possibly mismeasured) intermediate inputs.

To construct the moment functions we need to identify \(f_{y_t|k_t,k_{t-1},m_t,m_{t-1},\tilde{p}_t,m_{t-1},l_{t-1}}\) in \(g_1(\mathbf{x}_t; \beta)\), \(f_{y_t|k_t,k_{t-1},\tilde{p}_t,m_{t-1},l_{t-1}}\), \(f_{l_t|k_t,k_{t-1},\tilde{p}_t,m_{t-1},l_{t-1}}\), and \(f_{m_t|k_t,k_{t-1},\tilde{p}_t,m_{t-1},l_{t-1}}\) in \(g_2(\hat{x}_t; \theta)\). With the true intermediate input \(m_t\) being not observed, the only added complication is to identify the density \(f_{m_t|k_t,k_{t-1},\tilde{p}_t,m_{t-1},l_{t-1}}\) because both \(m_t\) and \(m_{t-1}\) are not observed, which means both the regressor and the regressand are unobserved in the regression context. But given identification of \(f_{k_t,k_{t-1},m_t,m_{t-1}|\tilde{p}_t,l_{t-1}}\), by Bayes Rule, it is trivial to obtain

\[
f_{m_t|k_t,k_{t-1},\tilde{p}_t,m_{t-1},l_{t-1}} = \frac{f_{k_t,k_{t-1},m_t,m_{t-1}|l_t,\tilde{p}_t,l_{t-1}}}{f_{k_t,k_{t-1},m_t,m_{t-1}|\tilde{p}_t,l_{t-1}}} = \frac{f_{k_t,k_{t-1},m_t,m_{t-1}|\tilde{p}_t,l_{t-1}}}{\int f_{k_t,k_{t-1},m_t,m_{t-1}|\tilde{p}_t,l_{t-1}} dm_t}.
\]

The conditional density functions, \(f_{k_t,k_{t-1},m_t,m_{t-1}|l_t,\tilde{p}_t,l_{t-1}}\) in \(Q(\theta)\) and \(f_{k_t,k_{t-1},m_t,m_{t-1}|\tilde{p}_t,l_{t-1}}\) in \((7)\) associated with latent true inputs, \((k_t, k_{t-1}, m_t, m_{t-1})\), can be also identified by similar arguments to the results in Section 4. Given identification results we then estimate \(\hat{g}(\mathbf{x}_t; \theta)\) and \(\hat{f}_{k_t,k_{t-1},m_t,m_{t-1}|l_t,\tilde{p}_t,l_{t-1}}\) in \(\hat{Q}_n(\theta)\) where \((k_t, k_{t-1}, m_t, m_{t-1})\) are unobserved, but \((k_t^*, k_{t-1}^*, m_t^*, m_{t-1}^*)\) are instead observed, using (e.g.) lagged values \((k_t^*_{t-2}, k_t^*_{t-3}, m_t^*_{t-2}, m_t^*_{t-3})\) as instruments.

### 4 Identification and Estimation

In this section we develop identification and estimation of conditional density functions that appear in the moment functions.
4.1 Identification of Density Functions

In order to implement the sieve MD estimation (6) we need to obtain the consistent estimators of $f_{y_t|k_t,k_{t-1},m_t,l_t,p_t,m_{t-1},l_{t-1}}$, $f_{y_t|k_t,k_{t-1},p_t,m_{t-1},l_{t-1}}$, $f_{l_t|k_t,k_{t-1},p_t,m_{t-1},l_{t-1}}$, $f_{m_t|k_t,k_{t-1},p_t,m_{t-1},l_{t-1}}$, and $f_{k_{t-1}|m_t,l_t,p_{t-1},m_{t-1},l_{t-1}}$. For ease of notation let $\tilde{\omega}_t = (\tilde{p}_t,m_{t-1},l_{t-1})$ and $\omega_t = (\tilde{\omega}_t,m_t,l_t)$. Let either (i) $r_t = y_t$ and $s_t = \omega_t$ or (ii) $r_t \in \{l_t,m_t\}$ and $s_t = \tilde{\omega}_t$, depending on objects of interest.

4.1.1 Identification of $f_{r_t|k_t,k_{t-1},s_t}$

We make the following assumptions for identification and then discussions follow.\footnote{We focus on a decomposition-based approach as Hu and Schennach (2008) to recover the conditional densities in the first stage. It is worth noting that our identification strategy of recovering production function parameters in the second stage is not only limited to this approach. Other nonlinear methods to recover density functions from observed data (e.g., repeated measurement or auxiliary data approach) can be also used in the first stage.}

**Assumption 4.1.**

1. (i) $f_{r_t|k_t,k_{t-1},k_t^*,k_{t-1}^*,k_{t-2}^*,k_{t-3}^*,s_t} = f_{r_t|k_t,k_{t-1},s_t}$; (ii) $f_{k_t^*,k_{t-1}^*,k_{t-2}^*,k_{t-3}^*,s_t} = f_{k_t^*,k_{t-1}^*,k_{t-2}^*,k_{t-3}^*,s_t}$.
2. The operators $L_{k_t^*,k_{t-1}^*,k_{t-2}^*,k_{t-3}^*,s_t}$ and $L_{k_t^*,k_{t-1}^*,k_{t-2}^*,k_{t-3}^*,s_t}$ (defined in Appendix A) are one-to-one.
3. For any $s_t \in \text{Supp}(s_t)$ and any $(\tilde{k}_t,\tilde{k}_{t-1}) \neq (k_t,k_{t-1}) \in \text{Supp}(k_t,k_{t-1})$, the set
   \[
   \{ r_t: f_{r_t|k_t,k_{t-1},s_t}(r_t|\tilde{k}_t,\tilde{k}_{t-1},s_t) \neq f_{r_t|k_t,k_{t-1},s_t}(r_t|k_t,k_{t-1},s_t) \}
   \]
   has positive probability.
4. For any given $s_t \in \text{Supp}(s_t)$, there exists a known functional $M$ such that for all $(k_t,k_{t-1}) \in \text{Supp}(k_t,k_{t-1})$, we have $M[f_{k_t^*,k_{t-1}^*,k_{t-2}^*,k_{t-3}^*,s_t}(\cdot|k_t,k_{t-1},s_t)] = (k_t,k_{t-1})$.

Assumption 4.1.1 (i) is an exclusion restriction that $(k_t^*,k_{t-1}^*,k_{t-2}^*,k_{t-3}^*)$ do not provide additional information on $r_t$ beyond what is known conditional on $(k_t,k_{t-1},s_t)$. Assumption 4.1.1 (ii) is a similar exclusion restriction that $(k_{t-2}^*,k_{t-3}^*)$ do not provide additional information on $(k_t^*,k_{t-1}^*)$ beyond what is known conditional on $(k_t,k_{t-1},s_t)$. Assumption 4.1.2 is a nonparametric rank condition of the instruments and is important for the invertibility of the integral operators we employ for our identification proof. This identification condition appears in several different forms in the literature. For example it has been phrased as singular value decompositions with nonzero singular values (Darolles, Fan, Florens, and Renault 2011), nonsingularity (Hall and Horowitz 2005, Horowitz 2006), or completeness condition (or bounded completeness) (Newey and Powell 2003, Blundell, Chen, and Kristensen 2007, Andrews 2011).

Assumption 4.1.3 states the conditional density functions of the output and other inputs are distinguishable at two distinct values of capital. Assumption 4.1.3 ensures the uniqueness of the decomposition of an integral operator associated with $f_{|k_t,k_{t-1},s_t}$. This condition is violated only if the distribution of $r_t$ conditional on $(k_t,k_{t-1},s_t)$ is identical at different values of $(k_t,k_{t-1})$. For output this condition holds as long as $\beta_k > 0$.

Assumption 4.1.4 places restrictions on some measure of the location of a density, denoted by $M[\cdot]$ such as the mean, mode, and quantiles of the distribution. For example the classical
measurement error imposes \( E[k_t^*|k_t] = k_t \) while the nonclassical measurement error extends to the case e.g. median\((k_t^*|k_t) = k_t\), which ensures the uniqueness of the decomposition of an integral operator associated with \( f|k_t,k_{t-1},s_t \).

Assumption 4.1 is sufficient for identification of the density functions \( f_{r_t|k_t,k_{t-1},s_t}, f_{k_t^*,k_{t-1}^*|k_t,k_{t-1},s_t}, \) and \( f_{k_t,k_{t-1}|k_{t-2}^*,k_{t-3}^*,s_t} \) as stated in the following theorem.

**Theorem 4.2.** Suppose Assumption 4.1 holds. Then the observations of \( (r_t, k_t^*, k_{t-1}^*, k_{t-2}^*, k_{t-3}^*, s_t) \) uniquely identify unobserved densities \( f_{r_t|k_t,k_{t-1},s_t}, f_{k_t^*,k_{t-1}^*|k_t,k_{t-1},s_t}, \) and \( f_{k_t,k_{t-1}|k_{t-2}^*,k_{t-3}^*,s_t} \).

See Appendix A for the proof.

### 4.1.2 Identification of \( f_{k_t,k_{t-1}|m_{t,l},\bar{p}_t,m_{t-1},l_{t-1}} \) and \( f_{y_t|k_t,k_{t-1},\bar{p}_t,m_{t-1},l_{t-1}} \)

Next we show that other conditional density functions that appear in the moment functions and the objective function are obtained from both observables and identified objects from the previous section. Using the results in Theorem 4.2 we also identify \( f_{k_t,k_{t-1}|m_{t,l},\bar{p}_t,m_{t-1},l_{t-1}} \) as

\[
\begin{align*}
f_{k_t,k_{t-1}|m_{t,l},\bar{p}_t,m_{t-1},l_{t-1}} & = \int f_{k_t,k_{t-1}|k_{t-2}^*,k_{t-3}^*,m_{t,l},\bar{p}_t,m_{t-1},l_{t-1}} \frac{f_{k_{t-2}^*,k_{t-3}^*,m_{t,l},\bar{p}_t,m_{t-1},l_{t-1}}}{f_{m_{t,l},\bar{p}_t,m_{t-1},l_{t-1}}} \, dk_{t-2}^* \, dk_{t-3}^* \\
& \text{since } f_{k_t,k_{t-1}|k_{t-2}^*,k_{t-3}^*,m_{t,l},\bar{p}_t,m_{t-1},l_{t-1}} \text{ is identified (Theorem 4.2) and both } f_{k_{t-2}^*,k_{t-3}^*,m_{t,l},\bar{p}_t,m_{t-1},l_{t-1}} \text{ and } f_{m_{t,l},\bar{p}_t,m_{t-1},l_{t-1}} \text{ are observed from the data. The following Corollary summarizes the result.}
\end{align*}
\]

**Corollary 4.3.** Suppose Assumption 4.1 holds. Then we identify \( f_{k_t,k_{t-1}|m_{t,l},\bar{p}_t,m_{t-1},l_{t-1}} \) from the observations of \( (y_t, k_t^*, k_{t-1}^*, k_{t-2}^*, k_{t-3}^*, m_{t,l}, \bar{p}_t, m_{t-1}, l_{t-1}) \).

Next, we show \( f_{y_t|k_t,k_{t-1},\bar{p}_t,m_{t-1},l_{t-1}} \) is also identified. Note that

\[
f_{y_t|k_t,k_{t-1},\bar{p}_t,m_{t-1},l_{t-1}} = \int f_{y_t|k_t,k_{t-1}|m_{t,l},\bar{p}_t,m_{t-1},l_{t-1}} f_{m_{t,l}|k_t,k_{t-1},\bar{p}_t,m_{t-1},l_{t-1}} \, dm_{t,l} \, dl_t
\]

and that

\[
f_{m_{t,l}|k_t,k_{t-1},\bar{p}_t,m_{t-1},l_{t-1}} = \int f_{m_{t,l}|k_t,k_{t-1},\bar{p}_t,m_{t-1},l_{t-1}} f_{\bar{p}_t,m_{t-1},l_{t-1}} \, dm_{t,l} \, dl_t
\]

Therefore we can write

\[
f_{y_t|k_t,k_{t-1},\bar{p}_t,m_{t-1},l_{t-1}} = \int f_{y_t|k_t,k_{t-1}|m_{t,l},\bar{p}_t,m_{t-1},l_{t-1}} f_{k_t,k_{t-1}|m_{t,l},\bar{p}_t,m_{t-1},l_{t-1}} \, dm_{t,l} \, dl_t.
\]
Finally note that
\[
\begin{align*}
    f_{k_t,k_{t-1},\bar{\pi}_t,m_{t-1},l_t-1} & = \int f_{k_t,k_{t-1},k_t^*,m_{t-1},l_t-1} dk_t^* dk_{t-1}^* dm_t dl_t \\
    & = \int f_{k_t,k_{t-1},k_t^*,m_{t},l_t,\bar{\pi}_t,m_{t-1},l_t-1} dk_t^* dk_{t-1}^* dm_t dl_t.
\end{align*}
\]  

Because (i) \( f_{y_t|k_t,k_{t-1},m_{t-1},l_{t-1}} \) and \( f_{k_t,k_{t-1}|m_{t-1},l_{t-1}} \) that appear in (8) and (9) are identified from Theorem 4.2, (ii) \( f_{k_t,k_{t-1}|m_{t},\bar{\pi}_t,m_{t-1},l_{t-1}} \) that appears in (8) is identified from Corollary 4.3, and (iii) \( f_{m_{t},l_{t},\bar{\pi}_t,m_{t-1},l_{t-1}} \) and \( f_{k_t^*,k_{t-1}^*,m_{t},l_{t},\bar{\pi}_t,m_{t-1},l_{t-1}} \) are observed from the data, we conclude \( f_{y_t|k_t,k_{t-1},\bar{\pi}_t,m_{t-1},l_{t-1}} \) is also identified from (8) and (9). The following Corollary summarizes the result.

**Corollary 4.4.** Suppose Assumption 4.1 holds. Then we identify \( f_{y_t|k_t,k_{t-1},\bar{\pi}_t,m_{t-1},l_{t-1}} \) from the observations of \((y_t, k_t^*, k_{t-1}^*, k_{t-2}^*, k_{t-3}^*, m_{t}, l_t, \bar{\pi}_t, m_{t-1}, l_{t-1})\).

### 4.2 Estimation of Conditional Densities

In the sieve MD estimation (6), we first estimate conditional densities that enter \( g_1(x_t; \beta) \) and \( g_2(\bar{x}_t; \theta) \) for given \( \theta \equiv (\beta, q) \) with \( \beta \equiv (\beta_t, \beta_k, \beta_{m}, \beta_{w}) \) and the conditional density that is required to integrate out unobserved variables, \((k_t, k_{t-1})\), in \( Q(\theta) \). Then we estimate \( \theta \) in the second stage. So in the first stage, we need to estimate conditional densities associated with unobserved variables, \((k_t, k_{t-1})\), such as \( f_{k_t,k_{t-1}|m_{t},l_{t},\bar{\pi}_t,m_{t-1},l_{t-1}} \), \( f_{\bar{y}_t|k_t,k_{t-1},m_{t},l_{t},\bar{\pi}_t,m_{t-1},l_{t-1}} = f_{y_t|k_t,k_{t-1},\bar{\pi}_t,m_{t-1},l_{t-1}} \), \( f_{m_{t},l_{t},\bar{\pi}_t,m_{t-1},l_{t-1}} \), and \( f_{m_{t},|k_t,k_{t-1},\bar{\pi}_t,m_{t-1},l_{t-1}} \). The first conditional density is required to integrate out unobserved variables and the second one is needed to estimate \( g_1(x_t; \beta) \). Last three conditional densities are needed to estimate \( g_2(\bar{x}_t; \theta) \). Recall that \( \bar{x}_t = (k_t, k_{t-1}, \bar{\pi}_t, m_{t-1}, l_{t-1}) \) and \( x_t = (\bar{x}_t, m_t, l_t) \). Also recall that when \( r_t = y_t \), we take \( s_t = w_t \) and when \( r_t \in \{l_t, m_t\} \), we take \( s_t = \bar{w}_t \) where \( \bar{w}_t = (\bar{\pi}_t, m_{t-1}, l_{t-1}) \) and \( w_t = (\bar{w}_t, m_t, l_t) \). Similarly, we let \( z_t = (k_t, k_{t-1}, s_t) \) and \( z_t^* = (k_{t-2}^*, k_{t-3}^*, s_t) \). We now discuss how to estimate those conditional densities.

#### 4.2.1 Estimation of \( f_{r_t|k_t,k_{t-1},s_t} \)

Let
\[
\begin{align*}
    f_{r_t,k_t^*,k_{t-1}^*|z_t}(r_t,k_t^*,k_{t-1}^*|z_t; \alpha_0) & = \int f_{r_t|z_t}(r_t|z_t; \psi_0)f_{k_t^*,k_{t-1}^*|z_t}(k_t^*,k_{t-1}^*|z_t)f_{k_t,k_{t-1}|z_t}(k_t,k_{t-1}|z_t^*)dk_t dk_{t-1},
\end{align*}
\]

where \( \alpha_0 \equiv (\psi_0, f_1, f_2) \in \mathcal{A} \equiv \mathcal{M} \times \mathcal{F}_1 \times \mathcal{F}_2, \psi_0 \equiv (\phi_0, \eta_0) \in \Psi \equiv \Phi \times \mathcal{M}, f_1 \equiv f_{k_t^*,k_{t-1}^*|z_t}(k_t^*,k_{t-1}^*|z_t), \) and \( f_2 \equiv f_{k_t,k_{t-1}|z_t}(k_t,k_{t-1}|z_t^*) \). Here \( \phi_0 \) denotes a finite-dimensional parameter while \( \eta_0 \) denotes an infinite-dimensional parameter that determines the density \( f_{r_t|z_t} \). Therefore in this specification we use a semi-nonparametric density for \( f_{r_t|z_t} \) while we use nonparametric densities for \( f_{k_t^*,k_{t-1}^*|z_t} \)
and \( f_{k_t k_{t-1}} \). We can also consider fully nonparametric densities with additional notation. We choose to use the current specification as (10) for flexibility.

