



























































































































$$\lambda^{-1/2} \begin{bmatrix} \text{vec } \underline{m}_{x\varepsilon} \\ \text{vech}(\underline{m}_{\varepsilon\varepsilon} - \underline{\Sigma}_{\varepsilon\varepsilon}) \end{bmatrix} \xrightarrow{L} N \left[ \begin{bmatrix} \underline{0} \\ \underline{\lambda}' \end{bmatrix}, \begin{bmatrix} \underline{\Sigma}_{\varepsilon\varepsilon} & \underline{\bar{m}}_{xx} & \underline{0} \\ \underline{0} & \underline{\Gamma}_{\varepsilon\varepsilon} \end{bmatrix} \right]$$

Thus, the result follows from the multivariate central limit theorem. □









































































$$= \tilde{S}_{\varepsilon\varepsilon}^{1/2} \hat{Q}_{(1)} (\Lambda_1^* - I) \hat{Q}_{(1)}' \tilde{S}_{\varepsilon\varepsilon}^{1/2} , \quad (3.55)$$

and

$$\begin{aligned} \tilde{\Sigma}_{\varepsilon\varepsilon} &= T^{-1} \hat{\Sigma}_{T\varepsilon\varepsilon} T^{-1} \\ &= \tilde{S}_{\varepsilon\varepsilon} + (n+d)^{-1} n \tilde{S}_{\varepsilon\varepsilon}^{1/2} \hat{Q}_{(2)} (\Lambda_2^* - I) \hat{Q}_{(2)}' \tilde{S}_{\varepsilon\varepsilon}^{1/2} \\ &= \tilde{S}_{\varepsilon\varepsilon} + (n+d)^{-1} n [n^{-1}(n-1) \tilde{\Sigma}_{ZZ} - \tilde{S}_{\varepsilon\varepsilon} \\ &\quad - \tilde{S}_{\varepsilon\varepsilon}^{1/2} \hat{Q}_{(1)} (\Lambda_1^* - I) \hat{Q}_{(1)}' \tilde{S}_{\varepsilon\varepsilon}^{1/2}] \\ &= (n+d)^{-1} d \tilde{S}_{\varepsilon\varepsilon} + (n+d)^{-1} n [n^{-1}(n-1) \tilde{\Sigma}_{ZZ} - \hat{\Sigma}_{ZZ}] . \end{aligned}$$

Also, by (3.55),

$$\hat{\beta} = (\hat{P}_{kk}')^{-1} \hat{P}_{rk}' = - \hat{T}_{kr} \hat{T}_{rr}^{-1} ,$$

$$\tilde{\Sigma}_{xx} = \hat{P}_{kk} (\Lambda_1^* - I) \hat{P}_{kk}' . \quad \square$$

Under the assumptions of Theorem 3.2, we can perform a goodness of fit test. The alternative to the model (3.1) is the unrestricted normal model where  $Z_t$  are independently and identically distributed with

unrestricted mean and covariance matrix and  $\hat{\Sigma}_{\epsilon\epsilon}$  is a multiple of Wishart matrix independent of  $\hat{Z}_t$ .

Theorem 3.3. Let the model (3.1) hold, and let Assumptions 3.2 and 3.4 through 3.8 hold. Then, the likelihood ratio statistic of the model (3.1) against the unrestricted model is

$$\chi^2 = (n+d) \sum_{i=k+1}^p \log\{(n+d)^{-1}[(n-1)\hat{\lambda}_i + d]\} - n \sum_{i=k+1}^p \log\{n^{-1}(n-1)\hat{\lambda}_i\}.$$

If the model specification is correct, then the test statistic is asymptotically distributed as a chi-square random variable with  $2^{-1}r(r+1)$  degrees of freedom.

Proof. Using the notation in the proof of Theorem 3.2, we have for the unrestricted model

$$\hat{\Sigma}_{TZZ} = \hat{\Lambda},$$

$$\hat{\Sigma}_{T\epsilon\epsilon} = I.$$

The expression for the test statistic follows by evaluating the function  $f(\theta)$  in (3.38) and taking the difference. The distribution of observations is a product of a normal distribution and a Wishart distribution. Also, the mean and the covariance matrices are twice differentiable functions of the parameters. Thus, by the standard likelihood ratio theory, the test statistic converges in law to a chi-square random variable. The degrees of freedom are given by the

difference of the number of parameters for the two models for  $\Sigma_{ZZ}$ . The model (3.1) has  $kr + 2^{-1} k(k+1)$  parameters in  $\Sigma_{ZZ}$ , and the unrestricted model has  $2^{-1} p(p+1)$  parameters. Therefore, the difference is  $2^{-1} r(r+1)$ .  $\square$

### 3. Maximum likelihood estimators adjusted for degrees of freedom

The maximum likelihood estimator does not take into account the number of parameters to be estimated. As in regression analysis, we often use an estimator which is obtained from the maximum likelihood estimator by making an adjustment for degrees of freedom. Usually, such an adjustment does not change the asymptotic properties of an estimator. However, more care is required when there exist the incidental parameters appearing in the distributions of finitely many observations. Consider the maximum likelihood estimator  $\tilde{\Sigma}_{\epsilon\epsilon}$  in (3.10) for the functional relationship model. The parameter  $\Sigma_{\epsilon\epsilon}$  is a structural parameter which appears in the distributions of infinitely many observations as  $n$  and  $d$  increase. But, the presence of the incidental parameters  $\tilde{x}_t$  for the functional model makes the maximum likelihood estimator  $\tilde{\Sigma}_{\epsilon\epsilon}$  inconsistent. In the expression (3.10) for the  $\tilde{\Sigma}_{\epsilon\epsilon}$ , the term

$$\hat{S}_{\epsilon v} \hat{S}_{vv}^{-1} \hat{m}_{vv} \hat{S}_{vv}^{-1} \hat{S}_{ve}$$

is estimating  $\Sigma_{\epsilon v} \Sigma_{vv}^{-1} \Sigma_{ve}$ , where

$$\tilde{\Sigma}_{\epsilon v} = \tilde{\Sigma}'_{v\epsilon} = \tilde{\Sigma}_{\epsilon\epsilon} (\tilde{I}, -\beta')'$$

$$\tilde{\Sigma}_{vv} = (\tilde{I}, -\beta') \tilde{\Sigma}_{\epsilon\epsilon} (\tilde{I}, -\beta')' .$$

Hence, even when both  $n$  and  $d$  tend to infinity, we do not expect the  $\tilde{\Sigma}_{\epsilon\epsilon}$  for the functional case to be a consistent estimator of  $\Sigma_{\epsilon\epsilon}$  .

However, it follows from (3.29), (3.30), and (3.31) that using the transformation used in the proof of Theorem 3.1 we can concentrate the inconsistency of  $\tilde{\Sigma}_{\epsilon\epsilon}$  in  $\tilde{\Sigma}_{ww.v}$  and obtain the consistent maximum likelihood estimators  $\tilde{\Sigma}_{vv}$  and  $\hat{\gamma}$  of  $\Sigma_{vv}$  and  $\Sigma_{wv} \Sigma_{vv}^{-1}$  .