From Theorem 4.2, the parameters can be recovered by solving the maximization problems:

\[
\alpha_0 = \arg \max_{\alpha \in (\psi, f_1, f_2) \in A} E \left[ \ln \int f_{T_i | \eta_i} (r_i | \eta_i; \psi_0) f_1(k^*_t, k^*_{t-1} | \eta_i) f_2(k_t, k_{t-1} | \eta_i) dk_t dk_{t-1} \right].
\]

Note that \( \Phi \) is finite-dimensional parameter space. But we need to impose some restrictions on the function spaces (\( M, F_1, F_2 \)) for infinite-dimensional parameters. Let \( \xi \in V \subset R^{d_k} \), \( \| \cdot \|_E \) denote the Euclidean norm, and \( \nabla^a g(\xi) = \frac{\partial^{a_1 + \cdots + a_k} g(\xi)}{\partial x_1^{a_1} \cdots \partial x_k^{a_k}} \) denote the sum of \( a \)-th derivative where \( a = (a_1, a_2, \ldots, a_k) \) is a vector of nonnegative integers. Let \( \gamma \) denote the largest integer satisfying \( \Gamma < \gamma \). The Hölder space \( \Lambda^\gamma(V) \) of order \( \gamma > 0 \) is a space of functions \( g : V \mapsto R \) such that the first \( \gamma \) derivatives are bounded and the \( \gamma \)-th derivative is Hölder continuous with the exponent \( \gamma - \gamma \in (0, 1] \), i.e., for all \( \xi, \xi' \in V \) and some constant \( c \)

\[
\max_{\sum a_i = \gamma} \left| \nabla^a g(\xi) - \nabla^a g(\xi') \right| \leq c (\| \xi - \xi' \|_E)^{\gamma - \gamma}.
\]

The space \( \Lambda^\gamma(V) \) becomes a Banach space under the Hölder norm:

\[
\|g\|_{\Lambda^\gamma} = \sup_{\xi} |g(\xi)| + \max_{\sum a_i = \gamma} \sup_{\xi \neq \xi'} \frac{\left| \nabla^a g(\xi) - \nabla^a g(\xi') \right|}{(\| \xi - \xi' \|_E)^{\gamma - \gamma}} < \infty.
\]

A Hölder ball (of radius \( c \)) is defined as \( \Lambda^\gamma_c(V) = \{g \in \Lambda^\gamma(V) : \|g\|_{\Lambda^\gamma} \leq c < \infty \} \). Let \( \nu(\cdot) \) be a positive continuous weight function on \( V \) where \( \nu(\xi) = (1 + \| \xi \|_E)^{-\varsigma/2} \), \( \varsigma > \gamma > 0 \). Denote \( \Lambda^{\gamma, \nu}_c(V) \) as the weighted Hölder space with a weighted Hölder norm \( \|g\|_{\Lambda^{\gamma, \nu}} \equiv \|g\|_{\Lambda^\gamma} \) for \( \tilde{g}(\xi) \equiv g(\xi) \nu(\xi) \). Then a weighted Hölder ball is defined as \( \Lambda^{\gamma, \nu}_c(V) = \{g \in \Lambda^{\gamma, \nu}(V) : \|g\|_{\Lambda^{\gamma, \nu}} \leq c < \infty \} \).

**Assumption 4.5.**

1. \( \eta \in \Lambda^{\gamma, \nu}_c(\text{Supp}(\cdot)) \) where \( \gamma_1 > 1 \)

2. \( f_1 \in \Lambda^{\gamma_1, \nu}_c(\text{Supp}(k^*_t, k^*_{t-1}, \eta_t)) \) where \( \gamma_1 > 1 \) and \( \int f_1(k^*_t, k^*_{t-1} | \eta_t) dk^*_t dk^*_{t-1} = 1 \) for all \( \eta_t \in \text{Supp}(\eta_t) \)

3. \( f_2 \in \Lambda^{\gamma_1, \nu}_c(\text{Supp}(k_t, k_{t-1}, \eta_t^*)) \) where \( \gamma_1 > 1 \) and \( \int f_2(k_t, k_{t-1} | \eta_t^*) dk_t dk_{t-1} = 1 \) for all \( \eta_t^* \in \text{Supp}(\eta_t^*) \).

We define the spaces which satisfy relevant restrictions:

\[
M = \{\eta(\cdot) : \text{Assumption 4.5.1 holds}\},
\]

\[
F_1 = \{f_1(\cdot, \cdot | \cdot) : \text{Assumption 4.1.2, 4.1.4, and 4.5.2 hold}\},
\]

\[
F_2 = \{f_2(\cdot, \cdot | \cdot) : \text{Assumption 4.1.2 and 4.5.3 hold}\}.
\]
Let \( \{u_j, j = 1, 2, \ldots \} \) denote a sequence of known univariate basis functions (e.g., Fourier series, power series, spline, wavelets, etc.). For \( i \in \{\eta, 1, 2\} \), we define \( u^{\kappa_{\eta i}} = (u_1, \ldots, u_{\kappa_{\eta i}}) \) as a tensor-product linear sieve basis. For the consistent estimation of \( \alpha_0 \), we replace the parameter spaces with finite-dimensional compact sieve spaces:

\[
\mathcal{M}_n = \{\eta(\cdot) = u^{\kappa_{\eta}(\cdot)} \pi_\eta \text{ for all } \pi_\eta \text{ s.t. Assumption 4.5.1 holds}\}, \\
\mathcal{F}_{1n} = \{f_1(k_t^*, k_{t-1}^*|z_t) = u^{\kappa_{\eta 1}}(k_t^*, k_{t-1}^*, z_t) \pi_1 \text{ for all } \pi_1 \text{ s.t. Assumption 4.1.2, 4.1.4, and 4.5.2 hold}\}, \\
\mathcal{F}_{2n} = \{f_2(k_t, k_{t-1}|z_t^*) = u^{\kappa_{\eta 2}}(k_t, k_{t-1}, z_t^*) \pi_2 \text{ for all } \pi_2 \text{ s.t. Assumption 4.1.2 and 4.5.3 hold}\}.
\]

We then can replace the densities, \( f_1 \) and \( f_2 \), with functions in the sieve spaces:

\[
f_1(k_t^*, k_{t-1}^*|z_t) = \sum_{j_1=0}^{jn} \sum_{j_2=0}^{jn} \sum_{j_3=0}^{jn} \varphi_{j_1j_2j_3} u_{j_1}(k_t^* - k_{t-1}^*) u_{j_2}(k_{t-1}^*) u_{j_3}(z_t),
\]

\[
f_2(k_t, k_{t-1}|z_t^*) = \sum_{j_1=0}^{jn} \sum_{j_2=0}^{jn} \sum_{j_3=0}^{jn} \gamma_{j_1j_2j_3} u_{j_1}(k_t - k_{t-1}) u_{j_2}(k_{t-1}) u_{j_3}(z_t^*).
\]

Let the projection of the true parameter \( \alpha_0 \) onto the space \( \mathcal{A}_n \) where \( \mathcal{A}_n = \Psi_n \times \mathcal{F}_{1n} \times \mathcal{F}_{2n} \) with \( \Psi_n = \Phi \times \mathcal{M}_n \):

\[
\Pi_n \alpha = \alpha_n \\
\alpha_n = \arg \max_{\alpha_n \in (\psi, f_1, f_2) \in \mathcal{A}_n} E \left[ \ln \int f_{r_t|z_t}(r_t|z_t; \psi) f_1(k_t^*, k_{t-1}^*|z_t) f_2(k_t, k_{t-1}|z_t^*) dk_t dk_{t-1} \right].
\]

Then the proposed sieve ML estimators of \( \alpha_0 \) is defined as follows:

\[
\hat{\alpha}_n = (\hat{\psi}_n, \hat{f}_1n, \hat{f}_2n) = \arg \max_{(\psi, f_1, f_2) \in \mathcal{A}_n} \frac{1}{n} \sum_{i=1}^{n} \ln \int f_{r_t|z_t}(r_{ti}|z_{ti}; \psi) f_1(k_t^*, k_{t-1}^*, z_{ti}) f_2(k_t, k_{t-1}|z_t^*) dk_t dk_{t-1}.
\]

**4.2.2 Estimation of** \( f_{k_t, k_{t-1}|m_t, \bar{\theta}, m_{t-1}, l_{t-1}} \) **and** \( f_{y_t|k_t, k_{t-1}, \bar{\theta}, m_{t-1}, l_{t-1}} \)

Based on Corollary 4.3, we propose the following estimator of \( f_{k_t, k_{t-1}|m_t, \bar{\theta}, m_{t-1}, l_{t-1}} \)

\[
\hat{f}_{k_t, k_{t-1}|m_t, \bar{\theta}, m_{t-1}, l_{t-1}} = \hat{f}_{k_t, k_{t-1}|k_{t-2}^*, k_{t-3}^*, m_t, \bar{\theta}, m_{t-1}, l_{t-1}} \hat{f}_{k_{t-2}^*, k_{t-3}^*, m_t, \bar{\theta}, m_{t-1}, l_{t-1}} dk_{t-2}^* dk_{t-3}^* \\
\hat{f}_{m_t, \bar{\theta}, m_{t-1}, l_{t-1}}
\]

where \( \hat{f}_{k_t, k_{t-1}|k_{t-2}^*, k_{t-3}^*, m_t, \bar{\theta}, m_{t-1}, l_{t-1}} \) is \( \hat{f}_{2n}(k_t, k_{t-1}|z_t^*) \) from the sieve ML estimation (11). Both \( \hat{f}_{k_{t-2}^*, k_{t-3}^*, m_t, \bar{\theta}, m_{t-1}, l_{t-1}} \) and \( \hat{f}_{m_t, \bar{\theta}, m_{t-1}, l_{t-1}} \) can be estimated by any consistent nonparametric density estimators of observables. Similarly, based on Corollary 4.4, we can estimate \( f_{y_t|k_t, k_{t-1}, \bar{\theta}, m_{t-1}, l_{t-1}} \).
\[ \hat{f}_{yt|kt,kt-1}\bar{p}_t,mt-1,t-1 = \frac{\int \int \int \int \int \int \hat{f}_{yt|kt,kt-1}m_{t},l,t,\bar{p}_t,mt-1,l-1 \hat{f}_{mt,l,t,\bar{p}_t,mt-1,l-1}dm_l dl_t}{\int \int \int \int \int \int \hat{f}_{kt,kt-1}k^*_t-2,k^*_t-3,m_{t},l,t,\bar{p}_t,mt-1,l-1 \hat{f}_{t}k^*_t-2,k^*_t-3,m_{t},l,t,\bar{p}_t,mt-1,l-1 \hat{f}_{kt,kt-1}k^*_t-2,k^*_t-3,m_{t},l,t,\bar{p}_t,mt-1,l-1 \hat{f}_{mt,l,t,\bar{p}_t,mt-1,l-1}dk^*_t-2dk^*_t-3dm_l dl_t} \] (13)

where \( \hat{f}_{yt|kt,kt-1}m_{t},l,t,\bar{p}_t,mt-1,l-1 \) is \( f_{r_t|z_t}r_t|z_t,\hat{\nu}_n \) for \( r_t = y_t \) from the sieve ML estimation (11).

To summarize, both \( \hat{f}_{kt,kt-1}k^*_t-2,k^*_t-3,m_{t},l,t,\bar{p}_t,mt-1,l-1 \) and \( \hat{f}_{yt|kt,kt-1}m_{t},l,t,\bar{p}_t,mt-1,l-1 \) can be estimated using the sieve ML estimators from (11) and using the consistent estimators of \( \hat{f}_{kt,kt-1}k^*_t-2,k^*_t-3,m_{t},l,t,\bar{p}_t,mt-1,l-1 \) and \( \hat{f}_{mt,l,t,\bar{p}_t,mt-1,l-1} \) from nonparametric density estimation with observed data.

### 5 Consistency and Convergence Rates

Our estimation approach is a two-step method. In the first stage we estimate conditional density functions associated with moment functions and in the second stage we estimate production function parameters. Our first stage to estimate conditional density functions follows Hu and Schennach (2008). Recall that \( \alpha \) denote collections of different kinds of conditional density functions and parameters that determine the conditional density functions, which we need to estimate in the first stage. Also recall that \( \beta = (\beta_1, \beta_2, \beta_3, \beta_4) \) and \( \theta = (\beta, q) \), and \((\beta_0, \theta_0)\) denote the true values. First we obtain consistency of \( \hat{\alpha} \) (and \( \hat{\theta} \)) under a strong metric, \( \| \cdot \|_{s,\alpha} \) (and \( \| \cdot \|_{s,\theta} \)), and then establish that \( \hat{\alpha} \) (and \( \hat{\theta} \)) converge to \( \alpha_0 \) (and \( \theta_0 \)) at a rate faster than \( n^{-1/4} \) under a weak metric, \( \| \cdot \|_{\alpha} \) (and \( \| \cdot \|_{\theta} \)), which is necessary to obtain \( \sqrt{n} \)-asymptotic normality for \( \hat{\beta} \). Our notation and results are built on Hu and Schennach (2008) and Ai and Chen (2003) (also see Chen 2007, Blundell, Chen, and Kristensen 2007, and Song 2015).

#### 5.1 Consistency

Let \( D_t \equiv (r_t, k^*_t, k^*_t-1, z^*_t) \). Let the smoothing parameter in the first stage be \( \kappa_{1n} = d_\phi + \kappa_{1n} + \kappa_{1n} \), from the sieve approximation for \( f_1 \) and \( f_2 \) where \( d_\phi = \dim(\Phi) \), \( \kappa_{1n} = \dim(M_n) \), \( \kappa_{1n} = \dim(F_n) \), and \( \kappa_{1n} = \dim(F_{2n}) \) and also let the smoothing parameter in the second stage be \( \kappa_{2n} = \dim(H_n) \) from the sieve approximation for \( q(\cdot) \). Define \( \| \theta \|_{s,\theta} \equiv \| \beta \|_E + \| q \|_{\infty,\nu}, \) \( \| \alpha \|_{s,\alpha} \equiv \| \phi \|_E + \| \eta \|_{\infty,\nu} + \| f_1 \|_{\infty,\nu} + \| f_2 \|_{\infty,\nu}, \) where \( \| g \|_{\infty,\nu} \equiv \sup_{\xi} | g(\xi) \nu(\xi) | \) with a weight function \( \nu(\xi) = (1 + \| \xi \|^2_{\nu})^{-\gamma/2} \), \( \gamma > \gamma_1 > 0 \). Because the supports of the mismeasured variables could be unbounded, we use the weighted sup-norm metric.

**Assumption 5.1.**

1. The data \( \{D_{ti} = (r_{ti}, k^*_t, k^*_t-1, z^*_t): i = 1, \ldots, n \} \) are i.i.d. for all \( t = 1, \ldots, T \)
2. The density of \( D_t, f_{D_t}, \) satisfies \( \int \nu(D_t)^{-2}f_{D_t}(D_t)d(D_t) < \infty \).

This assumption is about the nature of data. Next we assume \( \alpha \) belong to a class of function that is well approximated by sieves.
Assumption 5.2.

1. There is a metric $\| \cdot \|_{s,\alpha}$ such that $\mathcal{A} \equiv \Psi \times \mathcal{F}_1 \times \mathcal{F}_2$ is compact under $\| \cdot \|_{s,\alpha}$

2. For any $\alpha \in \mathcal{A}$, there exists $\Pi_n \alpha \in \mathcal{A}_n \equiv \Psi_n \times \mathcal{F}_{1n} \times \mathcal{F}_{2n}$ with $\Psi_n \equiv \Phi \times \mathcal{M}_n$ such that $\| \Pi_n \alpha - \alpha \|_{s,\alpha} = o(1)$.

We now impose an envelope condition and Hölder continuity on the log likelihood.

Assumption 5.3.

1. $E[|\ln f_{r_t, k^*_t, k^*_{t-1} | x_t^*(r_t, k^*_t, k^*_{t-1} | z_t^*)|^2]$ is bounded

2. There exists a measurable function $h_1(D_t)$ with $E[|h_1(D_t)|^2] < \infty$ such that for any $\bar{\alpha} = (\bar{\psi}, \bar{f}_1, \bar{f}_2) \in \mathcal{A}$,

   $$\frac{f^{[1]}_{r_t, k^*_t, k^*_{t-1} | x_t^*(r_t, k^*_t, k^*_{t-1} | z_t^*; \bar{\alpha}, \nu)}}{f^{[1]}_{r_t, k^*_t, k^*_{t-1} | x_t^*(r_t, k^*_t, k^*_{t-1} | z_t^*; \bar{\alpha})}} \leq h_1(D_t),$$

   where the pathwise first derivative $f^{[1]}_{r_t, k^*_t, k^*_{t-1} | x_t^*(r_t, k^*_t, k^*_{t-1} | z_t^*; \bar{\alpha})}$ is defined in the proof of Theorem 5.10.1.

Assumption 5.4. $\kappa_{1n}/n \rightarrow 0, \kappa_{2n}/n \rightarrow 0$ and $\kappa_{1n}, \kappa_{2n} \rightarrow \infty$.

This is about the rate conditions on the smoothing parameters for consistency. We next assume the parameter of interest $\theta$ is identified, which is satisfied under standard nonparametric rank conditions (e.g., completeness condition in Newey and Powell 2003).

Assumption 5.5. $\theta_0 \in \Theta$ is the only $\theta \in \Theta$ satisfying $g(x_t; \theta) = 0$.

We now assume a consistent estimator of the weighting function is available.

Assumption 5.6.

1. $\hat{A}(x_t) = A(x_t) + o_p(1)$ uniformly over $x_t \in \mathcal{X}_t \equiv \text{Supp}(x_t)$; 2. $A(x_t)$ is finite and strictly positive uniformly over $x_t \in \mathcal{X}_t$.

Next we assume the moment function has bounded second moment and is continuous.

Assumption 5.7.

1. $E[||g(x_t; \theta_0)||^2]$ is bounded; 2. $g(x_t; \theta)$ is Hölder continuous in $\theta \in \Theta$.

We next restrict the parameter space $\Theta$ to a certain class such that any element in the parameter space can be well-approximated by sieves under a metric $\| \cdot \|_{s,\theta}$.

Assumption 5.8.

1. There is a metric $\| \cdot \|_{s,\theta}$ such that $\Theta \equiv \mathcal{B} \times \mathcal{H}$ is compact under $\| \cdot \|_{s,\theta}$

2. For any $\theta \in \Theta$, there exists $\Pi_n \theta \in \Theta_n \equiv \mathcal{B} \times \mathcal{H}_n$ such that $\| \Pi_n \theta - \theta \|_{s,\theta} = o(1)$.  

16
We also impose consistency of \( \hat{f}_{k_{t-2}k_{t-3}m_{t-1}l_{t-1}} \) and \( \hat{f}_{m_{t-1}l_{t-1}} \), which is well established in the standard nonparametric density estimation literature (e.g. kernel or sieve estimation). Consistency of the density estimators is required in steps to estimate \( \bar{w}_t = (\bar{p}_t, m_{t-1}, l_{t-1}) \) and \( w_t = (\bar{w}_t, m_{t}, l_{t}) \).

**Assumption 5.9.** Let \( s_t = w_t \).

1. \( \hat{f}_{k_{t-2}k_{t-3}s_t} = f_{k_{t-2}k_{t-3}s_t} + o_p(1) \) uniformly over all \( z_t^* = (k_{t-2}^*, k_{t-3}^*, s_t) \in Z_t^* \equiv \text{Supp}(z_t^*) \)

2. \( \hat{f}_{s_t} = f_{s_t} + o_p(1) \) uniformly over all \( s_t \in S_t \equiv \text{Supp}(s_t) \).