Furthermore, the inconsistency of  $\tilde{\Sigma}_{ww.v}$  is due only to a proportion factor. Thus, making an obvious adjustment, we consider  $d^{-1}(d+n)\tilde{\Sigma}_{ww.v}$  as an estimator of  $\Sigma_{ww.v}$  . Therefore, transforming to the original parameterization  $\Sigma_{\epsilon\epsilon}$  , we obtain the maximum likelihood estimator of  $\Sigma_{\epsilon\epsilon}$  adjusted for the presence of the incidental parameters as follows. By (3.34),

$$\ddot{\Sigma}_{\epsilon\epsilon} = \hat{\tilde{\Gamma}}^{-1} \begin{bmatrix} \tilde{\Sigma}_{vv} & \tilde{\Sigma}_{vw} \\ \tilde{\Sigma}_{wv} & d^{-1}(d+n)\tilde{\Sigma}_{ww.v} + \tilde{\Sigma}_{wv}\tilde{\Sigma}_{vv}^{-1}\tilde{\Sigma}_{vw} \end{bmatrix} \hat{\tilde{\Gamma}}^{-1}$$

$$\begin{aligned}
&= \tilde{\Sigma}_{\varepsilon\varepsilon} + \hat{\Gamma}^{-1} (d+n)^{-1} n \begin{bmatrix} 0 & 0 \\ 0 & \hat{\Sigma}_{ww} - \hat{\Sigma}_{wv} \hat{\Sigma}_{vv}^{-1} \hat{\Sigma}_{vw} \end{bmatrix} \hat{\Gamma}^{-1} \\
&= \tilde{\Sigma}_{\varepsilon\varepsilon} + (d+n)^{-1} n (\hat{\beta}, \mathbf{I})' (\mathbf{I} + \hat{\beta} \hat{\beta}')^{-1/2} (\hat{\Sigma}_{ww} - \hat{\Sigma}_{wv} \hat{\Sigma}_{vv}^{-1} \hat{\Sigma}_{vw}) (\mathbf{I} + \hat{\beta} \hat{\beta}')^{-1/2} (\hat{\beta}, \mathbf{I}) \\
&= (d+n)^{-1} [d \tilde{\Sigma}_{\varepsilon\varepsilon} + (n-1) \hat{\Sigma}_{\varepsilon v} \hat{\Sigma}_{vv}^{-1} \hat{\Sigma}_{vv} \hat{\Sigma}_{v\varepsilon} \\
&\quad + n (\hat{\beta}, \mathbf{I})' (\mathbf{I} + \hat{\beta} \hat{\beta}')^{-1} (\hat{\beta}, \mathbf{I}) (\tilde{\Sigma}_{\varepsilon\varepsilon} - \hat{\Sigma}_{\varepsilon v} \hat{\Sigma}_{vv}^{-1} \hat{\Sigma}_{v\varepsilon}) (\hat{\beta}, \mathbf{I})' (\mathbf{I} + \hat{\beta} \hat{\beta}')^{-1} (\hat{\beta}, \mathbf{I})] \\
&= (d+n)^{-1} [d \tilde{\Sigma}_{\varepsilon\varepsilon} + (n-1) \tilde{\Sigma}_{\varepsilon\varepsilon} \hat{T}_{(2)} \hat{\Lambda}_{(2)} \hat{T}_{(2)}' \tilde{\Sigma}_{\varepsilon\varepsilon} \\
&\quad + n \hat{P}_{(1)} (\hat{P}_{(1)}' \hat{P}_{(1)})^{-1} \hat{P}_{(1)}' (\tilde{\Sigma}_{\varepsilon\varepsilon} - \tilde{\Sigma}_{\varepsilon\varepsilon} \hat{T}_{(2)} \hat{\Lambda}_{(2)} \hat{T}_{(2)}' \tilde{\Sigma}_{\varepsilon\varepsilon}) \hat{P}_{(1)} (\hat{P}_{(1)}' \hat{P}_{(1)})^{-1} \hat{P}_{(1)}] \\
&= (d+n)^{-1} [d \tilde{\Sigma}_{\varepsilon\varepsilon} + (n-1) \tilde{\Sigma}_{\varepsilon\varepsilon}^{1/2} \hat{Q}_{(2)} \hat{\Lambda}_{(2)} \hat{Q}_{(2)}' \tilde{\Sigma}_{\varepsilon\varepsilon}^{1/2} + n \tilde{\Sigma}_{\varepsilon\varepsilon}^{1/2} \hat{Q}_{(1)} \hat{Q}_{(1)}' \tilde{\Sigma}_{\varepsilon\varepsilon}^{1/2}] \\
&= (d+n)^{-1} [d \tilde{\Sigma}_{\varepsilon\varepsilon} + n \{n^{-1}(n-1) \hat{m}_{zz} - \tilde{\Sigma}_{zz}\}] , \tag{3.56}
\end{aligned}$$

where

$$\tilde{\Sigma}_{zz} = \tilde{\Sigma}_{\varepsilon\varepsilon}^{1/2} \hat{Q}_{(1)} \{n^{-1}(n-1) \hat{\Lambda}_1 - \mathbf{I}\} \hat{Q}_{(1)}' \tilde{\Sigma}_{\varepsilon\varepsilon}^{1/2} .$$

Thus, changing a divisor for a portion of the estimator (3.10) for the functional model, we have obtained  $\ddot{\tilde{\Sigma}}_{\epsilon\epsilon}$  which is identical to  $\tilde{\Sigma}_{\epsilon\epsilon}$  in (3.37), the maximum likelihood estimator for the structural model. By adjusting for degrees of freedom, we obtain the same maximum likelihood estimators of the parameters  $\beta_0$ ,  $\beta$ , and  $\Sigma_{\epsilon\epsilon}$  for both the functional relationship and the structural relationship.

Since the model (3.1) has the intercept term, we have used a divisor  $(n-1)$  for  $\mathfrak{m}_{ZZ}$ , the matrix of sums of squares and cross products. This choice of divisor led to the presence of a factor  $n^{-1}(n-1)$  in the maximum likelihood estimators  $\tilde{\Sigma}_{xx}$  in (3.36),  $\tilde{\Sigma}_{\epsilon\epsilon}$  in (3.37), and  $\ddot{\tilde{\Sigma}}_{\epsilon\epsilon}$  in (3.56). By analogy to usual regression analysis, we adjust the estimators  $\tilde{\Sigma}_{xx}$  and  $\tilde{\Sigma}_{\epsilon\epsilon}$ , and define the maximum likelihood estimators of  $\Sigma_{xx}$  and  $\Sigma_{\epsilon\epsilon}$  adjusted for degrees of freedom as

$$\hat{\tilde{\Sigma}}_{xx} = \hat{\tilde{P}}_{kk} (\hat{\Lambda}_1 - I) \hat{\tilde{P}}'_{kk}, \quad (3.57)$$

$$\hat{\tilde{\Sigma}}_{\epsilon\epsilon} = (n-1+d)^{-1} [d \hat{S}_{\epsilon\epsilon} + (n-1) (\mathfrak{m}_{ZZ} - \hat{\tilde{\Sigma}}_{ZZ})], \quad (3.58)$$

where

$$\hat{\tilde{\Sigma}}_{ZZ} = \hat{\tilde{P}}_{(1)} (\hat{\Lambda}_1 - I) \hat{\tilde{P}}'_{(1)}.$$

In the following section, we consider the properties of the maximum likelihood estimators adjusted for degrees of freedom. The estimators

are  $\hat{\beta}_0$  in (3.4),  $\hat{\beta}$  in (3.7), and  $\hat{\Sigma}_{\epsilon\epsilon}$  in (3.58) for both the functional and structural models, and  $\hat{\Sigma}_{xx}$  in (3.57) for the structural model. Since we consider only the asymptotic properties, the adjustments used in (3.57) and (3.58) do not change the results.