We state consistency of the proposed estimators of \( \alpha_0 \) and \( \theta_0 \) in the following theorem.

**Theorem 5.10.**

1. Suppose Assumptions 4.1, 4.5 and 5.1-5.4 hold. Then \( \| \hat{\alpha}_n - \alpha_0 \|_{S, \alpha} = o_p(1) \).

2. Suppose Assumptions 4.1, 4.5 and 5.1-5.9 hold. Then \( \| \hat{\theta}_n - \theta_0 \|_{S, \theta} = o_p(1) \).

See Appendix B and Supplementary Appendix for the proof.

### 5.2 Convergence Rates

In this section, we establish \( n^{-1/4} \) convergence rate of \( \hat{\alpha}_n \) (and \( \hat{\theta}_n \)) under a weaker metric, \( \| \cdot \|_\alpha \) (and \( \| \cdot \|_\theta \)), which is sufficient to obtain \( \sqrt{n}\)-asymptotic normality of \( \hat{\beta}_n \) (and also \( \hat{\phi}_n \)). We need to introduce additional notations. Again our notations and results are closely related to Hu and Schennach (2008), Ai and Chen (2003), and Song (2015).

Denote the first pathwise derivative of \( \ln f_{r_t,k_t^*,k_{t-1}^*}|z_t^*(r_t,k_t^*,k_{t-1}^*|z_t^*; \alpha) \) at the direction \( [\alpha - \alpha_0] \) evaluated at \( \alpha_0 \) by:

\[
\frac{d \ln f_{r_t,k_t^*,k_{t-1}^*}|z_t^*(r_t,k_t^*,k_{t-1}^*|z_t^*; \alpha)}{d\alpha}|_{\alpha = \alpha_0} \equiv \frac{d \ln f_{r_t,k_t^*,k_{t-1}^*}|z_t^*(r_t,k_t^*,k_{t-1}^*|z_t^*; (1-\tau)\alpha_0 + \tau\alpha)}{d\tau}|_{\tau = 0}
\]

almost everywhere. Also for \( \alpha_1, \alpha_2 \in A \) denote

\[
\frac{d \ln f_{r_t,k_t^*,k_{t-1}^*}|z_t^*(r_t,k_t^*,k_{t-1}^*|z_t^*; \alpha)}{d\alpha}|_{\alpha_1 = \alpha_2} \equiv \frac{d \ln f_{r_t,k_t^*,k_{t-1}^*}|z_t^*(r_t,k_t^*,k_{t-1}^*|z_t^*; \alpha)}{d\alpha}|_{\alpha = \alpha_0} \frac{d \ln f_{r_t,k_t^*,k_{t-1}^*}|z_t^*(r_t,k_t^*,k_{t-1}^*|z_t^*; \alpha)}{d\alpha}|_{\alpha = \alpha_0} - \frac{d \ln f_{r_t,k_t^*,k_{t-1}^*}|z_t^*(r_t,k_t^*,k_{t-1}^*|z_t^*; \alpha)}{d\alpha}|_{\alpha_2 = \alpha_0}.
\]
Then the pathwise derivative is written by:

\[
\frac{d \ln f_{r_t, k_t^*, k_{t-1}^*} | z_t^* (r_t, k_t^*, k_{t-1}^* | z_t^*; \alpha_0)}{d \alpha} [\alpha - \alpha_0]
\]

\[
= \frac{1}{f_{r_t, k_t^*, k_{t-1}^*} | z_t^* (r_t, k_t^*, k_{t-1}^* | z_t^*; \alpha_0)} \times \left\{ \int \frac{d}{d \psi} f_{r_t | z_t} (r_t | z_t; \psi_0) | \psi - \psi_0 | f_{k_t^*, k_{t-1}^* | z_t} (k_t^*, k_{t-1}^* | z_t) f_{k_t, k_{t-1} | z_t^*} (k_t, k_{t-1} | z_t^*) dk_t dk_{t-1}
\right.
\]

\[
+ \int f_{r_t | z_t} (r_t | z_t; \psi_0) \left[ f_1 (k_t^*, k_{t-1}^* | z_t) - f_{k_t^*, k_{t-1}^* | z_t} (k_t^*, k_{t-1}^* | z_t) \right] f_{k_t, k_{t-1} | z_t^*} (k_t, k_{t-1} | z_t^*) dk_t dk_{t-1}
\]

\[
+ \int f_{r_t | z_t} (r_t | z_t; \psi_0) f_{k_t^*, k_{t-1}^* | z_t} (k_t^*, k_{t-1}^* | z_t) \left[ f_2 (k_t, k_{t-1} | z_t^*) - f_{k_t, k_{t-1} | z_t^*} (k_t, k_{t-1} | z_t^*) \right] dk_t dk_{t-1} \right\}.
\]

For any \( \alpha_1, \alpha_2 \in A \), the metric \( \| \cdot \|_\alpha \) is defined as

\[
\| \alpha_1 - \alpha_2 \|_\alpha = \sqrt{E \left[ \left( \frac{d \ln f_{r_t, k_t^*, k_{t-1}^*} | z_t^* (r_t, k_t^*, k_{t-1}^* | z_t^*; \alpha_0)}{d \alpha} [\alpha_1 - \alpha_2] \right)^2 \right]}.
\]

We develop similar notation for \( \theta \). Let \( \rho(y_{t,t-1}; \theta) = (\rho_1(y_{t}; \beta) \rho_2(\tilde{y}_{t,t-1}; \theta))^t \) and denote its first pathwise derivative at the direction \( \theta_0 \) evaluated at \( \theta_0 \) by:

\[
\frac{d \rho(y_{t,t-1}; \theta_0)}{d \theta} |_{\theta = \theta_0} = \frac{d \rho(y_{t,t-1}; \theta_0) (1 - \tau) \theta_0 + \tau \theta |_{\tau = 0}}{d \tau}
\]

almost everywhere and for any \( \theta_1, \theta_2 \in \Theta \) denote

\[
\frac{d \rho(y_{t,t-1}; \theta_0)}{d \theta} |_{\theta_1 - \theta_2} = \frac{d \rho(y_{t,t-1}; \theta_0)}{d \theta} |_{\theta_2 - \theta_1} - \frac{d \rho(y_{t,t-1}; \theta_0)}{d \theta} |_{\theta_2 - \theta_0},
\]

\[
\frac{d g(x_t; \theta_0)}{d \theta} |_{\theta_1 - \theta_2} = \left( E \left[ \frac{d \rho_1 (y_{t}; \beta_0)}{d \theta} |_{\beta_1 = \beta_2} \right] x_t \right),
\]

\[
E \left[ \frac{d \rho_2 (\tilde{y}_{t,t-1}; \theta_0)}{d \theta} |_{\theta_1 = \theta_2} \right].
\]

Similarly, for any \( \theta_1, \theta_2 \in \Theta \), the metric \( \| \cdot \|_\theta \) is defined as:

\[
\| \theta_1 - \theta_2 \|_\theta = \sqrt{E \left[ \left( \frac{d g(x_t; \theta_0)}{d \theta} |_{\theta_1 - \theta_2} \right)^t A(x_t) \frac{d g(x_t; \theta_0)}{d \theta} |_{\theta_1 - \theta_2} \right]}.
\]

We now state sufficient conditions for the convergence rate of the proposed estimators of \( \alpha_0 \) and \( \theta_0 \). First we strengthen the requirement for the approximation order enough to obtain the convergence rate of estimators up to \( o_p(n^{-1/4}) \). This is also well known to be satisfied when \( \alpha \) and \( \theta \) belong to a class of bounded and smooth functions.

**Assumption 5.11.**

1. There exists a measurable function \( c(r_t, k_t^*, k_{t-1}^*, z_t^*) \) with \( E[c(\cdot)^4] < \infty \) such that for all
Assumption 5.14.\( (r_t, k_t^*, k_{t-1}^*, \mathbf{z}_t^*) \in \text{Supp}(r_t, k_t^*, k_{t-1}^*, \mathbf{z}_t^*) \) and \( \alpha \in \mathcal{A}_n \), \( |\ln f_{r_t, k_t^*, k_{t-1}^*|\mathbf{z}_t^*}(r_t, k_t^*, k_{t-1}^*|\mathbf{z}_t^*; \alpha)| \leq c() \)

2. \( \ln f_{r_t, k_t^*, k_{t-1}^*|\mathbf{z}_t^*}(r_t, k_t^*, k_{t-1}^*|\mathbf{z}_t^*; \alpha) \in \mathcal{N}_{\mu_1, \mu_2}(\text{Supp}(r_t, k_t^*, k_{t-1}^*, \mathbf{z}_t^*)) \) for some constant \( c > 0 \) with \( \mu_1 > d_{D_t}/2 \), for all \( \alpha \in \mathcal{A}_n \), where \( d_{D_t} \) is the dimension of \( D_t \).

This states an envelop condition and a smoothness condition on \( \ln f_{r_t, k_t^*, k_{t-1}^*|\mathbf{z}_t^*}(r_t, k_t^*, k_{t-1}^*|\mathbf{z}_t^*; \alpha) \). The next assumption quantifies the approximation error of \( \Pi_n \alpha \) to \( \alpha \) by the sieve approximation.

**Assumption 5.12.** There is a constant \( \mu_1 > 0 \) such that for any \( \alpha \in \mathcal{A} \), there exists \( \Pi_n \alpha \in \mathcal{A}_n \) satisfying \( \|\Pi_n \alpha - \alpha \|_\alpha = O(\kappa_{1n}^{\mu_1}) \), and \( \kappa_{1n}^{\mu_1} = o(n^{-1/4}) \).

Similarly, next assumptions are imposed for the estimation of \( \theta_0 \) in the second stage. Assumption 5.14 below is well known to be satisfied for sieve approximations when \( \Theta \) is a Hölder class.

**Assumption 5.13.**

1. There exist measurable functions \( c_1(y_t) \) and \( c_2(\mathbf{y}_{t,t-1}) \) with \( E[c_1(\cdot)^4] < \infty \) and \( E[c_2(\cdot)^4] < \infty \) such that (i) \( |\rho_1(y_t; \beta)| \leq c_1(\cdot) \) for all \( y_t \in \text{Supp}(y_t) \) and \( \beta \in \mathcal{B} \) and (ii) \( |\rho_2(\mathbf{y}_{t,t-1}; \theta)| \leq c_2(\cdot) \) for all \( \mathbf{y}_{t,t-1} \in \text{Supp}(\mathbf{y}_{t,t-1}) \) and \( \theta \in \Theta_n \).

2. Each element of \( g(x_t; \theta) \in \mathcal{N}_{\mu_2, \nu}(X_t) \) for some constant \( c > 0 \) with \( \mu_2 > d_{x_t}/2 \), for all \( \theta \in \Theta_n \), where \( d_{x_t} \) is the dimension of \( x_t \).

**Assumption 5.14.** There is a constant \( \mu_2 > 0 \) such that for any \( \theta \in \Theta \), there exists \( \Pi_n \theta \in \Theta_n \) satisfying \( \|\Pi_n \theta - \theta \|_\theta = O(\kappa_{2n}^{\mu_2}) \), and \( \kappa_{2n}^{\mu_2} = o(n^{-1/4}) \).

Next assumption restricts the estimation error of the weight matrix.

**Assumption 5.15.** \( \hat{A}(x_t) = A(x_t) + o_p(n^{-1/4}) \) uniformly over \( x_t \in \mathcal{X}_t \).

Let \( \xi_{0n} = \sup_{x_{11} \in \text{Supp}(\eta), x_{12} \in \text{Supp}(k_t^*, k_{t-1}^*, \mathbf{z}_t^*), x_{13} \in \text{Supp}(k_t^*, k_{t-1}^*, \mathbf{z}_t^*)} \|u^{\alpha n}(\xi_{11}), u^{\kappa_{n1}}(\xi_{12}), u^{\kappa_{n2}}(\xi_{13})\| \), which is non-decreasing in \( \kappa_{1n} \). Also \( N(\delta_1, \mathcal{A}_n, \| \cdot \|_{s, \alpha}) \) and \( N(\delta_2, \Theta_n, \| \cdot \|_{s, \theta}) \) define the minimal number of radius \( \delta_1 \) covering balls of \( \mathcal{A}_n \) under the \( \| \cdot \|_{s, \alpha} \) metric, and the minimal number of radius \( \delta_2 \) covering balls of \( \Theta_n \) under the \( \| \cdot \|_{s, \theta} \) metric, respectively. Two assumptions below impose conditions related to the sieve approximation in the first stage, such as size of sieve space, well-definedness of the norm, and the norm equivalence.

**Assumption 5.16.**

1. \( \kappa_{1n} \times \ln n \times \xi_{0n}^2 \times n^{-1/2} = o(1) \)

2. \( \ln[N(\delta_1, \mathcal{A}_n, \| \cdot \|_{s, \alpha})] \leq \text{const.} \times \kappa_{1n} \times \ln(\kappa_{1n}/\delta_1) \).

**Assumption 5.17.**

1. \( \mathcal{A} \) is convex in \( \alpha_0 \) and \( f_{r_t|\mathbf{z}_t}(r_t|\mathbf{z}_t; \psi) \) is pathwise differentiable at \( \psi_0 \)
2. For some $c_{11}, c_{12} > 0$,

$$c_{11} E \left[ \ln \frac{f_{r_t, k_t^*, k_{t-1}^*} | z_t^* (r_t, k_t^*, k_{t-1}^* | z_t^*; \alpha_0)}{f_{r_t, k_t^*, k_{t-1}^*} | z_t^* (r_t, k_t^*, k_{t-1}^* | z_t^*; \alpha)} \right] \leq \| \alpha - \alpha_0 \|_\alpha^2 \leq c_{12} E \left[ \ln \frac{f_{r_t, k_t^*, k_{t-1}^*} | z_t^* (r_t, k_t^*, k_{t-1}^* | z_t^*; \alpha_0)}{f_{r_t, k_t^*, k_{t-1}^*} | z_t^* (r_t, k_t^*, k_{t-1}^* | z_t^*; \alpha)} \right]$$

holds for all $\alpha \in \mathcal{A}_n$ with $\| \alpha - \alpha_0 \|_{s, \alpha} = o(1)$.

Similarly, the conditions associated with the sieve approximation in the second stage are imposed in the next two assumptions.

**Assumption 5.18.**

1. $\kappa_{2n} \times \ln n \times \xi_0^n \times n^{-1/2} = o(1)$
2. $\ln[N(\delta_2, \Theta_n, \| \cdot \|_{s, \theta})] \leq \text{const.} \times \kappa_{2n} \times \ln(\kappa_{2n}/\delta_2)$.

**Assumption 5.19.**

1. $\Theta$ is convex in $\theta_0$ and $\rho(y_{t,t-1}; \theta)$ is pathwise differentiable at $\theta_0$
2. For some $c_{21}, c_{22} > 0$,

$$c_{21} E \left[ g(x_t; \theta)' A(x_t) g(x_t; \theta) \right] \leq \| \theta - \theta_0 \|_\theta^2 \leq c_{22} E \left[ g(x_t; \theta)' A(x_t) g(x_t; \theta) \right]$$

holds for all $\theta \in \Theta_n$ with $\| \theta - \theta_0 \|_{s, \theta} = o(1)$.

We next impose convergence rates of $f_{t-2, k_{t-3}^*, m_{t-2}, i_{t-3}, m_{t-1}, l_{t-1}}$ and $f_{m_{t-1}, i_{t-1}, m_{t-1}, l_{t-1}}$, which appear in (12) and (13). Recall $\tilde{w}_t = (\tilde{p}_t, m_{t-1}, l_{t-1})$ and $w_t = (\tilde{w}_t, m_t, l_t)$.

**Assumption 5.20.** Let $s_t = w_t$.

1. $f_{k_{t-2}^*, k_{t-3}^*, s_t} = f_{k_{t-2}^*, k_{t-3}^*, s_t} + o_p(n^{-1/4})$ uniformly over all $z_t^* = (k_{t-2}^*, k_{t-1}^*, s_t) \in \mathcal{Z}_t^*$
2. $f_{s_t} = f_{s_t} + o_p(n^{-1/4})$ uniformly over all $s_t \in \mathcal{S}_t$.

Under these assumptions we obtain the following convergence rate results in the weaker metrics.

**Theorem 5.21.**

1. Suppose Assumptions 4.1, 4.5, 5.1-5.4, 5.11-5.12, and 5.16-5.17 hold. Then $\| \hat{\alpha}_n - \alpha_0 \|_\alpha = o_p(n^{-1/4})$.
2. Suppose Assumptions 4.1, 4.5, 5.1-5.9, and 5.11-5.20 hold. Then $\| \hat{\theta}_n - \theta_0 \|_\theta = o_p(n^{-1/4})$.