### C. Asymptotic Properties

#### 1. Strong consistency

In this section, we show the strong consistency of the maximum likelihood estimators adjusted for degrees of freedom under a broad class of assumptions. We begin by presenting two lemmas which will be used to show the strong consistency.

Lemma 3.1. Let the model (3.1) hold, and let Assumptions 3.5 and 3.7 hold. In addition, let either (a) Assumption 3.1, or (b) Assumption 3.2 hold. Then, for both the cases (a) and (b), as  $n \rightarrow \infty$ ,

$$S_{\epsilon\epsilon}^{-1/2} m_{ZZ} S_{\epsilon\epsilon}^{-1/2} \longrightarrow \Sigma_{\epsilon\epsilon}^{-1/2} \Sigma_{ZZ} \Sigma_{\epsilon\epsilon}^{-1/2}, \text{ a.s. },$$

where

$$\Sigma_{ZZ} = (\beta, I)' \Sigma_{xx} (\beta, I) + \Sigma_{\epsilon\epsilon}.$$

Proof. For case (b), the result follows from Assumption 3.7 and the strong law of large numbers. For case (a), we observe that

$$m_{ZZ} = m_{zz} + m_{z\epsilon} + m_{\epsilon z} + m_{\epsilon\epsilon}, \quad (3.59)$$

where

$$\mathbf{m}_{zz} = (\beta, \mathbf{I})' \mathbf{m}_{xx} (\beta, \mathbf{I}) \quad ,$$

$$\mathbf{m}_{z\varepsilon} = (\beta, \mathbf{I})' \mathbf{m}_{x\varepsilon}$$

$$= \mathbf{m}'_{\varepsilon z} \quad ,$$

$$\mathbf{m}_{x\varepsilon} = (n-1)^{-1} \sum_{t=1}^n (\mathbf{x}_t - \bar{\mathbf{x}})' \boldsymbol{\varepsilon}_t' \quad .$$

By Assumption 3.1,

$$\lim_{n \rightarrow \infty} \mathbf{m}_{zz} = (\beta, \mathbf{I})' \Sigma_{xx} (\beta, \mathbf{I}) \quad . \quad (3.60)$$

By the strong law of large numbers,

$$\mathbf{m}_{\varepsilon\varepsilon} \longrightarrow \Sigma_{\varepsilon\varepsilon} \quad \text{a.s.} \quad . \quad (3.61)$$

The  $(i,j)$ -th element of  $\mathbf{m}_{z\varepsilon}$  is

$$(n-1)^{-1} \sum_{t=1}^n (z_{it} - \bar{z}_i) \varepsilon_{jt} = (n-1)^{-1} \sum_{t=1}^n z_{it}^* \varepsilon_{jt} \quad ,$$

where  $z_{it}$  is the  $i$ -th element of  $\mathbf{z}_t$  ,

$$\bar{z}_i = n^{-1} \sum_{t=1}^n z_{it} ,$$

and

$$z_{it}^* = z_{it} - \bar{z}_i .$$

It follows from Assumption 3.1 that there exist constants  $K_i$  satisfying

$$\lim_{n \rightarrow \infty} (n-1)^{-1} \sum_{t=1}^n z_{it}^{*2} = K_i < \infty .$$

Hence, there exists a finite number  $K_0$  such that for any  $n$

$$(n-1)^{-1} \sum_{t=1}^n z_{it}^{*2} < K_0 . \quad (3.62)$$

By Abel's partial summation formula, for any two series  $\{a_t\}$  and  $\{b_t\}$  and any  $n > M > 0$  ,

$$\sum_{t=M}^n a_t (b_{t+1} - b_t) = a_n b_{n+1} - a_M b_M - \sum_{t=M+1}^n b_t (a_t - a_{t-1}) . \quad (3.63)$$

Letting

$$a_t = t^{-2} ,$$

$$b_t = \sum_{j=1}^{t-1} z_{ij}^{*2} ,$$

$$M = 2 ,$$

and using (3.62) we have for any  $n > 2$

$$\sum_{t=2}^n \frac{1}{t^2} z_{it}^{*2} = \frac{1}{n^2} \sum_{j=1}^n z_{ij}^{*2} - \frac{1}{4} z_{i1}^{*2} \sum_{t=3}^n \left[ \sum_{j=1}^{t-1} z_{ij}^{*2} \left( \frac{1}{t^2} - \frac{1}{(t-1)^2} \right) \right]$$

$$= \frac{1}{n} \left( \frac{1}{n} \sum_{j=1}^n z_{ij}^{*2} \right) - \frac{1}{4} z_{i1}^{*2} + \sum_{t=3}^n \frac{2t-1}{t^2(t-1)} \sum_{j=1}^{t-1} z_{ij}^{*2}$$

$$< \frac{1}{n} K_0 + 2 \sum_{t=3}^n \frac{1}{t(t-1)} \frac{1}{(t-1)} \sum_{j=1}^{t-1} z_{ij}^{*2}$$

$$\begin{aligned}
&< \frac{1}{n} K_0 + 2 \sum_{t=3}^n \frac{1}{t(t-1)} K_0 = K_0 \left(1 - \frac{1}{n}\right) \\
&< K_0 .
\end{aligned} \tag{3.64}$$

Since  $\sum_{t=1}^n t^{-2} z_{it}^{*2}$  is a series of nonnegative terms, (3.64) implies that for some  $V_0$

$$\lim_{n \rightarrow \infty} \sum_{t=1}^n \frac{1}{t^2} z_{it}^{*2} = V_0 < \infty .$$

Thus, as  $n \rightarrow \infty$ ,

$$\begin{aligned}
\sum_{t=1}^n \frac{1}{t^2} V\{z_{it}^{*2} \varepsilon_{jt}\} &= \sum_{t=1}^n \frac{1}{t^2} z_{it}^{*2} \sigma_{\varepsilon \varepsilon jj} \\
&\longrightarrow V_0 \sigma_{\varepsilon \varepsilon jj} < \infty .
\end{aligned} \tag{3.65}$$

By (3.65), the independence of  $z_{it}^* \varepsilon_{jt}$ , and the fact that

$$E\{z_{it}^* \varepsilon_{jt}\} = 0 ,$$

it follows from the standard probability theorem (See, for example, Chung (1974), p. 125.) that

$$(n-1)^{-1} \sum_{t=1}^n z_{it}^* \epsilon_{jt} \longrightarrow 0 \text{ a.s. .}$$

Therefore,

$$\begin{aligned} \underline{m}_{ZE} &= \underline{m}_{EZ}' \\ &\longrightarrow 0 \text{ a.s. .} \end{aligned} \quad (3.66)$$

By (3.59), (3.60), (3.61), and (3.66),

$$\underline{m}_{ZZ} \longrightarrow (\beta, \underline{I})' \underline{\Sigma}_{XX}(\beta, \underline{I}) + \underline{\Sigma}_{EE} \text{ a.s. .} \quad (3.67)$$

Also, by Assumption 3.7, as  $n \rightarrow \infty$ ,

$$\underline{S}_{EE} \longrightarrow \underline{\Sigma}_{EE} \text{ a.s. .} \quad (3.68)$$

By (3.67) and (3.68), the result for the case (a) follows.  $\square$

We introduce further notations. Let

$$\underline{\Sigma}_{pp} = [(\beta, \underline{I}) \underline{\Sigma}_{EE}^{-1} (\beta, \underline{I})']^{-1} .$$

Also, let  $v_1 > v_2 > \dots > v_k$  be the eigenvalues of  $\Sigma_{\rho\rho}^{-1/2} \Sigma_{xx} \Sigma_{\rho\rho}^{-1/2}$ ,  
and let  $\tilde{R}$  be the matrix of corresponding orthonormal eigenvectors.

Note that by Theorem 2.1

$$\Sigma_{\rho\rho} = \Sigma_{uu} - \Sigma_{uv} \Sigma_{vv}^{-1} \Sigma_{vu} \quad , \quad (3.69)$$

where

$$\Sigma_{vv} = (\tilde{I}, -\beta') \Sigma_{\varepsilon\varepsilon} (\tilde{I}, -\beta')' \quad ,$$

$$\Sigma_{uv} = \Sigma'_{vu}$$

$$= \Sigma_{ue} - \Sigma_{ur} \beta \quad .$$

Note also that  $v_k > 0$  since  $\Sigma_{\rho\rho}^{-1/2} \Sigma_{xx} \Sigma_{\rho\rho}^{-1/2}$  is positive definite.

Lemma 3.2. Let  $\lambda_1 > \lambda_2 > \dots > \lambda_p$  be the eigenvalues of the matrix

$\Sigma_{\varepsilon\varepsilon}^{-1/2} \Sigma_{ZZ} \Sigma_{\varepsilon\varepsilon}^{-1/2}$ , and denote the corresponding orthonormal eigenvectors by

$$Q = (Q_{(1)}, Q_{(2)}) \quad ,$$

where  $Q_{(1)}$  has  $k$  columns. Then,

$$\begin{aligned}\lambda_i &= 1 + v_i, \quad i = 1, 2, \dots, k, \\ &= 1, \quad i = k+1, k+2, \dots, p,\end{aligned}\tag{3.70}$$

$$\begin{aligned}Q_{(1)} &= \Sigma_{\epsilon\epsilon}^{-1/2} (\beta, \underline{I})' \Sigma_{\rho\rho}^{1/2} \underline{R}, \\ Q_{(2)} &= \Sigma_{\epsilon\epsilon}^{1/2} (\underline{I}, -\beta')' \Sigma_{vv}^{-1/2}.\end{aligned}\tag{3.71}$$

Proof. We observe that the columns of  $Q$  given by (3.71) are orthonormal. Also,

$$\begin{aligned}\Sigma_{xx} \Sigma_{\rho\rho}^{-1/2} \underline{R} &= \Sigma_{xx} \Sigma_{\rho\rho}^{-1} \Sigma_{\rho\rho}^{1/2} \underline{R} \\ &= \Sigma_{\rho\rho}^{1/2} \underline{D}_v, \end{aligned}$$

where

$$\underline{D}_v = \text{diag}\{v_1, v_2, \dots, v_k\}.$$

Thus,

$$\begin{aligned}\Sigma_{\epsilon\epsilon}^{-1/2} \Sigma_{ZZ} \Sigma_{\epsilon\epsilon}^{-1/2} Q_{(1)} &= \Sigma_{\epsilon\epsilon}^{-1/2} (\beta, \underline{I})' \Sigma_{xx} \Sigma_{\rho\rho}^{-1/2} \underline{R} + Q_{(1)} \\ &= Q_{(1)} (\underline{D}_v + \underline{I}).\end{aligned}\tag{3.72}$$

We also observe that

$$\begin{aligned} \Sigma_{\epsilon\epsilon}^{-1/2} \Sigma_{ZZ} \Sigma_{\epsilon\epsilon}^{-1/2} Q_{(2)} &= \Sigma_{\epsilon\epsilon}^{-1/2} (\beta, I)' \Sigma_{xx} (\beta, I) (I, -\beta')' + Q_{(2)} \\ &= Q_{(2)} \quad . \end{aligned} \quad (3.73)$$

The results follow from (3.72) and (3.73).  $\square$

In the following theorem, we prove the strong consistency of the maximum likelihood estimators adjusted for degrees of freedom. In the proof of Theorem 3.4, let  $\omega$  be a point in the probability space of all sequences of observations, and let the superscript  $(n)$  indicate that the quantity is calculated from the first  $n$  elements of  $\omega$ . This notation is discontinued once Theorem 3.4 is proven. Let

$$\Sigma_{ZZ} = (\beta, I)' \Sigma_{xx} (\beta, I) \quad ,$$

$$\hat{D}_{kk} = \hat{P}_{kk} \hat{\Lambda}_1 \hat{P}'_{kk} \quad ,$$

$$\hat{D}_{rr} = \hat{T}_{rr} \hat{\Lambda}_2 \hat{T}'_{rr} \quad .$$

Theorem 3.4. Let the model (3.1) hold, and let Assumptions 3.5 and 3.7 hold. In addition, let either (a) Assumption 3.1, or (b) Assumption 3.2 hold. Let the estimators  $\hat{\beta}_0^{(n)}$ ,  $\hat{\beta}^{(n)}$ ,  $\hat{\Sigma}_{\epsilon\epsilon}^{(n)}$ , and  $\hat{\Sigma}_{xx}^{(n)}$  be defined by (3.4), (3.7), (3.58), and (3.57), respectively. Then, for both the cases (a) and (b), as  $n \rightarrow \infty$ ,

$$\hat{\beta}_o^{(n)} \longrightarrow \beta_o, \text{ a.s. },$$

$$\hat{\beta}^{(n)} \longrightarrow \beta, \text{ a.s. },$$

$$\hat{\Sigma}_{\epsilon\epsilon}^{(n)} \longrightarrow \Sigma_{\epsilon\epsilon}, \text{ a.s. },$$

$$\hat{\Sigma}_{xx}^{(n)} \longrightarrow \Sigma_{xx}, \text{ a.s. },$$

$$\hat{\Sigma}_{kk}^{(n)} \longrightarrow \Sigma_{\eta\eta}, \text{ a.s. },$$

$$\hat{\Sigma}_{rr}^{(n)} \longrightarrow \Sigma_{vv}^{-1}, \text{ a.s. },$$

$$\hat{\Sigma}_{rr}^{(n)} \hat{\Sigma}_{rr}^{(n)'} \longrightarrow \Sigma_{vv}^{-1}, \text{ a.s. },$$

where

$$\begin{aligned} \Sigma_{\eta\eta} &= \Sigma_{xx} + \Sigma_{\rho\rho} \\ &= \Sigma_{xx} - \Sigma_{uv} \Sigma_{vv}^{-1} \Sigma_{vu}. \end{aligned}$$