See Appendix B and Supplementary Appendix for the proof.
6 Asymptotic Normality

We now establish the asymptotic normality and $\sqrt{n}$- consistency of $\hat{\phi}_n$ and $\hat{\beta}_n$. We adopt useful notation introduced in Ai and Chen (2003). Let $\nabla_1, \nabla_2$ denote the closure of the linear span of $A - \alpha_0$ under the metric $\| \cdot \|_\alpha$ and the closure of the linear span of $\Theta - \theta_0$ under the metric $\| \cdot \|_\theta$, respectively. Note $\nabla_1 = \mathcal{R}^{d_x} \times \nabla_1$ with $\nabla_1 \equiv \mathcal{M} \times \mathcal{F}_1 \times \mathcal{F}_2 - \{(\eta_0, f_{k_t^r,k_{t-1}^r}|z_t), f_{k_t,k_{t-1}|z_t}(k_t,k_{t-1}|z_t^*)\}$ and $\nabla_2 = \mathcal{R}^{d_\beta} \times \nabla_2$ with $\nabla_2 \equiv \mathcal{H} - \{q_0\}$. Then $(\nabla_1, \| \cdot \|_\alpha)$ becomes a Hilbert space with the inner product $\langle v_{11}, v_{12} \rangle_\alpha$ induced by $\| \cdot \|_\alpha$:

$$\langle v_{11}, v_{12} \rangle_\alpha = E \left[ \frac{d \ln f_{r_t,k_t^r,k_{t-1}^r|z_t^*(r_t,k_t^r,k_{t-1}^r|z_t^*;\alpha_0)}}{d \alpha} [v_{11}] \cdot \frac{d \ln f_{r_t,k_t^r,k_{t-1}^r|z_t^*(r_t,k_t^r,k_{t-1}^r|z_t^*;\alpha_0)}}{d \alpha} [v_{12}] \right].$$

Also $(\nabla_2, \| \cdot \|_\theta)$ is a Hilbert space with the inner product:

$$\langle v_{21}, v_{22} \rangle_\theta = E \left[ \left( \frac{d g(x_t;\theta_0)}{d \theta} [v_{21}] \right) ^\prime A(x_t) \frac{d g(x_t;\theta_0)}{d \theta} [v_{22}] \right].$$

The pathwise derivative at $\alpha_0$ is defined as

$$\frac{d \ln f_{r_t,k_t^r,k_{t-1}^r|z_t^*(r_t,k_t^r,k_{t-1}^r|z_t^*;\alpha_0)}}{d \alpha} [\alpha - \alpha_0] = \frac{d \ln f_{r_t,k_t^r,k_{t-1}^r|z_t^*(r_t,k_t^r,k_{t-1}^r|z_t^*;\alpha_0)}}{d \phi} (\phi - \phi_0) + \frac{d \ln f_{r_t,k_t^r,k_{t-1}^r|z_t^*(r_t,k_t^r,k_{t-1}^r|z_t^*;\alpha_0)}}{d \eta} [\eta - \eta_0]$$

and the pathwise derivative at $\theta_0$ is defined as

$$\frac{d g(x_t;\theta_0)}{d \theta} [\theta - \theta_0] = \frac{d g(x_t;\theta_0)}{d \beta} [\beta - \beta_0] + \frac{d g(x_t;\theta_0)}{d q} [q - q_0].$$

For each component $\phi_j$ of $\phi$, $j = 1, 2, \ldots, d_\phi$, we define $w_{1j}^* \in \nabla_1$ as

$$w_{1j}^* \equiv (\eta_j^*, f_{1j}^*, f_{2j}^*)' = \arg \min_{(\eta_j,f_{1j},f_{2j})' \in \nabla_1} E \left[ \left( \frac{d \ln f_{r_t,k_t^r,k_{t-1}^r|z_t^*(r_t,k_t^r,k_{t-1}^r|z_t^*;\alpha_0)}}{d \phi_j} - \frac{d \ln f_{r_t,k_t^r,k_{t-1}^r|z_t^*(r_t,k_t^r,k_{t-1}^r|z_t^*;\alpha_0)}}{d \eta} \right) [\eta_j] - \frac{d \ln f_{r_t,k_t^r,k_{t-1}^r|z_t^*(r_t,k_t^r,k_{t-1}^r|z_t^*;\alpha_0)}}{d f_1} [f_{1j}] - \frac{d \ln f_{r_t,k_t^r,k_{t-1}^r|z_t^*(r_t,k_t^r,k_{t-1}^r|z_t^*;\alpha_0)}}{d f_2} [f_{2j}] \right]^2. $$
Also for each component $\beta_j$ of $\beta$, $j = 1, 2, \ldots, d_\beta$, we define $w^*_2 = \mathbb{W}_2$ as
\[
\begin{align*}
  w^*_{2j} &= \arg \min_{w^*_{2j} \in \mathbb{W}_2} E \left[ \left( \frac{dg(x_t; \theta_0)}{d\beta_j} - \frac{dg(x_t; \theta_0)}{dq}[w^*_{2j}] \right) A(x_t) \left( \frac{dg(x_t; \theta_0)}{d\beta_j} - \frac{dg(x_t; \theta_0)}{dq}[w^*_{2j}] \right) \right].
\end{align*}
\]
We let $w^*_1 = (w^*_{11}, w^*_{12}, \ldots, w^*_{1d_\beta})$ and define
\[
\begin{align*}
  \frac{d \ln f_{r_t, k^*_t, k^*_t-1|z^*_t}(r_t, k^*_t, k^*_t-1|z^*_t; \alpha_0)}{df}[w^*_{1j}] &= \frac{d \ln f_{r_t, k^*_t, k^*_t-1|z^*_t}(r_t, k^*_t, k^*_t-1|z^*_t; \alpha_0)}{d\eta}[\eta^*_j] + \frac{d \ln f_{r_t, k^*_t, k^*_t-1|z^*_t}(r_t, k^*_t, k^*_t-1|z^*_t; \alpha_0)}{df_1}[f^*_1] \nonumber \\
  + \frac{d \ln f_{r_t, k^*_t, k^*_t-1|z^*_t}(r_t, k^*_t, k^*_t-1|z^*_t; \alpha_0)}{df_2}[f^*_2].
\end{align*}
\]
and
\[
\begin{align*}
  \frac{d \ln f_{r_t, k^*_t, k^*_t-1|z^*_t}(r_t, k^*_t, k^*_t-1|z^*_t; \alpha_0)}{df}[w^*_{1j}] &= \left( \frac{d \ln f_{r_t, k^*_t, k^*_t-1|z^*_t}(r_t, k^*_t, k^*_t-1|z^*_t; \alpha_0)}{df}[w^*_{11}], \ldots, \frac{d \ln f_{r_t, k^*_t, k^*_t-1|z^*_t}(r_t, k^*_t, k^*_t-1|z^*_t; \alpha_0)}{df}[w^*_{1d_\beta}] \right).
\end{align*}
\]
Also let $w^*_2 = (w^*_{21}, w^*_{22}, \ldots, w^*_{2d_\beta})$ and
\[
\frac{dg(x_t; \theta_0)}{dq}[w^*_{2j}] = \left( \frac{dg(x_t; \theta_0)}{dq}[w^*_{21}], \ldots, \frac{dg(x_t; \theta_0)}{dq}[w^*_{2d_\beta}] \right).
\]
We also define the row vectors
\[
\begin{align*}
  G_{w^*_1}(x_t, \theta_0) &\equiv \frac{d \ln f_{r_t, k^*_t, k^*_t-1|z^*_t}(r_t, k^*_t, k^*_t-1|z^*_t; \alpha_0)}{d\phi}[w^*_{1j}] - \frac{d \ln f_{r_t, k^*_t, k^*_t-1|z^*_t}(r_t, k^*_t, k^*_t-1|z^*_t; \alpha_0)}{df}[w^*_{11}], \\
  G_{w^*_2}(x_t, \theta_0) &\equiv \frac{dg(x_t; \theta_0)}{d\beta} - \frac{dg(x_t; \theta_0)}{dq}[w^*_{2j}].
\end{align*}
\]
Define $s_1(\alpha) \equiv \lambda^*_1(\phi - \phi_0) = (v^*_1, \alpha - \alpha_0)_\alpha$, for all $\alpha \in \mathcal{A}$ where $v^*_1 \equiv (v^*_\phi, v^*_f) \in \mathbb{N}_1$, $v^*_\phi = J^{-1}_1 \lambda_1$, $v^*_f = -w^*_1 \times v^*_\phi$ and that
\[
\begin{align*}
  s_2(\theta) - s_2(\theta_0) &\equiv \lambda^*_2(\theta - \theta_0) = (v^*_2, \theta - \theta_0)_\theta,
\end{align*}
\]
for all $\alpha \in \mathcal{A}$ where $v^*_1 \equiv (v^*_\phi, v^*_f) \in \mathbb{N}_1$, $v^*_\phi = J^{-1}_1 \lambda_1$, $v^*_f = -w^*_1 \times v^*_\phi$ and that
for all \( \theta \in \Theta \) where \( v_2^\alpha \equiv (v_{\beta}^\alpha, v_0^\alpha) \in \nabla v_2 \), \( v_0^\alpha = f^{-1}_2 \lambda_2 \), \( v_0^\beta = -w_2^\alpha \times v_\beta^\alpha \).

We denote, for any \( v_1 \in \nabla v_1 \),
\[
\frac{d \ln f_{r_t, k_t^*, k_{t-1}^*} z_t^* (r_t, k_t^*, k_{t-1}^*; \alpha)}{d \alpha} \bigg|_{\alpha = 0} = \frac{d \ln f_{r_t, k_t^*, k_{t-1}^*} z_t^* (r_t, k_t^*, k_{t-1}^*; \alpha + \tau v_1)}{d \tau} \bigg|_{\tau = 0} \quad \text{a.s. in } D_t.
\]

We also denote
\[
\frac{d \rho(y_{t, t-1}; \theta)}{d \theta} \bigg|_{\theta = 0} = \frac{d \rho(y_{t, t-1}; \theta + \tau v_2)}{d \tau} \bigg|_{\tau = 0} \quad \text{a.s. in } y_{t, t-1},
\]
and
\[
\frac{d g(x_t; \theta)}{d \theta} \bigg|_{\theta = 0} = E \left[ \frac{d \rho(y_{t, t-1}; \theta)}{d \theta} \bigg|_{\theta = 0} \right] x_t \quad \text{a.s. in } x_t,
\]
for any \( v_2 \in \nabla v_2 \).

Define \( N_{01n} = \{ \alpha : \alpha \in A_n, \| \alpha - \alpha_0 \|_{s, \alpha} = o(n^{-1/4}), \| \alpha - \alpha_0 \|_{s, \alpha} = o(1) \} \) and define \( N_{01} \) in the same way except \( A_n \) being replaced with \( A \). We denote a local alternative \( \alpha^*(\alpha, \varepsilon_n) = (1 - \varepsilon_n)\alpha + \varepsilon_n (v_0^* + \alpha_0) \) with \( \varepsilon_n = o(n^{-1/2}) \), for \( \alpha \in N_{01n} \). Let \( \Pi_n \alpha^*(\alpha, \varepsilon_n) \) be the projection of \( \alpha^*(\alpha, \varepsilon_n) \) onto \( A_n \). Also define \( N_{02n} = \{ \theta : \theta \in \Theta_n, \| \theta - \theta_0 \|_{s, \theta} = o(n^{-1/4}), \| \theta - \theta_0 \|_{s, \theta} = o(1) \} \) and define \( N_{02} \) in the same way except \( \Theta_n \) being replaced with \( \Theta \).

We state the sufficient conditions for the \( \sqrt{n}\)-normality of \( \hat{\phi}_n \) and \( \hat{\beta}_n \).

**Assumption 6.1.**

1. \( E[G_{w_1^*}(D, \alpha_0) G_{w_1^*}(D, \alpha_0)] \) exists, is bounded, and is positive-definite.
2. \( \phi_0 \in \text{int}(\Phi) \).

**Assumption 6.2.** There is \( v_1^* = (v_\beta^*, -\Pi_n w_1^* \times v_0^*) \in A_n \) such that \( \| v_1^* \|_{s, \alpha} = o(n^{-1/4}) \).

Assumption 6.1 contains a standard rank condition and the bounded second moment condition that ensure the existence of the Riesz representation. Assumption 6.2 states that the Riesz representer is also well approximated by the sieves. Next we impose an envelope condition on the second derivative of the log likelihood, which is associated with the stochastic equicontinuity.

**Assumption 6.3.** There exists measurable function \( h_2(D_t) \) with \( E[(h_2(D_t))^2] < \infty \) such that for any \( \bar{\alpha} = (\bar{\psi}, \bar{f}_1, \bar{f}_2) \in N_{01} \),
\[
\left| \frac{f_{1[1]}^*(r_t, k_t^*, k_{t-1}^*; z_t^*; \alpha, \nu)}{f_{3[1]}^*(r_t, k_t^*, k_{t-1}^*; z_t^*; \alpha)} \right|^2 + \left| \frac{f_{1[2]}^*(r_t, k_t^*, k_{t-1}^*; z_t^*; \alpha, \nu)}{f_{3[2]}^*(r_t, k_t^*, k_{t-1}^*; z_t^*; \alpha)} \right|^2 \leq h_2(D_t)
\]

where the pathwise first derivative \( f_{1[1]}^*(r_t, k_t^*, k_{t-1}^*; z_t^*; \alpha, \nu) \) and the pathwise second derivative \( f_{1[2]}^*(r_t, k_t^*, k_{t-1}^*; z_t^*; \alpha, \nu) \) are defined in the proofs of Theorem 5.10.1 and Theorem 6.12.1, respectively.
We now introduce some notation for the assumptions below. Let

\[
\frac{d \ln f_{r_t, k_t^*, k_{t-1}^*} | z_t^* (r_t, k_t^*, k_{t-1}^* | z_t^*; \alpha_0)}{d\bar{\alpha}} \Big|_{u_{\kappa_1}^{\kappa_1 n}} = \left( \frac{d \ln f_{r_t, k_t^*, k_{t-1}^*} | z_t^* (r_t, k_t^*, k_{t-1}^* | z_t^*; \alpha_0)}{d\bar{\phi}} \right) \left( \frac{d \ln f_{r_t, k_t^*, k_{t-1}^*} | z_t^* (r_t, k_t^*, k_{t-1}^* | z_t^*; \alpha_0)}{d\bar{\eta}} \right) \left( \frac{d \ln f_{r_t, k_t^*, k_{t-1}^*} | z_t^* (r_t, k_t^*, k_{t-1}^* | z_t^*; \alpha_0)}{d\bar{f}} \right),
\]

where for \( \bar{f} = \eta, f_1, \) or \( f_2, \)

\[
\frac{d \ln f_{r_t, k_t^*, k_{t-1}^*} | z_t^* (r_t, k_t^*, k_{t-1}^* | z_t^*; \alpha_0)}{d\bar{f}} = \left( \frac{d \ln f_{r_t, k_t^*, k_{t-1}^*} | z_t^* (r_t, k_t^*, k_{t-1}^* | z_t^*; \alpha_0)}{d\bar{\phi}} \right) \left( \frac{d \ln f_{r_t, k_t^*, k_{t-1}^*} | z_t^* (r_t, k_t^*, k_{t-1}^* | z_t^*; \alpha_0)}{d\bar{\eta}} \right) \left( \frac{d \ln f_{r_t, k_t^*, k_{t-1}^*} | z_t^* (r_t, k_t^*, k_{t-1}^* | z_t^*; \alpha_0)}{d\bar{f}} \right),
\]

Let

\[
\Omega_{\kappa_1 n} = E \left\{ \left( \frac{d \ln f_{r_t, k_t^*, k_{t-1}^*} | z_t^* (r_t, k_t^*, k_{t-1}^* | z_t^*; \alpha_0)}{d\bar{\alpha}} \right) \left( \frac{d \ln f_{r_t, k_t^*, k_{t-1}^*} | z_t^* (r_t, k_t^*, k_{t-1}^* | z_t^*; \alpha_0)}{d\bar{\eta}} \right) \left( \frac{d \ln f_{r_t, k_t^*, k_{t-1}^*} | z_t^* (r_t, k_t^*, k_{t-1}^* | z_t^*; \alpha_0)}{d\bar{f}} \right) \right\}.
\]

We now impose a normalization on the approximating functions, which trivially holds for e.g. orthonormalized approximating basis functions.

**Assumption 6.4.** The smallest eigenvalue of the matrices \( \Omega_{\kappa_1 n} \) is bounded away from zero, and \( \| u_{j}^{\kappa_1 n} \|_{\infty, \nu} < \infty \) for \( j = 1, 2, \ldots, \kappa_1 n, \) uniformly in \( \kappa_1 n. \)

Next we impose Lipschitz conditions on the fourth derivatives of the log densities.

**Assumption 6.5.** For all \( \alpha, \tilde{\alpha} \in N_{01n}, \) there exists a measurable function \( h_3(D_t) \) with \( E[|h_3(D_t)|] < \infty \) such that

\[
\left| \frac{d^4}{d\tau^4} \ln f_{r_t, k_t^*, k_{t-1}^*} | z_t^* (r_t, k_t^*, k_{t-1}^* | z_t^*; \tilde{\alpha} + \tau(\alpha - \alpha_0)) \right|_{\tau=0} \leq h_3(D_t) \| \alpha - \alpha_0 \|_4^4.
\]

Now we add conditions on the moment function. First, we impose a local identification condition for \( \beta_0. \) This also ensures the existence of the Riesz representer.

**Assumption 6.6.**

1. \( E[G_{w_2}(x_t, \theta_0) A(x_t) G_{w_2}(x_t, \theta_0)] \) exists, is bounded, and is positive-definite.
2. \( \beta_0 \in \text{int}(B) \)

3. \( \Sigma_0(x_t) \equiv \text{var}[\rho(y_{t,t-1}; \theta_0) \mid x_t] \) is positive-definite for all \( x_t \in \mathcal{X}_t \).

**Assumption 6.7.** There is \( v_{2n}^* = (v_{\beta_1}^*, -\Pi_n w_{2n}^* \times v_{\beta_1}^*) \in \Theta_n - \{ \theta_0 \} \) such that \( \| v_{2n}^* - v_2^* \|_\theta = o(n^{-1/4}) \).

This condition is required to ensure that the sieve space \( \mathcal{H}_n \) approximates the space \( \mathcal{H} - \{ q_0 \} \), so the Riesz representer is also well approximated by the sieves. Now we impose an envelope condition for \( \rho(y_{t,t-1}; \theta) \).

**Assumption 6.8.**

1. For all \( \theta \in \mathcal{N}_{02} \), the pathwise first derivative \( (dp(y_{t,t-1}; \theta)/d\theta)[v_2] \) exists a.s. in \( y_{t,t-1} \)

2. Each element of \( (dp(y_{t,t-1}; \theta)/d\theta)[v_{2n}^*] \) satisfies an envelope condition and is Hölder continuous in \( \theta \in \mathcal{N}_{02n} \)

3. Each element of \( (dg(x_t; \theta)/d\theta)[v_{2n}^*] \) is in \( \Lambda^2_\gamma (\mathcal{X}_t), \gamma > d_x/2 \) for all \( \theta \in \mathcal{N}_{02} \).

Next, the following two assumptions are necessary to control the asymptotic bias when \( \rho(\cdot) \) is a highly nonlinear function of \( \theta \).

**Assumption 6.9.** Uniformly over \( \theta \in \mathcal{N}_{02n} \), we have

\[
E \left[ \left\| \frac{dg(x_t; \theta)}{d\theta} \left[ v_{2n}^* \right] - \frac{dg(x_t; \theta_0)}{d\theta} \left[ v_{2n}^* \right] \right\|^2 \right] = o(n^{-1/2}).
\]

We also impose

**Assumption 6.10.** Uniformly over \( \theta \in \mathcal{N}_{02}, \bar{\theta} \in \mathcal{N}_{02n} \), we have

\[
E \left[ \left\{ \frac{dg(x_t; \theta_0)}{d\theta} \left[ v_{2n}^* \right] \right\} A(x_t) \left\{ \frac{dg(x_t; \theta)}{d\theta} \left[ \bar{\theta} - \theta_0 \right] - \frac{dg(x_t; \theta_0)}{d\theta} \left[ \bar{\theta} - \theta_0 \right] \right\} \right] = o(n^{-1/2}).
\]

The following condition is required to have the higher order terms asymptotically negligible in the stochastic expansions to obtain the asymptotic normality.

**Assumption 6.11.** For all \( \theta \in \mathcal{N}_{02n} \), the pathwise second derivative \( d^2p(y_{t,t-1}; \theta + \tau v_{2n}^*)/d\tau^2 \big|_{\tau=0} \) exists a.s. in \( y_{t,t-1} \), and is bounded by a measurable function \( c(y_{t,t-1}) \) with \( E[\| c(y_{t,t-1}) \|_E^2] < \infty \).

We now conclude the \( \sqrt{n} \)-normality of \( (\hat{\beta}_n, \hat{\phi}_n) \) in the following theorem.

**Theorem 6.12.**

1. Suppose Assumptions 4.1, 4.5, 5.1-5.4, 5.11-5.12, 5.16-5.17, and 6.1-6.5 hold. Then we have \( \sqrt{n}(\hat{\phi}_n - \phi_0) \xrightarrow{d} N(0, J_1^{-1}) \), where \( J_1 = E[G_{w_1^*}(D_t, \alpha_0)^\gamma G_{w_2^*}(D_t, \alpha_0)] \).
2. Suppose Assumptions 4.1, 4.5, 5.1-5.9, 5.11-5.20, and 6.1-6.11 hold. Then \( \sqrt{n}(\hat{\beta}_n - \beta_0) \xrightarrow{d} N(0, J_2^{-1}) \), where

\[
J_2 = E[G_{w_2^*}(x_t, \theta_0)'A(x_t)G_{w_2^*}(x_t, \theta_0)] \times (E[G_{w_2^*}(x_t, \theta_0)'A(x_t)\Sigma_0(x_t)A(x_t)G_{w_2^*}(x_t, \theta_0)])^{-1} \\
\times E[G_{w_2^*}(x_t, \theta_0)'A(x_t)G_{w_2^*}(x_t, \theta_0)].
\]

See Appendix B and Supplementary Appendix for the proof. As Ai and Chen (2003) one may also obtain an efficient two-step estimator of \( \beta_0 \) by using the optimal weight matrix as \( A(x_t) = \Sigma_0(x_t)^{-1} \). In this case the asymptotic variance reduces to \( J_2^{-1} = \{E[G_{w_2^*}(x_t, \theta_0)'\Sigma_0(x_t)^{-1}G_{w_2^*}(x_t, \theta_0)]\}^{-1} \).

7 Conclusion

The control function estimators for production functions of Olley and Pakes (1996) and Levinsohn and Petrin (2003) are not consistent if the arguments that enter the control function for productivity are measured with error. The problem is complicated by the fact that the unobserved productivity control function is an unknown nonlinear function of capital and other inputs, so standard instrumental variable techniques cannot be employed. We combine results from the nonclassical measurement error literature (e.g. Hu and Schennach 2008) with the sieve minimum distance estimation techniques (e.g. Ai and Chen 2003, Chen 2007, Song 2015) to recover the production function parameters in the Wooldridge (2009) setting. We develop the asymptotic distribution for our estimator, which accounts for the multiple stages and the infinite-dimensional nuisance parameters.
A  Identification of Conditional Densities: Proof of Theorem 4.2

We first define integral operators. Let $\mathcal{G}(\mathcal{C})$ denote a space of functions with a domain $\mathcal{C}$. $L_{R_1,R_2|R_3,R_4,R_5}$ denote an integral operator mapping $g \in \mathcal{G}(\text{Supp}(r_3,r_4))$ to $L_{R_1,R_2|R_3,R_4,R_5} g \in \mathcal{G}(\text{Supp}(r_1,r_2))$ for a given $R_5 = r_5$:

$$[L_{R_1,R_2|R_3,R_4,R_5} g](r_1,r_2) = \int f_{R_1,R_2|R_3,R_4,R_5}(r_1,r_2 \mid r_3,r_4,r_5) g(r_3,r_4) dr_3 dr_4.$$ 

In addition, $\triangle_{R_1|R_2,R_3,R_4}$ denotes a diagonal operator mapping $g \in \mathcal{G}(\text{Supp}(r_3,r_4))$ to $\triangle_{R_1|R_2,R_3,R_4} g \in \mathcal{G}(\text{Supp}(r_2,r_3))$ for a given $R_4 = r_4$:

$$\triangle_{R_1|R_2,R_3,R_4} g = \int f_{R_1|R_2,R_3,R_4}(r_1 \mid r_2,r_3,r_4) g(r_2,r_3).$$

Proof. (Theorem 4.2) Recall that $r_t = y_t$ and $s_t = w_t$ or $r_t \in \{l_t, m_t\}$ and $s_t = \tilde{w}_t$ where $\tilde{w}_t = (\tilde{p}_t, m_{t-1}, l_{t-1})$ and $w_t = (\tilde{w}_t, m_t, l_t)$. We have

$$f_{r_t,k_t^*,k_t^*|k_{t-1}^*,k_{t-1}^*|k_{t-2}^*,k_{t-3}^*,s_t} = \int f_{r_t,k_t,k_{t-1},k_t^*,k_{t-1}^*|k_{t-2}^*,k_{t-3}^*,s_t} dk_t dk_{t-1}$$

$$= \int f_{r_t|k_t,k_{t-1},s_t} f_{k_t^*,k_{t-1}^*|k_{t-2}^*,k_{t-3}^*,s_t} dk_t dk_{t-1}$$

$$= \int f_{r_t|k_t,k_{t-1},s_t} f_{k_t^*,k_{t-1}^*|k_{t-2}^*,k_{t-3}^*,s_t} f_{k_t,k_{t-1}|k_{t-2}^*,k_{t-3}^*,s_t} dk_t dk_{t-1}$$

$$= \int f_{r_t|k_t,k_{t-1},s_t} f_{k_t^*,k_{t-1}^*|k_{t-2}^*,k_{t-3}^*,s_t} f_{k_t,k_{t-1}|k_{t-2}^*,k_{t-3}^*,s_t} f_{k_t,k_{t-1}|k_{t-2}^*,k_{t-3}^*,s_t} dk_t dk_{t-1}$$

by Assumption 4.1.1. Now we need to show uniqueness of the decomposition. Using operator notation, we have

$$[L_{r_t,k_t^*,k_t^*|k_{t-1}^*,k_{t-1}^*|k_{t-2}^*,k_{t-3}^*,s_t} g](k_t^*,k_{t-1}^*) = \int f_{r_t,k_t,k_{t-1},s_t} f_{k_t^*,k_{t-1}^*|k_{t-2}^*,k_{t-3}^*,s_t} g(k_t^*,k_{t-1}^*) dk_t dk_{t-1}$$

$$= \int f_{r_t|k_t,k_{t-1},s_t} f_{k_t^*,k_{t-1}^*|k_{t-2}^*,k_{t-3}^*,s_t} g(k_t^*,k_{t-1}^*) dk_t dk_{t-1}$$

$$= \int f_{r_t|k_t,k_{t-1},s_t} f_{k_t^*,k_{t-1}^*|k_{t-2}^*,k_{t-3}^*,s_t} f_{k_t,k_{t-1}|k_{t-2}^*,k_{t-3}^*,s_t} g(k_t^*,k_{t-1}^*) dk_t dk_{t-1}$$

$$= \int f_{r_t|k_t,k_{t-1},s_t} f_{k_t^*,k_{t-1}^*|k_{t-2}^*,k_{t-3}^*,s_t} f_{k_t,k_{t-1}|k_{t-2}^*,k_{t-3}^*,s_t} f_{k_t,k_{t-1}|k_{t-2}^*,k_{t-3}^*,s_t} g(k_t^*,k_{t-1}^*) dk_t dk_{t-1}$$

$$= [L_{r_t,k_t^*,k_t^*|k_{t-1}^*,k_{t-1}^*|k_{t-2}^*,k_{t-3}^*,s_t} g](k_t^*,k_{t-1}^*).$$
Thus we have \( L_{r_t,k_t^*,k_t^-1|k_{t-2}^*,k_{t-3}^*,s_t} = L_{k_t^*,k_t^-1|k_{t-1},s_t} \Delta r_t|k_{t-1},s_t L_{k_{t-1}|k_{t-2}^*,k_{t-3}^*,s_t} \). From the integration of the above equation over all \( r_t \), we have

\[
L_{k_t^*,k_t^-1|k_{t-2}^*,k_{t-3}^*,s_t} = L_{k_t^*,k_t^-1|k_{t-1},s_t} L_{k_{t-1}|k_{t-2}^*,k_{t-3}^*,s_t}.
\]

Then \( L_{k_{t-1}|k_{t-2}^*,k_{t-3}^*,s_t} = \frac{1}{L_{k_t^*,k_t^-1|k_{t-1},s_t} L_{k_{t-1}|k_{t-2}^*,k_{t-3}^*,s_t}} \), since \( L_{k_t^*,k_t^-1|k_{t-1},s_t} \) is one-to-one. Since \( L_{k_t^*,k_t^-1|k_{t-2}^*,k_{t-3}^*,s_t} \) is also one-to-one, rearranging two equations above gives

\[
L_{r_t,k_t^*,k_t^-1|k_{t-2}^*,k_{t-3}^*,s_t} L_{k_{t-1}|k_{t-2}^*,k_{t-3}^*,s_t} = L_{k_t^*,k_t^-1|k_{t-1},s_t} \Delta r_t|k_{t-1},s_t L_{k_{t-1}|k_{t-2}^*,k_{t-3}^*,s_t}.
\]

Thus the observed operator on the left hand side has the form of an eigenvalue-eigenfunction decomposition, with the eigenvalues corresponding to the density \( f_{r_t|k_{t-1},s_t} \) and the eigenfunctions corresponding to the density \( f_{k_{t-1},s_t} \).

Assumption 4.1 ensures that this decomposition is unique, by a similar argument to Hu and Schennach (2008). Given the identification of \( f_{r_t|k_{t-1},s_t} \) and \( f_{k_t^*,k_t^-1|k_{t-1},s_t} \), \( f_{k_{t-1}|k_{t-2}^*,k_{t-3}^*,s_t} \) is also identified from \( L_{k_{t-1}|k_{t-2}^*,k_{t-3}^*,s_t} = L_{k_{t-1}^1|k_{t-1},s_t} L_{k_{t-2}^*,k_{t-3}^*,s_t} \) where \( L_{k_{t-1}^1|k_{t-1}^*,s_t} \) is observable.

\[\square\]

**B  Proof of Asymptotic Results**

We provide proofs of Asymptotic Results: Consistency (Theorem 5.10), Convergence Rates (Theorem 5.21), and Asymptotic Normality (Theorem 6.12) in a Supplementary Appendix.

Here we provide only a brief summary of our proof results. For the consistency results, Theorem 5.10.1 follows from a similar argument to the proof of Lemma 2 in Hu and Schennach (2008). Theorem 5.10.2 can be proved by verifying the three conditions i)-iii) of Lemma A.1 of Newey and Powell (2003). For the convergence rate results, Theorem 5.21.1 follows from a similar argument to Theorem 2 in Hu and Schennach (2008). Theorem 5.21.2 follows from a similar argument to Ai and Chen (2003) and Song (2015). For the asymptotic normality results, Theorem 6.12.1 follows from a similar argument to Theorem 3 in Hu and Schennach (2008). Theorem 6.12.2 follows from a similar argument to Ai and Chen (2003) and Song (2015).

**References**


29
Supplementary Appendix:

Estimating Production Functions with Control Functions When Capital Is Measured with Error

Kyoo il Kim
Michigan State University

Amil Petrin
University of Minnesota, Twin Cities and NBER

Suyong Song
University of Iowa

August 2015
C Supplementary Appendix

Here we provide proofs of our Asymptotic results: Consistency (Theorem 5.10), Convergence Rates (Theorem 5.21), and Asymptotic Normality (Theorem 6.12).

Recall that \( \tilde{x}_t = (k_t, k_{t-1}, \tilde{p}_t, m_{t-1}, l_{t-1}) \), \( x_t = (\tilde{x}_t, m_t, l_t) \), \( z_t = (k_t, k_{t-1}, s_t) \) and \( z_t^* = (k_{t-2}, k_{t-3}, s_t) \). Also recall that \( r_t = y_t \) and \( s_t = w_t \) or \( r_t \in \{l_t, m_t\} \) and \( s_t = \tilde{w}_t \) where \( \tilde{w}_t = (\tilde{p}_t, m_{t-1}, l_{t-1}) \) and \( w_t = (\tilde{w}_t, m_t, l_t) \). Define

\[
Q(\theta) = E[g(x_t; \theta)'A(x_t)g(x_t; \theta)] = \int g(x_t; \theta)'A(x_t)g(x_t; \theta)f_{k_t, k_{t-1}|w_t}f_{w_t}dx_t
\]

\[
Q_n(\theta) = \frac{1}{n} \sum_{i=1}^{n} \int g(x_{ti}; \theta)'A(x_{ti})g(x_{ti}; \theta)f_{k_t, k_{t-1}|w_t}dk_tdk_{t-1}
\]

\[
\hat{Q}_n(\theta) = \frac{1}{n} \sum_{i=1}^{n} \int \hat{g}(x_{ti}; \theta)'\hat{A}(x_{ti})\hat{g}(x_{ti}; \theta)\hat{f}_{k_t, k_{t-1}|w_t}dk_tdk_{t-1}
\]

C.1 Consistency: Proof of Theorem 5.10

The first result (Theorem 5.10.1) follows from a similar argument to the proof of Lemma 2 in Hu and Schennach (2008). In the proof we obtain the bound for the pathwise first derivative by

\[
\left| \frac{d}{d\tau} \ln f_{r_t, k_{t-1}^*, k_{t-2}^*|z_t^*}(r_t, k_{t-1}^*, k_{t-2}^*|z_t^*; \tilde{\alpha} + \tau(\alpha_1 - \alpha_2)) \right|_{\tau=0}
\]

\[
\leq \frac{1}{f_{r_t, k_{t-1}^*, k_{t-2}^*|z_t^*}(r_t, k_{t-1}^*, k_{t-2}^*|z_t^*; \tilde{\alpha})}
\times \left\{ \int \left| \frac{d}{d\psi} f_{r_t|z_t}(r_t|z_t; \tilde{\psi})\nu^{-1}(\cdot)\tilde{f}_1(k_t^*, k_{t-1}^*|z_t)\tilde{f}_2(k_t, k_{t-1}|z_t^*) \right| dk_tdk_{t-1}
\right.

\left. + \int f_{r_t|z_t}(r_t|z_t; \tilde{\psi})\nu^{-1}(k_t^*, k_{t-1}^*, z_t)\tilde{f}_2(k_t, k_{t-1}|z_t^*) \right| dk_tdk_{t-1}
\left. + \int f_{r_t|z_t}(r_t|z_t; \tilde{\psi})\tilde{f}_1(k_t^*, k_{t-1}^*|z_t)\nu^{-1}(k_t, k_{t-1}, z_t, z_t^*) \right| dk_tdk_{t-1}
\left. \right\} ||\alpha_1 - \alpha_2||_{s, \alpha}
\]

\[
\equiv \frac{f_{r_t, k_{t-1}^*, k_{t-2}^*|z_t^*}(r_t, k_{t-1}^*, k_{t-2}^*|z_t^*; \tilde{\alpha}, \nu)}{f_{r_t, k_{t-1}^*, k_{t-2}^*|z_t^*}(r_t, k_{t-1}^*, k_{t-2}^*|z_t^*; \tilde{\alpha})} ||\alpha_1 - \alpha_2||_{s, \alpha}
\]

where \( \tilde{\alpha} = (\tilde{\psi}, \tilde{f}_1, \tilde{f}_2) \in \mathcal{A} \).

The second result (Theorem 5.10.2) can be proved by verifying the three conditions i)-iii) of Lemma A.1 of Newey and Powell (2003). Condition i) is explicitly assumed in Assumption 5.5. Condition iii) is satisfied by choosing \( \Pi_n \theta \in \Theta_n \) such that \( ||\Pi_n \theta - \theta||_{s, \theta} = o(1) \) in Assumption 5.8.2.

We now verify Condition ii) of their Lemma A.1 by checking Conditions (i)-(iii) in Lemma A.2 of Newey and Powell (2003). First, Assumption 5.8.1 assumes their compactness of a parameter space in Lemma A.2 Condition (i). Next we verify their Lemma A.2 Condition (ii) about the pointwise convergence of the sample objective function to the population one in \( \theta \). Note that for a
by Assumption 5.6. Since generic positive constant $C$,

$$
\left| \hat{Q}_n(\theta) - Q_n(\theta) \right| \\
\leq \frac{C}{n} \cdot \sum_{i=1}^{n} \int \left( \left\| \hat{g}(x_{ti}; \theta) \hat{f}_{k_t, k_{t-1}}^\frac{1}{2} \right\|_E^2 - \left\| g(x_{ti}; \theta) f_{k_t, k_{t-1}}^\frac{1}{2} \right\|_E^2 \right) \, dk_t \, dk_{t-1} \\
\leq \frac{C}{n} \cdot \int \left( \sum_{i=1}^{n} \left\| \hat{g}(x_{ti}; \theta) \hat{f}_{k_t, k_{t-1}}^\frac{1}{2} - g(x_{ti}; \theta) f_{k_t, k_{t-1}}^\frac{1}{2} \right\|_E^2 + 2 \sum_{i=1}^{n} \left\| g(x_{ti}; \theta) f_{k_t, k_{t-1}}^\frac{1}{2} \right\|_E \sum_{i=1}^{n} \left\| \hat{g}(x_{ti}; \theta) \hat{f}_{k_t, k_{t-1}}^\frac{1}{2} - g(x_{ti}; \theta) f_{k_t, k_{t-1}}^\frac{1}{2} \right\|_E \right) \, dk_t \, dk_{t-1},
$$

by Assumption 5.6. Since $\sum_{i=1}^{n} \left\| g(x_{ti}; \theta) f_{k_t, k_{t-1}}^\frac{1}{2} \right\|_E^2 / n = O_p(1)$ by the Markov inequality from Assumption 5.7.1, it suffices to show

$$
\sum_{i=1}^{n} \left\| \hat{g}(x_{ti}; \theta) \hat{f}_{k_t, k_{t-1}}^\frac{1}{2} - g(x_{ti}; \theta) f_{k_t, k_{t-1}}^\frac{1}{2} \right\|_E^2 / n = o_p(1)
$$

to verify $|\hat{Q}_n(\theta) - Q_n(\theta)| = o_p(1)$. Then note that for any given $(k_t, k_{t-1})$