Proof. Since the eigenvectors  $\lambda_i^{(n)}$  are locally continuous functions of the elements in  $S_{\epsilon\epsilon}^{-1/2} \mathbb{M}_{ZZ} S_{\epsilon\epsilon}^{-1/2}$ , it follows from Lemma 3.1 that for  $i = 1, 2, \dots, p$ ,

$$\hat{\lambda}_i^{(n)} \longrightarrow \lambda_i, \text{ a.s. } .$$

Fix  $\omega$  such that

$$\hat{\lambda}_i^{(n)}(\omega) \longrightarrow \lambda_i, \quad (3.74)$$

$$m_{ZZ}^{(n)}(\omega) \longrightarrow \Sigma_{ZZ}, \quad (3.75)$$

$$S_{\epsilon\epsilon}^{(n)}(\omega) \longrightarrow \Sigma_{\epsilon\epsilon}, \text{ as } n \rightarrow \infty. \quad (3.76)$$

The set of  $\omega$  satisfying (3.74), (3.75), and (3.76) has probability one. Since  $\hat{Q}^{(n)}(\omega)$  is orthogonal for all  $n$ , each element of  $\hat{Q}^{(n)}(\omega)$  is bounded. Thus, for every subsequence  $\{\hat{Q}^{(n_i)}(\omega)\}$ , there exists a convergent sub-subsequence  $\{\hat{Q}^{(n_{ij})}(\omega)\}$ . The limit of such a convergent sub-subsequence depends on the sub-subsequence, and we denote the limit by

$$\tilde{H}_{ij}(\omega) = (\tilde{H}_{ij}^{(1)}(\omega), \tilde{H}_{ij}^{(2)}(\omega)) .$$

Since each  $\hat{Q}^{(n_{ij})}(\omega)$  is orthogonal, so is  $\tilde{H}_{ij}(\omega)$ . For all  $j$ ,

$$S_{\epsilon\epsilon}^{(n_{ij}) - 1/2}(\omega) m_{ZZ}^{(n_{ij})}(\omega) S_{\epsilon\epsilon}^{(n_{ij}) - 1/2}(\omega) \hat{Q}^{(n_{ij})}(\omega)$$

$$= \hat{Q}^{(n_{ij})}(\omega) \text{diag}\{\hat{\lambda}_1^{(n_{ij})}(\omega), \dots, \hat{\lambda}_p^{(n_{ij})}(\omega)\} . \quad (3.77)$$

Taking limits over  $j$  on both sides of the equality in (3.77), and using (3.74), (3.75), and (3.76), we obtain

$$\Sigma_{\varepsilon\varepsilon}^{-1/2} \Sigma_{ZZ} \Sigma_{\varepsilon\varepsilon}^{-1/2} \tilde{H}_{ij}(\omega) = \tilde{H}_{ij}(\omega) \text{diag}\{\lambda_1, \dots, \lambda_p\} .$$

where  $\lambda_1$  are defined in (3.70). Hence, the columns of  $\tilde{H}_{ij}(\omega)$  are orthogonal eigenvectors of  $\Sigma_{\varepsilon\varepsilon}^{-1/2} \Sigma_{ZZ} \Sigma_{\varepsilon\varepsilon}^{-1/2}$ . Since by Lemma 3.2

$$\lambda_k = 1 + v_k > 1 = \lambda_i, \quad i = k+1, \dots, p,$$

the two eigensubspaces of  $\Sigma_{\varepsilon\varepsilon}^{-1/2} \Sigma_{ZZ} \Sigma_{\varepsilon\varepsilon}^{-1/2}$  corresponding to the first  $k$  eigenvalues and the last  $r$  eigenvalues, respectively, are unique. Thus, there exist orthogonal matrices  $G_{ij}^{(1)}$ ;  $k \times k$ , and  $G_{ij}^{(2)}$ ;  $r \times r$ , such that

$$\begin{aligned} \tilde{H}_{ij}^{(1)}(\omega) &= Q_{(1)} G_{ij}^{(1)} \\ &= Q_{(1)} \text{block diag}\{G_{ij}^{(11)}, G_{ij}^{(12)}, \dots, G_{ij}^{(1s)}\}, \end{aligned}$$

and

$$\tilde{H}_{ij}^{(2)}(\omega) = Q_{(2)} \tilde{G}_{ij}^{(2)},$$

where  $s$  is the number of distinct roots  $v_u$ ,  $\tilde{G}_{ij}^{(1u)}$  is orthogonal with dimension corresponding to the multiplicity of  $v_u$ , and  $Q_{(1)}$  and  $Q_{(2)}$  are defined by (3.71). Therefore, by (3.70) and (3.71), as  $j \rightarrow \infty$ ,

$$\begin{aligned} \hat{Q}_{(1)}^{(n_{ij})}(\omega) [\hat{Q}_{(1)}^{(n_{ij})}(\omega)]' &\longrightarrow \tilde{H}_{ij}^{(1)}(\omega) (\tilde{H}_{ij}^{(1)}(\omega))', \\ &= Q_{(1)} \tilde{G}_{ij}^{(1)} \tilde{G}_{ij}^{(1)'} Q_{(1)}' \\ &= Q_{(1)} Q_{(1)}' \\ &= \tilde{\Sigma}_{\epsilon\epsilon}^{-1/2} (\beta, I)' \tilde{\Sigma}_{\rho\rho}^{1/2} R R' \tilde{\Sigma}_{\rho\rho}^{1/2} (\beta, I) \tilde{\Sigma}_{\epsilon\epsilon}^{-1/2} \\ &= \tilde{\Sigma}_{\epsilon\epsilon}^{-1/2} (\beta, I)' \tilde{\Sigma}_{\rho\rho} (\beta, I) \tilde{\Sigma}_{\epsilon\epsilon}^{-1/2}, \end{aligned} \quad (3.78)$$