$$
E \left[ \sum_{i=1}^{n} \left\| \hat{g}(x_{ti}; \theta) \hat{f}_{k_t, k_{t-1}}^\frac{1}{2} - g(x_{ti}; \theta) f_{k_t, k_{t-1}}^\frac{1}{2} \right\|_E^2 / n \right]
\leq \frac{1}{n} \sum_{i=1}^{n} \left\{ \left( \int y_t f_{yt|k_t, k_{t-1}, w_{ti}} \, dy_t - \beta_k l_t - \beta_k m_t - c(m_t, k_t, p_t) \right) \hat{f}_{k_t, k_{t-1}}^\frac{1}{2} \right\}^2
+ \left\{ \int y_t f_{yt|k_t, k_{t-1}, w_{ti}} \, dy_t - \beta_k l_t - \beta_k m_t - c(m_t, k_t, p_t) \hat{f}_{k_t, k_{t-1}}^\frac{1}{2} \right\}^2
\leq \frac{1}{n} \sum_{i=1}^{n} \left\{ \left( \int y_t f_{yt|k_t, k_{t-1}, w_{ti}} \, dy_t - \beta_k l_t - \beta_k m_t - c(m_t, k_t, p_t) \right) \hat{f}_{k_t, k_{t-1}}^\frac{1}{2} \right\}^2
+ \left\{ \int y_t f_{yt|k_t, k_{t-1}, w_{ti}} \, dy_t - \beta_k l_t - \beta_k m_t - c(m_t, k_t, p_t) \hat{f}_{k_t, k_{t-1}}^\frac{1}{2} \right\}^2
\quad + \beta_m \int m_t f_{m_t|k_t, k_{t-1}, w_{ti}} \, dm_t - q(c(m_{t-1, i}, k_{t-1}, p_{t-1, i}) \beta_\omega) \hat{f}_{k_t, k_{t-1}}^\frac{1}{2}
\quad - \beta_m \int m_t f_{m_t|k_t, k_{t-1}, w_{ti}} \, dm_t - q(c(m_{t-1, i}, k_{t-1}, p_{t-1, i}) \beta_\omega) \hat{f}_{k_t, k_{t-1}}^\frac{1}{2}
\quad - \beta_m \int m_t f_{m_t|k_t, k_{t-1}, w_{ti}} \, dm_t - q(c(m_{t-1, i}, k_{t-1}, p_{t-1, i}) \beta_\omega) \hat{f}_{k_t, k_{t-1}}^\frac{1}{2}
\leq \frac{1}{n} \sum_{i=1}^{n} \left\{ \left( \int y_t f_{yt|k_t, k_{t-1}, w_{ti}} \, dy_t - \beta_k l_t - \beta_k m_t - c(m_t, k_t, p_t) \right) \hat{f}_{k_t, k_{t-1}}^\frac{1}{2} \right\}^2
+ \left\{ \int y_t f_{yt|k_t, k_{t-1}, w_{ti}} \, dy_t - \beta_k l_t - \beta_k m_t - c(m_t, k_t, p_t) \hat{f}_{k_t, k_{t-1}}^\frac{1}{2} \right\}^2
\leq \frac{1}{n} \sum_{i=1}^{n} \left\{ \left( \int y_t f_{yt|k_t, k_{t-1}, w_{ti}} \, dy_t - \beta_k l_t - \beta_k m_t - c(m_t, k_t, p_t) \right) \hat{f}_{k_t, k_{t-1}}^\frac{1}{2} \right\}^2
+ \left\{ \int y_t f_{yt|k_t, k_{t-1}, w_{ti}} \, dy_t - \beta_k l_t - \beta_k m_t - c(m_t, k_t, p_t) \hat{f}_{k_t, k_{t-1}}^\frac{1}{2} \right\}^2
$$
\[ E \left[ \frac{1}{n} \sum_{i=1}^{n} \left\{ \left( \int y_t \left( f_{y_t[k_t,k_{t-1},w_{ti}]} \frac{\partial}{\partial w_{ti}} f_{k_t,k_{t-1}}(w_{ti}) - f_{y_t[k_t,k_{t-1},w_{ti}]} \frac{\partial}{\partial w_{ti}} f_{k_t,k_{t-1}}(w_{ti}) \right) \right)^2 \right\} \right] \]

\[ = E \left[ \frac{1}{n} \sum_{i=1}^{n} \left\{ \left( \int y_t \left( \hat{f}_{y_t[k_t,k_{t-1},w_{ti}]} - \frac{\partial}{\partial w_{ti}} f_{k_t,k_{t-1}}(w_{ti}) \right) \right)^2 \right\} \right] \]

\[ = E \left[ \frac{1}{n} \sum_{i=1}^{n} \left\{ \left( \int y_t \hat{f}_{y_t[k_t,k_{t-1},w_{ti}]} dy_t \hat{f}_{k_t,k_{t-1}}(w_{ti}) - \frac{\partial}{\partial w_{ti}} f_{k_t,k_{t-1}}(w_{ti}) \right)^2 \right\} \right] \]

\[ \leq E \left[ \frac{1}{n} \sum_{i=1}^{n} \left\{ \left( \int y_t \hat{f}_{y_t[k_t,k_{t-1},w_{ti}]} dy_t \left| \nu^{-1}(k_t,k_{t-1},z^*_i) \right| \left\| \frac{\partial}{\partial w_{ti}} f_{k_t,k_{t-1}}(w_{ti}) \right\| \right)^2 \right\} \right] \]

\[ + \int \left| \frac{df_{y_t[k_t,k_{t-1},w_{ti}]}^{-1}}{w_{ti}} \left( k_t,k_{t-1},z^*_i \right) \right| \left| \psi^{-1}(k_t,k_{t-1},z^*_i) \right| \left\| \psi^{-1}(k_t,k_{t-1},z^*_i) \right\| dy_t \left[ \frac{\partial}{\partial w_{ti}} f_{k_t,k_{t-1}}(w_{ti}) \right] \]

\[ + \left| \beta \hat{e}_{ti} + \beta_k \hat{e}_{k_t} + \beta_m \hat{e}_{m_t} + \mathbf{c}(m_t,k,p_t)' \beta_w \cdot \nu^{-1}(k_t,k_{t-1},z^*_i) \right| \left\| \frac{\partial}{\partial w_{ti}} f_{k_t,k_{t-1}}(w_{ti}) \right\| \left\| \frac{\partial}{\partial w_{ti}} f_{k_t,k_{t-1}}(w_{ti}) \right\| ^2 \]

\[ + \left( \int y_t \hat{f}_{y_t[k_t,k_{t-1},w_{ti}]} dy_t \left| \nu^{-1}(k_t,k_{t-1},z^*_i) \right| \left\| \frac{\partial}{\partial w_{ti}} f_{k_t,k_{t-1}}(w_{ti}) \right\| \right)^2 \]
because \( \| \hat{\alpha} - \alpha \|_{s, \alpha}^2 = o(1) \), and also because we can bound \( B_{1n} \) as below:

\[
B_{1n} = \left[ \frac{1}{n} \sum_{i=1}^{n} E \left\{ \left( \int y_t f_y \left[ k_t, k_{t-1}, \bar{w}_{ti} \right] d\nu^{-1} (k_t, k_{t-1}, z_t^*) \right)^2 + \int \frac{df_{y|k_t, k_{t-1}, w_{ti}}}{d\psi} \left[ \nu^{-1} (k_t, k_{t-1}, z_t^*) \right] \right\} \right] = O(1).
\]

We therefore obtain

\[
\sum_{n} \left\| \hat{g}(\mathbf{x}_n; \theta) f_{k_t, k_{t-1} | \mathbf{w}_{ti}^*} - g(\mathbf{x}_n; \theta) f_{k_t, k_{t-1} | \mathbf{w}_{ti}^*} \right\|_E^2 / n = o_p(1)
\]

by the Markov inequality. Note that \( Q_u(\theta) = Q(\theta) + o_p(1) \) by the weak law of large numbers. Therefore, we have verified Lemma A.2 Condition (ii) of Newey and Powell (2003) by the triangle inequality.

Next we verify Condition (iii) of their Lemma A.2. Since \( \| \cdot \|_{s, \theta}^r \) is bounded on \( \Theta \times \Theta \) by the compactness of the parameter space, there are constants \( C, \tilde{C} \) such that
\[
\left| \hat{Q}_n(\hat{\theta}) - \bar{Q}_n(\theta) \right| \\
\leq \frac{C}{n} \cdot \left( \sum_{i=1}^{n} \int \left( \left\| \hat{g}(x_{ti}; \hat{\theta}) \hat{f}_{k_t, k_t-1|x_{ti}}^{\frac{1}{2}} - \hat{g}(x_{ti}; \theta) \hat{f}_{k_t, k_t-1|x_{ti}}^{\frac{1}{2}} \right\|_E^2 \right) \, dk_t dk_{t-1} \right)
\]
\[
\leq \frac{C}{n} \cdot \left( \sum_{i=1}^{n} \int \left( \left\| \hat{g}(x_{ti}; \hat{\theta}) \hat{f}_{k_t, k_t-1|x_{ti}}^{\frac{1}{2}} - \hat{g}(x_{ti}; \theta) \hat{f}_{k_t, k_t-1|x_{ti}}^{\frac{1}{2}} \right\|_E^2 \right) \, dk_t dk_{t-1} \right)
\]
\[
+ 2 \sum_{i=1}^{n} \int \left( \left\| \hat{g}(x_{ti}; \theta) \hat{f}_{k_t, k_t-1|x_{ti}}^{\frac{1}{2}} \right\|_E \right) \, dk_t dk_{t-1}
\]
\[
\times \sum_{i=1}^{n} \int \left( \left\| \hat{g}(x_{ti}; \hat{\theta}) \hat{f}_{k_t, k_t-1|x_{ti}}^{\frac{1}{2}} - \hat{g}(x_{ti}; \theta) \hat{f}_{k_t, k_t-1|x_{ti}}^{\frac{1}{2}} \right\|_E \right) \, dk_t dk_{t-1}
\]
\[
= \frac{C}{n} \cdot \left( \sum_{i=1}^{n} \int \left( \left\| \hat{g}(x_{ti}; \hat{\theta}) \hat{f}_{k_t, k_t-1|x_{ti}}^{\frac{1}{2}} - \hat{g}(x_{ti}; \theta) \hat{f}_{k_t, k_t-1|x_{ti}}^{\frac{1}{2}} \right\|_E^2 \right) \, dk_t dk_{t-1} \right)
\]
\[
+ 2 \sum_{i=1}^{n} \int \left( \left\| \hat{g}(x_{ti}; \theta) \hat{f}_{k_t, k_t-1|x_{ti}}^{\frac{1}{2}} \right\|_E \right) \, dk_t dk_{t-1}
\]
\[
\times \sum_{i=1}^{n} \int \left( \left\| \hat{g}(x_{ti}; \hat{\theta}) \hat{f}_{k_t, k_t-1|x_{ti}}^{\frac{1}{2}} - \hat{g}(x_{ti}; \theta) \hat{f}_{k_t, k_t-1|x_{ti}}^{\frac{1}{2}} \right\|_E \right) \, dk_t dk_{t-1}
\]
\[
+ 2 \sum_{i=1}^{n} \int \left( \left\| \hat{g}(x_{ti}; \theta) \hat{f}_{k_t, k_t-1|x_{ti}}^{\frac{1}{2}} \right\|_E \right) \, dk_t dk_{t-1}
\]
\[
\times \sum_{i=1}^{n} \int \left( \left\| \hat{g}(x_{ti}; \hat{\theta}) \hat{f}_{k_t, k_t-1|x_{ti}}^{\frac{1}{2}} - \hat{g}(x_{ti}; \theta) \hat{f}_{k_t, k_t-1|x_{ti}}^{\frac{1}{2}} \right\|_E \right) \, dk_t dk_{t-1}
\]
\[
= \frac{C}{n} \left\{ \sum_{i=1}^{n} \left[ \left( \hat{g}_1(x_{ti}; \hat{\theta}) \hat{f}_{k_t, k_t-1|x_{ti}}^{\frac{1}{2}} - \hat{g}_1(x_{ti}; \beta) \hat{f}_{k_t, k_t-1|x_{ti}}^{\frac{1}{2}} \right)^2 \right. \right.
\]
\[
+ \left. \left( \hat{g}_2(\tilde{x}_{ti}; \hat{\theta}) \hat{f}_{k_t, k_t-1|x_{ti}}^{\frac{1}{2}} - \hat{g}_2(\tilde{x}_{ti}; \beta) \hat{f}_{k_t, k_t-1|x_{ti}}^{\frac{1}{2}} \right)^2 \right] \right\}^{1/2} \, dk_t dk_{t-1}
\]
\[
+ 2 \left\{ \sum_{i=1}^{n} \left[ \left( \hat{g}_1(x_{ti}; \hat{\theta}) \hat{f}_{k_t, k_t-1|x_{ti}}^{\frac{1}{2}} - \hat{g}_1(x_{ti}; \beta) \hat{f}_{k_t, k_t-1|x_{ti}}^{\frac{1}{2}} \right)^2 \right. \right.
\]
\[
+ \left. \left( \hat{g}_2(\tilde{x}_{ti}; \hat{\theta}) \hat{f}_{k_t, k_t-1|x_{ti}}^{\frac{1}{2}} - \hat{g}_2(\tilde{x}_{ti}; \beta) \hat{f}_{k_t, k_t-1|x_{ti}}^{\frac{1}{2}} \right)^2 \right] \right\}^{1/2} \, dk_t dk_{t-1}
\]
\[
+ 2 \left\{ \sum_{i=1}^{n} \left[ \left( \hat{g}_1(x_{ti}; \hat{\theta}) \hat{f}_{k_t, k_t-1|x_{ti}}^{\frac{1}{2}} - \hat{g}_1(x_{ti}; \beta) \hat{f}_{k_t, k_t-1|x_{ti}}^{\frac{1}{2}} \right)^2 \right. \right.
\]
\[
+ \left. \left( \hat{g}_2(\tilde{x}_{ti}; \hat{\theta}) \hat{f}_{k_t, k_t-1|x_{ti}}^{\frac{1}{2}} - \hat{g}_2(\tilde{x}_{ti}; \beta) \hat{f}_{k_t, k_t-1|x_{ti}}^{\frac{1}{2}} \right)^2 \right] \right\}^{1/2} \, dk_t dk_{t-1}
\]
\[
+ 2 \left\{ \sum_{i=1}^{n} \left[ \left( \hat{g}_1(x_{ti}; \hat{\theta}) \hat{f}_{k_t, k_t-1|x_{ti}}^{\frac{1}{2}} - \hat{g}_1(x_{ti}; \beta) \hat{f}_{k_t, k_t-1|x_{ti}}^{\frac{1}{2}} \right)^2 \right. \right.
\]
\[
+ \left. \left( \hat{g}_2(\tilde{x}_{ti}; \hat{\theta}) \hat{f}_{k_t, k_t-1|x_{ti}}^{\frac{1}{2}} - \hat{g}_2(\tilde{x}_{ti}; \beta) \hat{f}_{k_t, k_t-1|x_{ti}}^{\frac{1}{2}} \right)^2 \right] \right\}^{1/2} \, dk_t dk_{t-1}\}^\frac{1}{2}
\[
\begin{align*}
&= \frac{C}{n} \left\{ \sum_{i=1}^{n} \left[ \left( (\rho_1(y_i; \tilde{\beta}) - \rho_1(y_i; \beta)) \hat{f}_{y_i|x_t} dy_t \hat{f}_{k_t,k_{t-1}|w_{t_i}}^{1/2} \right)^2 \right. \\
&\quad + \left( (\rho_2(\tilde{y}_{t-1}; \hat{\tilde{\theta}}) - \rho_2(\tilde{y}_{t-1}; \theta)) \hat{f}_{\tilde{y}_{t-1}|x_{t_i}} d\tilde{y}_{t-1} \hat{f}_{k_{t-1}|w_{t_i}}^{1/2} \right)^2 \right] dk_t dk_{t-1} \\
&\quad + 2 \left\{ \sum_{i=1}^{n} \left[ \left( (\rho_1(y_i; \beta) - \rho_1(y_i; \beta_0)) \hat{f}_{y_i|x_t} dy_t \hat{f}_{k_t,k_{t-1}|w_{t_i}}^{1/2} \right)^2 \right. \\
&\quad + \left( (\rho_2(\tilde{y}_{t-1}; \tilde{\theta}) - \rho_2(\tilde{y}_{t-1}; \theta_0)) \hat{f}_{\tilde{y}_{t-1}|x_{t_i}} d\tilde{y}_{t-1} \hat{f}_{k_{t-1}|w_{t_i}}^{1/2} \right)^2 \right] \right\} \frac{1}{2} dk_t dk_{t-1} \\
&\quad + 2 \left\{ \sum_{i=1}^{n} \left[ \left( (\rho_1(y_i; \beta) - \rho_1(y_i; \beta_0)) \hat{f}_{y_i|x_t} dy_t \hat{f}_{k_t,k_{t-1}|w_{t_i}}^{1/2} \right)^2 \right. \\
&\quad + \left( (\rho_2(\tilde{y}_{t-1}; \tilde{\theta}) - \rho_2(\tilde{y}_{t-1}; \theta_0)) \hat{f}_{\tilde{y}_{t-1}|x_{t_i}} d\tilde{y}_{t-1} \hat{f}_{k_{t-1}|w_{t_i}}^{1/2} \right)^2 \right] \right\} \frac{1}{2} dk_t dk_{t-1} \\
&\quad + \left\{ \sum_{i=1}^{n} \left[ \left( (\rho_1(y_i; \beta) - \rho_1(y_i; \beta_0)) \hat{f}_{y_i|x_t} dy_t \hat{f}_{k_t,k_{t-1}|w_{t_i}}^{1/2} \right)^2 \right. \\
&\quad + \left( (\rho_2(\tilde{y}_{t-1}; \tilde{\theta}) - \rho_2(\tilde{y}_{t-1}; \theta_0)) \hat{f}_{\tilde{y}_{t-1}|x_{t_i}} d\tilde{y}_{t-1} \hat{f}_{k_{t-1}|w_{t_i}}^{1/2} \right)^2 \right] \right\} \frac{1}{2} dk_t dk_{t-1} \\
&\quad \leq \frac{C}{n} \left\{ \sum_{i=1}^{n} \left[ \left( b_1(y_i) \hat{f}_{y_i|x_t} dy_t \hat{f}_{k_t,k_{t-1}|w_{t_i}}^{1/2} \right)^2 \cdot \| \hat{\theta} - \theta \|^{2\nu}_{s,\theta} \right] dk_t dk_{t-1} \\
&\quad + \left\{ \sum_{i=1}^{n} \left[ \left( b_2(\tilde{y}_{t-1}) \hat{f}_{\tilde{y}_{t-1}|x_{t_i}} d\tilde{y}_{t-1} \hat{f}_{k_{t-1}|w_{t_i}}^{1/2} \right)^2 \cdot \| \hat{\theta} - \theta \|^{2\nu}_{s,\theta} \right] dk_t dk_{t-1} \\
&\quad + 2 \left\{ \sum_{i=1}^{n} \left[ \left( b_1(y_i) \hat{f}_{y_i|x_t} dy_t \hat{f}_{k_t,k_{t-1}|w_{t_i}}^{1/2} \right)^2 \cdot \left( b_2(\tilde{y}_{t-1}) \hat{f}_{\tilde{y}_{t-1}|x_{t_i}} d\tilde{y}_{t-1} \hat{f}_{k_{t-1}|w_{t_i}}^{1/2} \right)^2 \right] \right\} \frac{1}{2} dk_t dk_{t-1} \\
&\quad + 2 \left\{ \sum_{i=1}^{n} \left[ \left( b_1(y_i) \hat{f}_{y_i|x_t} dy_t \hat{f}_{k_t,k_{t-1}|w_{t_i}}^{1/2} \right)^2 \cdot \left( b_2(\tilde{y}_{t-1}) \hat{f}_{\tilde{y}_{t-1}|x_{t_i}} d\tilde{y}_{t-1} \hat{f}_{k_{t-1}|w_{t_i}}^{1/2} \right)^2 \right] \right\} \frac{1}{2} dk_t dk_{t-1} \\
&\quad + 2 \left\{ \sum_{i=1}^{n} \left[ \left( b_1(y_i) \hat{f}_{y_i|x_t} dy_t \hat{f}_{k_t,k_{t-1}|w_{t_i}}^{1/2} \right)^2 \cdot \left( b_2(\tilde{y}_{t-1}) \hat{f}_{\tilde{y}_{t-1}|x_{t_i}} d\tilde{y}_{t-1} \hat{f}_{k_{t-1}|w_{t_i}}^{1/2} \right)^2 \right] \right\} \frac{1}{2} dk_t dk_{t-1} \\
&\quad \leq B_{2n} \| \hat{\theta} - \theta \|^{\nu}_{s,\theta}. \right. \right. \right. \right. \end{align*}
\]
where \(b_1(y_t)\) and \(b_2(\tilde{y}_{t,t-1})\) are some measurable functions with bounded second moments and 
\(B_{2n} = \tilde{C}B_n\), for which by Assumption 5.7 we obtain

\[
\tilde{B}_n = \frac{1}{n} \left\{ \sum_{i=1}^{n} \left[ (f b_1(y_t) \hat{f}_{yt|x_t} dy_t \hat{f}_{kt,kt-1|w_t})^2 + (f b_2(\tilde{y}_{t,t-1}) \hat{f}_{yt,t-1|\tilde{x}_t} dy_{t,t-1} \hat{f}_{kt,kt-1|w_t})^2 \right] dk_t dk_{t-1} \right. \\
+ 2 \sum_{i=1}^{n} \left\{ \left[ (f \rho_1(y_t; \beta_0) \hat{f}_{yt|x_t} dy_t)^2 + (f \rho_2(\tilde{y}_{t,t-1}; \theta_0) \hat{f}_{yt,t-1|\tilde{x}_t} d\tilde{y}_{t,t-1})^2 \right] \right\}^{1/2} \hat{f}_{kt,kt-1|w_t} dk_t dk_{t-1} \\
\left. \cdot \left\{ \sum_{i=1}^{n} \left[ (f b_1(y_t) \hat{f}_{yt|x_t} dy_t \hat{f}_{kt,kt-1|w_t})^2 \right] \right\}^{1/2} dk_t dk_{t-1} \right\}
\]

\(= Op(1)\).

Therefore, we satisfy Condition (iii) of Lemma A.2. We have verified all three conditions in Lemma A.1 of Newey and Powell (2003) and hence the result follows.

### C.2 Convergence Rate: Proof of Theorem 5.21

The first result (Theorem 5.21.1) follows from a similar argument to Theorem 2 in Hu and Schennach (2008). The second result (Theorem 5.21.2) follows from a similar argument to Ai and Chen (2003) and Song (2015)’s convergence rate results. First we show two lemmas: Lemma C.1 and Lemma C.2 below, which are useful to prove Theorem 5.21.

**Lemma C.1.** Suppose Assumptions for Theorem 5.10 hold. Then (a) under Assumptions 5.4, 5.7, 5.11, 5.12, 5.13, 5.16, 5.17, 5.18, 5.20, 
\[
\sum_{i=1}^{n} \left[ \int \left\| \hat{g}(x_t; \theta) \hat{f}_{kt,kt-1|w_t} - g(x_t; \theta) f_{kt,kt-1|w_t} \right\|^2_{\alpha} dE dk_t dk_{t-1} \right] / n
\]

\(= op(n^{-1/2})\) uniformly over \(\theta \in \Theta\); (b) Under Assumptions 5.5 and 5.7, 
\[
\sum_{i=1}^{n} \left[ \int \left\| \hat{g}(x_t; \theta_0) \hat{f}_{kt,kt-1|w_t} \right\|^2_{\alpha} dE dk_t dk_{t-1} / n \right] = op(\kappa_1 n / n).
\]

**Proof.** (a) Since \(\left\| \hat{g} - \alpha \right\|_{\alpha} = op(n^{-1/4})\) by a similar argument to Theorem 2 in Hu and Schennach (2008), \(\left\| \hat{f}_{kt,kt-1|w_t} - f_{kt,kt-1|w_t} \right\|^2_{\alpha} = op(n^{-1/2})\). From

\[
\sum_{i=1}^{n} \left[ \int \left\| \hat{g}(x_t; \theta) \hat{f}_{kt,kt-1|w_t} - g(x_t; \theta) f_{kt,kt-1|w_t} \right\|^2_{\alpha} dE dk_t dk_{t-1} \right] / n \leq C \cdot \sum_{i=1}^{n} \left[ \int \left\| \hat{g}(x_t; \theta) - g(x_t; \theta) \right\|^2_{\alpha} dE dk_t dk_{t-1} \right] + \left\| \hat{f}_{kt,kt-1|w_t} - f_{kt,kt-1|w_t} \right\|^2_{\alpha} / n + op(n^{-1/2}),
\]

for a generic constant \(C\), it suffices to show \(\sum_{i=1}^{n} \left[ \int \left\| \hat{g}(x_t; \theta) - g(x_t; \theta) \right\|^2_{\alpha} dE dk_t dk_{t-1} \right] / n = op(n^{-1/2})\) uniformly over \(\theta \in \Theta\). For \(\tilde{w}_t = (\tilde{p}_t, m_{t-1}, l_{t-1})\) and \(w_t = (\tilde{w}_t, m_t, l_t)\), we note that from Corollary 4.4, equation (13), and Assumption 5.20, \(\left\| \hat{f}_{yt|kt,kt-1,\tilde{w}_t} - f_{yt|kt,kt-1,\tilde{w}_t} \right\|^2_{\alpha} = C\left\| \hat{f}_{yt|kt,kt-1,w_t} - f_{yt|kt,kt-1,w_t} \right\|^2_{\alpha} = op(n^{-1/2})\).
\[ \| f_{y_t|k_t, k_{t-1}, \omega_t} \|_\alpha + o_p(n^{-1/4}) \] for a generic constant \( C \). We also note that \( \| \tilde{f}_{r_t|k_t, k_{t-1}, \omega_t} - f_{r_t|k_t, k_{t-1}, \omega_t} \|_\alpha \leq \| \tilde{f}_{y_t|k_t, k_{t-1}, \omega_t} - f_{y_t|k_t, k_{t-1}, \omega_t} \|_\alpha \) for \( r_t \in \{l_t, m_t\} \). Let \( \rho(\cdot)(j) \) denote the \( j \)-th element of \( \rho(\cdot) \). It follows that

\[
E \sum_{i=1}^{n} \left[ \int \| \hat{\theta}(x_{ti}; \theta) - g(x_{ti}; \theta) \|_E dk_t dk_{t-1} \right]^2 / n
\]

\[
= \frac{1}{n} \sum_{i=1}^{n} \sum_{j=1}^{2} E \left( \int \rho(y_{t, t-1}; \theta)(j) \tilde{f}_{y_{t, t-1}|x_{ti}} dy_{t, t-1} - \rho(y_{t, t-1}; \theta)(j) f_{y_{t, t-1}|x_{ti}} dy_{t, t-1} \right)^2 dk_t dk_{t-1}
\]

\[
= \frac{1}{n} \sum_{i=1}^{n} \sum_{j=1}^{2} E \left[ \left( \int y_t \tilde{f}_{y|k_t, k_{t-1}, \omega_t} dy_t - \beta_t l_t - \beta_k k_t - \beta_m m_t - c(m_t, k_t, p_t) \beta_\omega \right)
- \left( \int y_t f_{y|k_t, k_{t-1}, \omega_t} dy_t - \beta_t l_t - \beta_k k_t - \beta_m m_t - c(m_t, k_t, p_t) \beta_\omega \right) \right] dk_t dk_{t-1}
\]

\[
= \frac{1}{n} \sum_{i=1}^{n} \sum_{j=1}^{2} E \left\{ \left( \int y_t \tilde{f}_{y|k_t, k_{t-1}, \omega_t} dy_t - \beta_t l_t - \beta_k k_t - \beta_m m_t - c(m_t, k_t, p_t) \beta_\omega \right) \right\}^2 dk_t dk_{t-1}
\]

\[
\leq C \frac{1}{n} \sum_{i=1}^{n} E \left\{ \left( \int y_t (\tilde{f}_{y|k_t, k_{t-1}, \omega_t} - f_{y|k_t, k_{t-1}, \omega_t}) dy_t dk_t dk_{t-1} \right)^2 \right\} + o_p(n^{-1/2})
\]

\[
\leq C \frac{1}{n} \sum_{i=1}^{n} E \left\{ \left( \int y_t (\tilde{f}_{k^*_t, k_{t-1}^*|z_t} - f_{k^*_t, k_{t-1}^*|z_t}) dy_t dk_t dk_{t-1} \right)^2 \right\} + o_p(n^{-1/2})
\]

\[
\leq C \frac{1}{n} \sum_{i=1}^{n} E \left( \int y_t (\tilde{f}_{k^*_t, k_{t-1}^*|z_t} - f_{k^*_t, k_{t-1}^*|z_t}) dy_t dk_t dk_{t-1} \right)^2 + o_p(n^{-1/2})
\]
Proof. 1. The first result follows from a similar argument to the proof of Theorem 2 in Hu and Schennach (2008).

2. From Lemma C.1 (a) and Assumption 5.6, we obtain \( \hat{Q}_n(\theta) - Q_n(\theta) = o_p(n^{-1/4}) \) uniformly over \( \theta \in \Theta_n \). In addition, from Lemma C.1 and Lemma C.2 (b), we obtain \( \hat{Q}_n(\theta) - \hat{Q}_n(\theta_0) - \{Q_n(\theta) - Q_n(\theta_0)\} = o_p(\eta_n n^{-1/4}) \) uniformly over \( \theta \in \{\Theta_n : \|\theta - \theta_0\|_\theta = o(\eta_n)\} \), where \( \eta_n = n^{-\tau} \) with \( \tau \leq 1/4 \). Then the second result follows by a similar argument to the proof of Theorem 3.1 in Ai and Chen (2003).

Lemma C.2. (a) Under Assumptions 5.1, 5.3, 5.5, 5.6.2, 5.11, 5.13.1, and 5.19, we obtain
\[
\sum_{i=1}^{n} \|g(x_{ti}; \theta)\|^2_E/n - E \left[\|g(x_{ti}; \theta)\|^2_E\right] = o_p(n^{-1/2}) \text{ uniformly over } \theta \in \{\Theta_n : \|\theta - \theta_0\|_\theta = o(1)\};
\]
(b) Under Assumptions 5.1, 5.3, 5.5, 5.6, 5.7, 5.9, 5.11, 5.13, 5.15, 5.18, 5.19, we obtain
\[
\sum_{i=1}^{n} \|g(x_{ti}; \theta)\|^2_E/n = o_p(\eta_n^2) \text{ and } \sum_{i=1}^{n} \|g(x_{ti}; \theta)\|^2_E/n = o_p(\eta_n^2) \text{ uniformly over } \theta \in \{\Theta_n : \|\theta - \theta_0\|_\theta = o(\eta_n)\},
\]
where \( \eta_n = n^{-\tau} \) with \( \tau \leq 1/4 \).

Proof. (a) The result follows from Corollary A.2 (i) in Ai and Chen (2003).

(b) Note that \( E[\|g(x_{ti}; \theta)\|^2_E] = o(\eta_n^2) \) by Assumptions 5.6.2 and 5.19. Then the result follows from applying Lemma C.1 (a) and Lemma C.2 (a).
C.3 Asymptotic Normality: Proof of Theorem 6.12

The first result (Theorem 6.12.1) follows from a similar argument to Theorem 3 in Hu and Schennach (2008). The second result (Theorem 6.12.2) follows from a similar argument to Ai and Chen (2003) and Song (2015)'s asymptotic normality results. First we show two lemmas: Lemma C.3 and Lemma C.4 below, which are useful to prove Theorem 6.12.

Define

\[
\frac{dQ_n(\theta)}{d\theta}[v^*_2] = \frac{1}{n} \sum_{i=1}^{n} \left\{ \frac{d g(x_{ti}; \theta)}{d\theta}[v^*_2] \right\} A(x_{ti}) g(x_{ti}; \theta) f_{k_t,k_{t-1}|w_t} dk_t dk_{t-1},
\]

\[
\frac{d\hat{Q}_n(\theta)}{d\theta}[v^*_2] = \frac{1}{n} \sum_{i=1}^{n} \left\{ \frac{d \hat{g}(x_{ti}; \theta)}{d\theta}[v^*_2] \right\} \hat{A}(x_{ti}) \hat{g}(x_{ti}; \theta) \hat{f}_{k_t,k_{t-1}|w_t} dk_t dk_{t-1},
\]

and similarly define

\[
\frac{d^2Q_{1n}(\theta)}{d\theta d\theta}[v^*_2, v^*_2] = \frac{1}{n} \sum_{i=1}^{n} \left\{ \frac{d g(x_{ti}; \theta)}{d\theta}[v^*_2] \right\} \left\{ \frac{d g(x_{ti}; \theta)}{d\theta}[v^*_2] \right\} A(x_{ti}) g(x_{ti}; \theta) f_{k_t,k_{t-1}|w_t} dk_t dk_{t-1},
\]

\[
\frac{d^2\hat{Q}_{1n}(\theta)}{d\theta d\theta}[v^*_2, v^*_2] = \frac{1}{n} \sum_{i=1}^{n} \left\{ \frac{d \hat{g}(x_{ti}; \theta)}{d\theta}[v^*_2] \right\} \left\{ \frac{d \hat{g}(x_{ti}; \theta)}{d\theta}[v^*_2] \right\} \hat{A}(x_{ti}) \hat{g}(x_{ti}; \theta) \hat{f}_{k_t,k_{t-1}|w_t} dk_t dk_{t-1},
\]

\[
\frac{d^2Q_{2n}(\theta)}{d\theta d\theta}[v^*_2, v^*_2] = \frac{1}{n} \sum_{i=1}^{n} \left\{ \frac{d^2 g(x_{ti}; \theta)}{d\theta d\theta}[v^*_2,v^*_2] \right\} A(x_{ti}) g(x_{ti}; \theta) f_{k_t,k_{t-1}|w_t} dk_t dk_{t-1},
\]

\[
\frac{d^2\hat{Q}_{2n}(\theta)}{d\theta d\theta}[v^*_2, v^*_2] = \frac{1}{n} \sum_{i=1}^{n} \left\{ \frac{d^2 \hat{g}(x_{ti}; \theta)}{d\theta d\theta}[v^*_2,v^*_2] \right\} \hat{A}(x_{ti}) \hat{g}(x_{ti}; \theta) \hat{f}_{k_t,k_{t-1}|w_t} dk_t dk_{t-1},
\]

where

\[
\frac{d g(x_{ti}; \theta)}{d\theta}[v^*_2] = \int \frac{d \rho(y_{t,t-1}; \theta)}{d\theta}[v^*_2] f_{y_{t,t-1}|x_t} dy_{t,t-1},
\]

\[
\frac{d^2 g(x_{ti}; \theta)}{d\theta d\theta}[v^*_2,v^*_2] = \int \frac{d^2 \rho(y_{t,t-1}; \theta)}{d\theta d\theta}[v^*_2,v^*_2] f_{y_{t,t-1}|x_t} dy_{t,t-1},
\]

Lemma C.3. (a) Under Assumptions 5.1, 5.3-5.4, 5.6-5.7, 5.9, 6.6.2-6.9, we have

\[
\sup_{\theta \in \Theta_{02n}} \frac{d^2 \hat{Q}_{1n}(\theta)}{d\theta d\theta}[v^*_2, v^*_2] = \frac{d^2 Q_{1n}(\theta_0)}{d\theta d\theta}[v^*_2, v^*_2] + o_p(n^{-1/4}).
\]

(b) Under Assumptions 5.1-5.4, 5.6-5.9, 5.11-5.20, 6.11, we have

\[
\sup_{\theta \in \Theta_{02n}} \frac{d^2 \hat{Q}_{2n}(\theta)}{d\theta d\theta}[v^*_2, v^*_2] = o_p(n^{-1/4}).
\]
Proof. (a) We have that uniformly over \( \tilde{\theta} \in \mathcal{N}_{02n} \),

\[
\frac{d^2 \hat{Q}_{1n}(\tilde{\theta})}{d\tilde{\theta}d\theta} [v_{2n}^*, v_{2n}^*] - \frac{d^2 Q_{1n}(\theta_0)}{d\theta d\theta} [v_{2n}^*, v_{2n}^*]
\]

\[
= \frac{1}{n} \sum_{i=1}^{n} \left\{ \left( \frac{d\hat{g}(x_{ti}; \theta)}{d\theta} [v_{2n}^*] - \frac{dg(x_{ti}; \theta_0)}{d\theta} [v_{2n}^*] \right) \hat{A}(x_{ti}) \left( \frac{d\hat{g}(x_{ti}; \theta)}{d\theta} [v_{2n}^*] \right) \hat{f}_{k_t, k_{t-1}|\theta_{ti}} \right.
\]

\[
+ \left\{ \left( \frac{dg(x_{ti}; \theta_0)}{d\theta} [v_{2n}^*] \right) \hat{A}(x_{ti}) \left( \frac{d\hat{g}(x_{ti}; \theta)}{d\theta} [v_{2n}^*] - \frac{dg(x_{ti}; \theta_0)}{d\theta} [v_{2n}^*] \right) \hat{f}_{k_t, k_{t-1}|\theta_{ti}} \right.
\]

\[
+ \left\{ \left( \frac{dg(x_{ti}; \theta_0)}{d\theta} [v_{2n}^*] \right) \hat{A}(x_{ti}) \left( \frac{dg(x_{ti}; \theta_0)}{d\theta} [v_{2n}^*] \right) \hat{f}_{k_t, k_{t-1}|\theta_{ti}} \right\} dk_t dk_{t-1}
\]

\[
\equiv A_1 + A_2 + A_3 + A_4.
\]

Since

\[
E \left[ \frac{1}{n} \sum_{i=1}^{n} \left\{ \left( \frac{d\hat{g}(x_{ti}; \theta)}{d\theta} [v_{2n}^*] - \frac{dg(x_{ti}; \theta_0)}{d\theta} [v_{2n}^*] \right) \right\}^2 \right]
\]

\[
= \frac{1}{n} \sum_{i=1}^{n} \sum_{j=1}^{n} E \left[ \left( \int \frac{dp(y_{t,t-1}; \theta)(j)}{d\theta} [v_{2n}^*] \left( \hat{f}_{y_{t,t-1}|\theta_{ti}} - f_{y_{t,t-1}|\theta_{ti}} \right) dy_{t,t-1} \right)^2 \right]
\]

\[
= o(n^{-1/2}),
\]

uniformly over \( \tilde{\theta} \in \mathcal{N}_{02n} \) from \( \|\hat{\alpha} - \alpha_0\|_\alpha = o_p(n^{-1/4}) \), we obtain \( A_1 = A_3 = A_4 = o_p(n^{-1/4}) \) by the Markov inequality, Assumption 5.6, and Lemma C.2 (b). Also we obtain \( A_2 = o_p(n^{-1/4}) \) from Assumption 5.15 and Lemma C.2 (b).