$$\begin{aligned} \hat{Q}_{(2)}^{(n_{ij})}(\omega) [\hat{Q}_{(2)}^{(n_{ij})}(\omega)]' &\longrightarrow Q_{(2)} Q_{(2)}' \\ &= \tilde{\Sigma}_{\epsilon\epsilon}^{1/2} (I, -\beta')' \tilde{\Sigma}_{\nu\nu}^{-1} (I, -\beta') \tilde{\Sigma}_{\epsilon\epsilon}^{1/2}, \end{aligned} \quad (3.79)$$

$$\hat{Q}_{(1)}^{(n_{ij})}(\omega) \hat{\Lambda}_1^{(n_{ij})}(\omega) [\hat{Q}_{(1)}^{(n_{ij})}(\omega)]'$$

$$\begin{aligned}
& \longrightarrow Q_{(1)} \text{block diag}\{G_{ij}^{(1)}, \dots, G_{ij}^{(1s)}\} \text{block diag}\{\lambda_1 I, \dots, \lambda_s I\} \\
& \quad \text{block diag}\{G_{ij}^{(11)'}, \dots, G_{ij}^{(1s)'}\} Q'_{(1)} \\
& = Q_{(1)} \text{diag}\{\lambda_1, \dots, \lambda_k\} Q'_{(1)} \\
& = \Sigma_{\varepsilon\varepsilon}^{-1/2} (\beta, I)' \Sigma_{\rho\rho}^{1/2} R (I + D_v) R' \Sigma_{\rho\rho}^{1/2} (\beta, I) \Sigma_{\varepsilon\varepsilon}^{-1/2} \\
& = \Sigma_{\varepsilon\varepsilon}^{-1/2} (\beta, I)' \Sigma_{\rho\rho} (\beta, I) \Sigma_{\varepsilon\varepsilon}^{-1/2} + \Sigma_{\varepsilon\varepsilon}^{-1/2} \Sigma_{zz} \Sigma_{\varepsilon\varepsilon}^{-1/2}, \quad (3.80)
\end{aligned}$$

$$\begin{aligned}
& \hat{Q}_{(2)}^{(n_{ij})}(\omega) \hat{A}_2^{(n_{ij})}(\omega) [\hat{Q}_{(2)}^{(n_{ij})}(\omega)]' \longrightarrow Q_{(2)} G_{ij}^{(2)} I G_{ij}^{(2)'} Q'_{(2)} \\
& = Q_{(2)} Q'_{(2)} \\
& = \Sigma_{\varepsilon\varepsilon}^{1/2} (I, -\beta)' \Sigma_{vv}^{-1} (I, -\beta') \Sigma_{\varepsilon\varepsilon}^{1/2}. \quad (3.81)
\end{aligned}$$

Thus, by (3.68) and (3.80), as  $j \rightarrow \infty$ ,

$$\hat{D}_{kk}^{(n_{ij})}(\omega) \longrightarrow \Sigma_{\eta\eta}. \quad (3.82)$$

By (3.68) and (3.81), as  $j \rightarrow \infty$ ,

$$\hat{\tilde{D}}_{rr}^{(n_{ij})}(\omega) \longrightarrow \Sigma_{vv}^{-1} . \quad (3.83)$$

By (3.68) and (3.79),

$$\hat{\tilde{T}}_{rr}^{(n_{ij})}(\omega) \hat{\tilde{T}}_{rr}^{(n_{ij})'}(\omega) \longrightarrow \Sigma_{vv}^{-1} . \quad (3.84)$$

Also, by (3.68), (3.78), and (3.80), as  $j \rightarrow \infty$ ,

$$\hat{\tilde{\Sigma}}_{zz}^{(n_{ij})}(\omega) \longrightarrow \Sigma_{zz} , \quad (3.85)$$

and thus

$$\hat{\tilde{\beta}}^{(n_{ij})}(\omega) \longrightarrow \beta , \quad (3.86)$$

$$\hat{\tilde{\Sigma}}_{xx}^{(n_{ij})}(\omega) \longrightarrow \Sigma_{xx} . \quad (3.87)$$

By (3.67), (3.68), (3.85), and Assumption 3.7, as  $j \rightarrow \infty$ ,

$$\hat{\tilde{\Sigma}}_{\varepsilon\varepsilon}^{(n_{ij})}(\omega) \longrightarrow \Sigma_{\varepsilon\varepsilon} . \quad (3.88)$$

If every subsequence has a sub-subsequence converging to the same limit, then the sequence converges to the limit for such an  $\omega$ . Since the set of such  $\omega$  has probability one, the results for  $\hat{\tilde{D}}_{kk}^{(n)}$ ,  $\hat{\tilde{D}}_{rr}^{(n)}$ ,  $\hat{\tilde{T}}_{rr}^{(n)}$ ,  $\hat{\tilde{T}}_{rr}^{(n)'}(\omega)$ ,  $\hat{\tilde{\beta}}^{(n)}$ ,  $\hat{\tilde{\Sigma}}_{xx}^{(n)}$ , and  $\hat{\tilde{\Sigma}}_{\varepsilon\varepsilon}^{(n)}$  follows from (3.82), (3.83),

(3.84), (3.86), (3.87), and (3.88). Finally, for either the case (a) or (b), as  $n \rightarrow \infty$ ,

$$\bar{\tilde{x}} \longrightarrow \mu_x, \text{ a.s.,}$$

$$\bar{\tilde{y}} \longrightarrow \beta_0 + \mu_x \beta, \text{ a.s.,}$$

and thus the result for  $\hat{\beta}_0^{(n)}$  follows from that for  $\hat{\beta}^{(n)}$ .  $\square$

## 2. Limiting distribution

The asymptotic normality of the maximum likelihood estimators adjusted for degrees of freedom holds without the normality assumption on  $\tilde{\varepsilon}_t$  and  $\tilde{x}_t$ . But, to write down an explicit form of the limiting covariance matrix, we assume that  $\tilde{\varepsilon}_t$  are normal and that  $\tilde{\Sigma}_{\varepsilon\varepsilon}$  is a multiple of a Wishart matrix. The following theorem shows that inferences about  $\beta_0$ ,  $\beta$ , and  $\tilde{\Sigma}_{\varepsilon\varepsilon}$  based on the limiting results are the same for a wide range of assumptions on  $\tilde{x}_t$ .

Theorem 3.5. Let the model (3.1) hold, and let Assumptions 3.5 through 3.8 hold.

(a) If either Assumption 3.1 or Assumption 3.2 holds, then

$$\frac{1}{n^{1/2}} \begin{bmatrix} (\hat{\beta}_0 - \beta_0)' \\ \text{vec}(\hat{\beta} - \beta) \\ \text{vech}(\hat{\tilde{\Sigma}}_{\varepsilon\varepsilon} - \tilde{\Sigma}_{\varepsilon\varepsilon}) \end{bmatrix} \xrightarrow{L} N \left[ \begin{bmatrix} 0 \\ 0 \\ 0 \end{bmatrix}, \begin{bmatrix} \tilde{V}_{00} & \tilde{V}_{0\beta} & \tilde{V}_{0\varepsilon} \\ \tilde{V}'_{0\beta} & \tilde{V}_{\beta\beta} & \tilde{V}_{\beta\varepsilon} \\ \tilde{V}'_{0\varepsilon} & \tilde{V}'_{\beta\varepsilon} & \tilde{V}_{\varepsilon\varepsilon} \end{bmatrix} \right],$$

where

$$\tilde{V}_{oo} = \tilde{\Sigma}_{vv} + (\tilde{I}_{rxr} \boxtimes \mu_x) \tilde{V}_{\beta\beta} (\tilde{I}_{rxr} \boxtimes \mu'_x) ,$$

$$\tilde{V}_{o\beta} = - (\tilde{I}_{rxr} \boxtimes \mu_x) \tilde{V}_{\beta\beta} ,$$

$$\tilde{V}_{o\varepsilon} = - (\tilde{I}_{rxr} \boxtimes \mu_x) \tilde{V}_{\beta\varepsilon} ,$$

$$\tilde{V}_{\beta\beta} = \tilde{\Sigma}_{vv} \boxtimes \{ \tilde{\Sigma}_{xx}^{-1} [\tilde{\Sigma}_{xx} + (1+c) \tilde{\Sigma}_{\rho\rho}] \tilde{\Sigma}_{xx}^{-1} \} ,$$

$$\tilde{V}_{\beta\varepsilon} = -2c \{ \tilde{\Sigma}_{ve} \boxtimes [\tilde{\Sigma}_{xx}^{-1} \tilde{\Sigma}_{\rho\rho} (\beta, \tilde{I})] \} \psi'_p ,$$

$$\tilde{V}_{\varepsilon\varepsilon} = 2c \psi_p [ (\tilde{\Sigma}_{\varepsilon\varepsilon} \boxtimes \tilde{\Sigma}_{\varepsilon\varepsilon}) - (1+c^{-1})^{-1} (\tilde{\Sigma}_{\varepsilon v} \tilde{\Sigma}_{vv}^{-1} \tilde{\Sigma}_{ve}) \boxtimes (\tilde{\Sigma}_{\varepsilon v} \tilde{\Sigma}_{vv}^{-1} \tilde{\Sigma}_{ve}) ] \psi_p ,$$

$$\tilde{\Sigma}_{\rho\rho} = \tilde{\Sigma}_{uu} - \tilde{\Sigma}_{uv} \tilde{\Sigma}_{vv}^{-1} \tilde{\Sigma}_{vu} ,$$

$$\tilde{\Sigma}_{vv} = (\tilde{I}, -\beta') \tilde{\Sigma}_{\varepsilon\varepsilon} (\tilde{I}, -\beta')' ,$$

$$\tilde{\Sigma}_{uv} = \tilde{\Sigma}'_{vu} = \tilde{\Sigma}_{ue} - \tilde{\Sigma}_{uu} \beta ,$$

$$\tilde{\Sigma}_{ve} = \tilde{\Sigma}'_{v\varepsilon} = (\tilde{I}, -\beta') \tilde{\Sigma}_{\varepsilon\varepsilon} ,$$

$$c = \lim_{n \rightarrow \infty} d^{-1} n .$$

(b) If Assumption 3.2 and Assumption 3.3 hold, then

$$n^{1/2} \begin{bmatrix} (\hat{\beta}_0 - \beta_0)' \\ \text{vec}(\hat{\beta} - \beta) \\ \text{vech}(\hat{\Sigma}_{\epsilon\epsilon} - \Sigma_{\epsilon\epsilon}) \\ \text{vech}(\hat{\Sigma}_{xx} - \Sigma_{xx}) \end{bmatrix} \xrightarrow{L} N \left[ \begin{bmatrix} 0 \\ 0 \\ 0 \\ 0 \end{bmatrix}, \begin{bmatrix} \tilde{V}_{00} & \tilde{V}_{0\beta} & \tilde{V}_{0\epsilon} & \tilde{V}_{0x} \\ \tilde{V}'_{0\beta} & \tilde{V}_{\beta\beta} & \tilde{V}_{\beta\epsilon} & \tilde{V}_{\beta x} \\ \tilde{V}'_{0\epsilon} & \tilde{V}'_{\beta\epsilon} & \tilde{V}_{\epsilon\epsilon} & \tilde{V}_{\epsilon x} \\ \tilde{V}'_{0x} & \tilde{V}'_{\beta x} & \tilde{V}'_{\epsilon x} & \tilde{V}_{xx} \end{bmatrix} \right],$$

where

$$\tilde{V}_{0x} = -(\tilde{\Gamma}_{rxr} \boxtimes \mu_x)' \tilde{V}_{\beta x},$$

$$\tilde{V}_{\beta x} = 2\{\tilde{\Sigma}_{vu} \boxtimes \Sigma_{xx}^{-1}[\tilde{\Sigma}_{xx} + (1+c)\tilde{\Sigma}_{\rho\rho}]\}\psi'_k,$$

$$\tilde{V}_{\epsilon x} = 2c\psi_p[(\hat{\Sigma}_{\epsilon u} \boxtimes \hat{\Sigma}_{\epsilon u}) - (\tilde{\Sigma}_{\epsilon u} \boxtimes \tilde{\Sigma}_{\epsilon u})]\psi'_k,$$

$$\begin{aligned} \tilde{V}_{xx} &= \tilde{\Gamma}_x + 2\psi_k[\tilde{\Sigma}_{xx} \boxtimes \tilde{\Sigma}_{uu} + \tilde{\Sigma}_{uu} \boxtimes \tilde{\Sigma}_{xx} \\ &\quad + (1+c)(\tilde{\Sigma}_{uu} \boxtimes \tilde{\Sigma}_{uu} - \hat{\Sigma}_{uu} \boxtimes \hat{\Sigma}_{uu})]\psi'_k, \end{aligned}$$

$$\tilde{\Gamma}_x = V\{\text{vech}[(x_t - \mu_x)'(x_t - \mu_x)]\},$$

$$\hat{\Sigma}_{uu} = \tilde{\Sigma}_{uv} \tilde{\Sigma}_{vv}^{-1} \tilde{\Sigma}_{vu},$$

$$\hat{\Sigma}_{ue} = \tilde{\Sigma}_{uv} \tilde{\Sigma}_{vv}^{-1} \tilde{\Sigma}_{ve}.$$

Proof. (a) We observe that

$$\begin{aligned}
 S_{\varepsilon\varepsilon}^{-1} m_{ZZ} &= S_{\varepsilon\varepsilon}^{-1/2} S_{\varepsilon\varepsilon}^{-1/2} m_{ZZ} S_{\varepsilon\varepsilon}^{-1/2} S_{\varepsilon\varepsilon}^{1/2} \\
 &= S_{\varepsilon\varepsilon}^{-1/2} (\hat{Q}_{(1)} \hat{A}_1 \hat{Q}'_{(1)} + \hat{Q}_{(2)} \hat{A}_2 \hat{Q}'_{(2)}) S_{\varepsilon\varepsilon}^{1/2} \\
 &= S_{\varepsilon\varepsilon}^{-1} \hat{P}_{(1)} \hat{A}_1 \hat{P}'_{(1)} + \hat{T}_{(2)} \hat{A}_2 \hat{T}'_{(2)} S_{\varepsilon\varepsilon} \\
 &= S_{\varepsilon\varepsilon}^{-1} (\hat{\beta}, \hat{I})' \hat{D}_{kk} (\hat{\beta}, \hat{I}) + (\hat{I}, -\hat{\beta}')' \hat{D}_{rr} (\hat{I}, -\hat{\beta}') S_{\varepsilon\varepsilon} .
 \end{aligned} \tag{3.89}$$