(b) By Assumption 5.6 and Lemma C.2 (b), we obtain that for some generic constant \( C \),

\[
\left| \frac{d^2 \hat{Q}_{2n}(\tilde{\theta})}{d\tilde{\theta}d\theta} [v_{2n}^*, v_{2n}^*] \right|
\]

\[
\leq C \sqrt{ \frac{1}{n} \sum_{i=1}^{n} \int \left\| \frac{d^2 \hat{g}(x_{ti}; \tilde{\theta})}{d\theta d\theta} [v_{2n}^*, v_{2n}^*] \right\|_E^2 dk_t dk_{t-1} \cdot \sqrt{ \frac{1}{n} \sum_{i=1}^{n} \int \left\| \hat{g}(x_{ti}; \tilde{\theta}) \right\|_E^2 dk_t dk_{t-1}}
\]

\[
= o_p(n^{-1/4}),
\]

since Assumption 6.11 implies that uniformly over \( \tilde{\theta} \in \mathcal{N}_{02n} \), for a measurable function \( c(y_{t,t-1}) \) with bounded second moment,

\[
\left\| \frac{d^2 \hat{g}(x_{ti}; \tilde{\theta})}{d\theta d\theta} [v_{2n}^*, v_{2n}^*] \right\|_E^2 \leq \frac{1}{n} \sum_{i=1}^{n} c(y_{t,t-1,i})^2 = O_p(1).
\]

\( \square \)
Lemma C.4. (a) Under Assumptions 5.1-5.9, 6.6.2-6.9, we have uniformly over \( \tilde{\theta} \in \mathcal{N}_{0^2n} \),

\[
\frac{d\hat{Q}_n(\tilde{\theta})}{d\theta} [v_{2n}^*] = \frac{1}{n} \sum_{i=1}^{n} \left\{ \frac{dg(x_{ti}; \theta_0)}{d\theta} [v_2^*] \right\}' A(x_{ti}) \hat{g}(x_{ti}; \tilde{\theta}) f_{k_t,k_{t-1}|w_{ti}} dk_t dk_{t-1}.
\]

(b) Under Assumptions 5.1, 5.3, 5.12, 6.6.2-6.9, we have uniformly over \( \tilde{\theta} \in \mathcal{N}_{0^2n} \),

\[
\frac{1}{n} \sum_{i=1}^{n} \left\{ \frac{dg(x_{ti}; \theta_0)}{d\theta} [v_2^*] \right\}' A(x_{ti}) \{ \hat{g}(x_{ti}; \tilde{\theta}) - \hat{g}(x_{ti}; \theta_0) \} f_{k_t,k_{t-1}|w_{ti}} dk_t dk_{t-1} = (v_2^*, \tilde{\theta} - \theta_0)_{\tilde{\theta}} + o_p(n^{-1/2}).
\]

(c) Under Assumptions 5.1, 5.3, 5.4, 5.6.2, 5.12, 5.16, 6.6.3, we have

\[
\frac{1}{n} \sum_{i=1}^{n} \left\{ \frac{dg(x_{ti}; \theta_0)}{d\theta} [v_2^*] \right\}' A(x_{ti}) \hat{g}(x_{ti}; \theta_0) f_{k_t,k_{t-1}|w_{ti}} dk_t dk_{t-1} = \frac{1}{n} \sum_{i=1}^{n} \left\{ \frac{dg(x_{ti}; \theta_0)}{d\theta} [v_2^*] \right\}' A(x_{ti}) \rho(y_{t-1,i}; \theta_0) f_{k_t,k_{t-1}|w_{ti}} dk_t dk_{t-1} + o_p(n^{-1/2}).
\]

Proof. (a) Note that uniformly over \( \tilde{\theta} \in \mathcal{N}_{0^2n} \),

\[
\frac{d\hat{Q}_n(\tilde{\theta})}{d\theta} [v_{2n}^*] = \frac{1}{n} \sum_{i=1}^{n} \left\{ \frac{dg(x_{ti}; \theta_0)}{d\theta} [v_2^*] \right\}' A(x_{ti}) \hat{g}(x_{ti}; \tilde{\theta}) f_{k_t,k_{t-1}|w_{ti}} dk_t dk_{t-1}
\]

\[
= \frac{1}{n} \sum_{i=1}^{n} \left\{ \frac{d\hat{g}(x_{ti}; \theta_0)}{d\theta} [v_{2n}^*] - \frac{dg(x_{ti}; \theta_0)}{d\theta} [v_{2n}^*] \right\}' \hat{A}(x_{ti}) \hat{g}(x_{ti}; \tilde{\theta}) f_{k_t,k_{t-1}|w_{ti}}
\]

\[
+ \left\{ \frac{d\hat{g}(x_{ti}; \theta_0)}{d\theta} [v_{2n}^*] \right\}' \{ \hat{A}(x_{ti}) - A(x_{ti}) \} \hat{g}(x_{ti}; \tilde{\theta}) f_{k_t,k_{t-1}|w_{ti}}
\]

\[
+ \left\{ \frac{dg(x_{ti}; \theta_0)}{d\theta} [v_2^*] \right\}' A(x_{ti}) \{ \hat{g}(x_{ti}; \tilde{\theta}) f_{k_t,k_{t-1}|w_{ti}} - f_{k_t,k_{t-1}|w_{ti}} \}
\]

\[
+ \left\{ \frac{dg(x_{ti}; \theta_0)}{d\theta} [v_{2n}^* - v_2^*] \right\}' A(x_{ti}) \hat{g}(x_{ti}; \tilde{\theta}) f_{k_t,k_{t-1}|w_{ti}} dk_t dk_{t-1}
\]

\[
= D_1 + D_2 + D_3 + D_4.
\]

Also note that from the proof of Lemma C.3 (a)

\[
\sup_{\tilde{\theta} \in \mathcal{N}_{0^2n}} \frac{1}{n} \sum_{i=1}^{n} \left\| \frac{d\hat{g}(x_{ti}; \theta)}{d\theta} [v_{2n}^*] - \frac{dg(x_{ti}; \theta)}{d\theta} [v_{2n}^*] \right\|_{E}^2 = o_p(n^{-1/2})
\]

42
and from Corollary C.1 (ii) in Ai and Chen (2003)

\[ \sup_{\theta \in \mathcal{N}_{02n}} \frac{1}{n} \sum_{i=1}^{n} \left\| \frac{dg(x_{ti}; \theta)}{d\theta} [v_{2n}^*] - \frac{dg(x_{ti}; \hat{\theta})}{d\theta} [v_{2n}^*] \right\|_E^2 = o_p(n^{-1/2}). \]

Thus \( D_1 = o_p(n^{-1/2}) \) by Lemma C.2 (b) and Assumption 5.6. \( D_2 = o_p(n^{-1/2}) \) by Assumption 5.15 and Lemma C.2 (b). \( D_3 = o_p(n^{-1/2}) \) by \( \| \hat{\theta} - \theta_0 \|_\alpha = o_p(n^{-1/4}) \) and Lemma C.2 (b). \( D_4 = o_p(n^{-1/2}) \) by Assumption 6.7 and Lemma C.2 (b). Thus the result follows.

(b) Define \( \varphi(x_t, v_2^*) = \left\{ \frac{dg(x_t; \theta)}{d\theta} [v_{2n}^*] \right\}' A(x_t) \) and

\[ \mathcal{F} = \left\{ \varphi(x_t, v_2^*) \hat{g}(x_t; \theta) : \theta \in \mathcal{N}_{02n}, \hat{g} \in \Lambda^n(\mathcal{x}_t) \text{ s.t.} \sup_{x_t \in \mathcal{x}_t, \theta \in \mathcal{N}_{02n}} |\hat{g}(x_t; \theta) - g(x_t; \theta)| = o(1) \right\}. \]

By the argument of Corollary C.3 (ii) in Ai and Chen (2003), \( \mathcal{F} \) is a Donsker class. We also note that \( E[\|\varphi(x_{ti}, v_2^*) \{\hat{g}(x_{ti}; \theta) - g(x_{ti}; \theta)\}\|^2] = o_p(1) \) and \( \hat{g}(x_{ti}; \theta) \in \Lambda^n(\mathcal{x}_t) \) uniformly over \( \theta \in \mathcal{N}_{02n} \) by Assumptions 5.1, 5.3, 5.4, 5.7, 5.18. Thus we obtain uniformly over \( \theta \in \mathcal{N}_{02n} \),

\[ \frac{1}{n} \sum_{i=1}^{n} \varphi(x_{ti}, v_2^*) \{\hat{g}(x_{ti}; \theta) - g(x_{ti}; \theta)\} - E[\varphi(x_{ti}, v_2^*) \{\hat{g}(x_{ti}; \theta) - g(x_{ti}; \theta)\}] = o_p(n^{-\frac{1}{2}}) \quad (14) \]

\[ \frac{1}{n} \sum_{i=1}^{n} \varphi(x_{ti}, v_2^*) \{\hat{g}(x_{ti}; \theta_0) - g(x_{ti}; \theta_0)\} - E[\varphi(x_{ti}, v_2^*) \{\hat{g}(x_{ti}; \theta_0) - g(x_{ti}; \theta_0)\}] = o_p(n^{-\frac{1}{2}}) \quad (15) \]

by Lemma 1 of Chen, Linton, and van Keilegom (2003). Combining (14) and (15) we obtain (because \( \hat{\theta} \in \mathcal{N}_{02n} \))

\[ \frac{1}{n} \sum_{i=1}^{n} \varphi(x_{ti}, v_2^*) \{\hat{g}(x_{ti}; \hat{\theta}) - \hat{g}(x_{ti}; \theta_0)\} - \frac{1}{n} \sum_{i=1}^{n} \varphi(x_{ti}, v_2^*) \{g(x_{ti}; \hat{\theta}) - g(x_{ti}; \theta_0)\} = E[\varphi(x_{ti}, v_2^*) \{\hat{g}(x_{ti}; \hat{\theta}) - \hat{g}(x_{ti}; \theta_0)\}] - E[\varphi(x_{ti}, v_2^*) \{g(x_{ti}; \hat{\theta}) - g(x_{ti}; \theta_0)\}] + o_p(n^{-1/2}). \]

Now define \( \hat{\varphi}(x_t, v_2^*) = \int \varphi(x_t, v_2^*) \hat{f}_{y_{t-1}|x_t} dy_{t-1} \). We then obtain

\[ E[\varphi(x_{ti}, v_2^*) \{\hat{g}(x_{ti}; \hat{\theta}) - \hat{g}(x_{ti}; \theta_0)\}] = E[\varphi(x_{ti}, v_2^*) \int \rho(y_{t-1}; \hat{\theta}) - \rho(y_{t-1}; \theta_0) \hat{f}_{y_{t-1}|x_t} dy_{t-1}] \]

\[ = E[\int \varphi(x_{ti}, v_2^*) \hat{f}_{y_{t-1}|x_t} dy_{t-1} E[\rho(y_{t-1}; \hat{\theta}) - \rho(y_{t-1}; \theta_0) | x_{ti}] \]

\[ = E[\hat{\varphi}(x_{ti}, v_2^*) \{g(x_{ti}; \hat{\theta}) - g(x_{ti}; \theta_0)\}] \]

and

\[ E[\hat{\varphi}(x_{ti}, v_2^*) \{g(x_{ti}; \hat{\theta}) - g(x_{ti}; \theta_0)\}] = E[\hat{\varphi}(x_{ti}, v_2^*) \{g(x_{ti}; \hat{\theta}) - g(x_{ti}; \theta_0)\}] + o_p(n^{-1/2}). \]
It follows that
\[
\frac{1}{n} \sum_{i=1}^{n} \varphi(x_{ti}, v_2^*) \{\hat{g}(x_{ti}; \hat{\theta}) - \hat{g}(x_{ti}; \theta_0)\} - \frac{1}{n} \sum_{i=1}^{n} \varphi(x_{ti}, v_2^*) \{g(x_{ti}; \hat{\theta}) - g(x_{ti}; \theta_0)\} = o_p(n^{-1/2}).
\]
Then for some \( \tilde{\theta} \in \mathcal{N}_{02n} \), which lies between \( \hat{\theta} \) and \( \theta_0 \), the mean value theorem (applied in the third equality) yields
\[
\frac{1}{n} \sum_{i=1}^{n} \varphi(x_{ti}, v_2^*) \{g(x_{ti}; \tilde{\theta}) - g(x_{ti}; \theta_0)\} = E[\varphi(x_{ti}, v_2^*) \{g(x_{ti}; \tilde{\theta}) - g(x_{ti}; \theta_0)\}] + o_p(n^{-1/2})
\]
\[
= \langle v_2^*, \tilde{\theta} - \theta_0 \rangle + E \left[ \varphi(x_{ti}, v_2^*) \left( \frac{dg(x_{ti}; \tilde{\theta})}{d\theta} [\tilde{\theta} - \theta_0] - \frac{dg(x_{ti}; \theta_0)}{d\theta} [\tilde{\theta} - \theta_0] \right) \right] + o_p(n^{-1/2})
\]
\[
= \langle v_2^*, \tilde{\theta} - \theta_0 \rangle + o_p(n^{-1/2})
\]
where the second equality holds because the class of functions \( \{\varphi(x_{ti}, v_2^*) \{g(x_{ti}; \theta) - g(x_{ti}; \theta_0)\} : \theta \in \mathcal{N}_{02n}\} \) is a Donsker class and the last result holds by Assumption 6.10. Thus the result follows.

(c) By the definition of \( \hat{\varphi}(x_t, v_2^*) \), we obtain
\[
\frac{1}{n} \sum_{i=1}^{n} \int \varphi(x_{ti}, v_2^*) \{\hat{g}(x_{ti}; \theta_0) - \rho(y_{t,t-1}; \theta_0)\} f_{k_t, k_{t-1}|w_{ti}} dk_t dk_{t-1}
\]
\[
= \frac{1}{n} \sum_{i=1}^{n} \int \hat{\varphi}(x_{ti}, v_2^*) - \varphi(x_{ti}, v_2^*) \rho(y_{t,t-1}; \theta_0) f_{k_t, k_{t-1}|w_{ti}} dk_t dk_{t-1}.
\]
Thus the result follows by the Markov inequality since \( \hat{\varphi}(x_{ti}, v_2^*) - \varphi(x_{ti}, v_2^*) = o_p(1) \) uniformly over \( x_{ti} \in \mathcal{X}_t \) and \( g(x_{ti}; \theta_0) = 0 \), as in Ai and Chen (2003).

We now prove Theorem 6.12.

**Proof.** 1. The result follows from a similar argument to the proof of Theorem 3 in Hu and Schennach (2008). In the proof the bound for the pathwise second derivative is obtained by
\[
\sup_{\alpha \in \mathcal{N}_{02n}} \frac{d^2}{d\alpha^2} \left( \ln f_{r, k_t^*, k_{t-1}^*|z_t^*} (r_t, k_t^*, k_{t-1}^*|z_t^*; \alpha) \right)_{v=1n, (\alpha - \alpha_0)} \]
\[
\leq \sup_{\alpha \in \mathcal{N}_{02n}} \left[ \frac{f^{[1]}_{r, k_t^*, k_{t-1}^*|z_t^*} (r_t, k_t^*, k_{t-1}^*|z_t^*; \alpha, \nu)}{f_{r, k_t^*, k_{t-1}^*|z_t^*} (r_t, k_t^*, k_{t-1}^*|z_t^*; \alpha)} \right]^2 + \left[ \frac{f^{[2]}_{r, k_t^*, k_{t-1}^*|z_t^*} (\cdot; z_t^*; \alpha, \nu)}{f_{r, k_t^*, k_{t-1}^*|z_t^*} (\cdot; z_t^*; \alpha)} \right] \| \alpha - \alpha_0 \|_{s, \alpha}, \| v=1n \|_{s, \alpha}
\]
where \( f^{[2]}_{r, k_t^*, k_{t-1}^*|z_t^*} (r_t, k_t^*, k_{t-1}^*|z_t^*; \alpha, \nu) \) is defined as \( \frac{d^2}{d\nu^2} f_{r, k_t^*, k_{t-1}^*|z_t^*} (r_t, k_t^*, k_{t-1}^*|z_t^*; \alpha + \nu^{-1}) \) with the terms inside the integrals being replaced by the absolute values of \( \bar{f}_1, \bar{f}_2, \frac{d}{d\alpha} \bar{f}_r|z_t^* \), and \( \frac{d^2}{d\alpha^2} \bar{f}_r|z_t^* \).
similar to the definition of $f_{r_t, k_t^*, k_{t-1}^*|z_t^*}^1(r_t, k_t^*, k_{t-1}^*|z_t^*; \bar{\alpha}, \nu)$.

2. Let $\varepsilon_n = o(n^{-1/2}) > 0$ and $u_{2n}^* = \pm v_{2n}^*$. By a Taylor expansion around $\hat{\theta}$, we obtain that for $\theta_s \in (\hat{\theta}, \hat{\theta} + \varepsilon_n u_{2n}^*)$,

$$\frac{d\hat{Q}_n(\hat{\theta}) - \hat{Q}_n(\hat{\theta} + \varepsilon_n u_{2n}^*)}{d\theta} = \frac{d\hat{Q}_n(\hat{\theta})}{d\theta} [\varepsilon_n u_{2n}^*] + \frac{d^2\hat{Q}_n(\hat{\theta})}{d\theta d\theta} [\varepsilon_n u_{2n}^*, \varepsilon_n u_{2n}^*] + \frac{d^2\hat{Q}_n(\hat{\theta})}{d\theta d\theta} [\varepsilon_n u_{2n}^*, \varepsilon_n u_{2n}^*].$$

Then by Lemma C.3 and Lemma C.4, we have

$$\sqrt{n} \langle v_{2n}^*, \hat{\theta} - \theta_0 \rangle_\theta = -\frac{1}{\sqrt{n}} \sum_{i=1}^{n} \int \left\{ \frac{dg(x_{ti}; \theta_0)}{d\theta} [v_{2n}^*] \right\}' \cdot A(x_{ti}) \rho(y_{t,t-1,i}; \theta_0) f_{k_t,k_{t-1}|w_{ti}dk_tdk_{t-1}} + o_p(1).$$

Thus the result follows by the Lindeberg-Lévy central limit theorem.

References