Therefore,

$$\begin{aligned}
 (\hat{\beta}, \hat{I}) S_{\varepsilon\varepsilon}^{-1} m_{ZZ} (\hat{I}, -\hat{\beta}')' &= (\hat{\beta}, \hat{I}) S_{\varepsilon\varepsilon}^{-1} (\hat{\beta}, \hat{I})' \hat{D}_{kk} (\hat{\beta} - \hat{\beta}) \\
 &\quad - (\hat{\beta} - \hat{\beta}) \hat{D}_{rr} (\hat{I}, -\hat{\beta}') S_{\varepsilon\varepsilon} (\hat{I}, -\hat{\beta}')' \\
 &= U_{kk} (\hat{\beta} - \hat{\beta}) + (\hat{\beta} - \hat{\beta}) U_{rr} ,
 \end{aligned}$$

where

$$\begin{aligned}
 U_{kk} &= (\hat{\beta}, \hat{I}) S_{\varepsilon\varepsilon}^{-1} (\hat{\beta}, \hat{I})' \hat{D}_{kk} , \\
 U_{rr} &= - \hat{D}_{rr} (\hat{I}, -\hat{\beta}') S_{\varepsilon\varepsilon} (\hat{I}, -\hat{\beta}) .
 \end{aligned}$$

Hence, using Theorem 2.2, we obtain

$$\begin{aligned}
\text{vec}[(\beta, I) S_{\varepsilon\varepsilon}^{-1} (I, -\beta')'] &= \text{vec}[U_{kk}(\hat{\beta} - \beta) I_{rxr}] + \text{vec}[I_{kxk}(\hat{\beta} - \beta) U_{rr}] \\
&= (I_{rxr} \otimes U_{kk}) \text{vec}(\hat{\beta} - \beta) \\
&\quad + (U'_{rr} \otimes I_{kxk}) \text{vec}(\hat{\beta} - \beta) \\
&= (I_{kxk} \otimes U_{kk} + U'_{rr} \otimes I_{kxk}) \text{vec}(\hat{\beta} - \beta) .
\end{aligned} \tag{3.90}$$

By Assumption 3.7, Theorem 2.1, and Theorem 3.4,

$$\begin{aligned}
I_{rxr} \otimes U_{kk} + U'_{rr} \otimes I_{kxk} &\xrightarrow{\text{a.s.}} I_{rxr} \otimes (\Sigma_{\rho\rho}^{-1} \Sigma_{\eta\eta}) - (\Sigma_{vv}^{-1} \Sigma_{vv}) \otimes I_{kxk} \\
&= I_{rxr} \otimes (\Sigma_{\rho\rho}^{-1} \Sigma_{xx}) ,
\end{aligned} \tag{3.91}$$

and the limit is nonsingular. Thus,

$$I_{rxr} \otimes U_{kk} + U'_{rr} \otimes I_{kxk}$$

is nonsingular for large  $n$ . Hence, by (3.90),

$$\begin{aligned}
\text{vec}(\hat{\beta} - \beta) &= (I_{rxr} \otimes U_{kk} + U'_{rr} \otimes I_{kxk})^{-1} \text{vec}[(\beta, I) S_{\varepsilon\varepsilon}^{-1} (I, -\beta')'] \\
&= (I_{rxr} \otimes U_{kk} + U'_{rr} \otimes I_{kxk})^{-1} \text{vec}\{(\beta, I) S_{\varepsilon\varepsilon}^{-1} [(\beta, I)' \Sigma_{x\varepsilon}^{-1}
\end{aligned}$$

$$+ \mathfrak{m}_{\varepsilon\varepsilon} ](\mathfrak{I}, -\beta')' \} \quad . \quad (3.92)$$

Using Theorem 2.13 if Assumption 3.1 holds, and using Theorem 2.10 if Assumption 3.2 holds, we obtain

$$(\beta, \mathfrak{I})' \mathfrak{m}_{\mathbf{x}\varepsilon} + \mathfrak{m}_{\varepsilon\varepsilon} - \Sigma_{\varepsilon\varepsilon} = o_p(n^{-1/2}) \quad . \quad (3.93)$$

By Assumption 3.8,

$$\mathfrak{S}_{\varepsilon\varepsilon} - \Sigma_{\varepsilon\varepsilon} = o_p(n^{-1/2}) \quad . \quad (3.94)$$

Using the Taylor expansion of a matrix, we have

$$\mathfrak{S}_{\varepsilon\varepsilon}^{-1} = \Sigma_{\varepsilon\varepsilon}^{-1} - \Sigma_{\varepsilon\varepsilon}^{-1}(\mathfrak{S}_{\varepsilon\varepsilon} - \Sigma_{\varepsilon\varepsilon})\Sigma_{\varepsilon\varepsilon}^{-1} + o_p(n^{-1}) \quad .$$

Hence,

$$\begin{aligned} & (\beta, \mathfrak{I})' \mathfrak{S}_{\varepsilon\varepsilon}^{-1} [(\beta, \mathfrak{I})' \mathfrak{m}_{\mathbf{x}\varepsilon} + \mathfrak{m}_{\varepsilon\varepsilon}] (\mathfrak{I}, -\beta')' \\ &= (\beta, \mathfrak{I})' [\Sigma_{\varepsilon\varepsilon}^{-1} - \Sigma_{\varepsilon\varepsilon}^{-1}(\mathfrak{S}_{\varepsilon\varepsilon} - \Sigma_{\varepsilon\varepsilon})\Sigma_{\varepsilon\varepsilon}^{-1}] [(\beta, \mathfrak{I})' \mathfrak{m}_{\mathbf{x}\varepsilon} + \mathfrak{m}_{\varepsilon\varepsilon} - \Sigma_{\varepsilon\varepsilon} + \Sigma_{\varepsilon\varepsilon}] (\mathfrak{I}, -\beta')' \\ & \quad + o_p(n^{-1}) \\ &= (\beta, \mathfrak{I})' \Sigma_{\varepsilon\varepsilon}^{-1} [(\beta, \mathfrak{I})' \mathfrak{m}_{\mathbf{x}\varepsilon} + \mathfrak{m}_{\varepsilon\varepsilon} - \Sigma_{\varepsilon\varepsilon}] (\mathfrak{I}, -\beta')' - (\beta, \mathfrak{I})' \Sigma_{\varepsilon\varepsilon}^{-1} (\mathfrak{S}_{\varepsilon\varepsilon} - \Sigma_{\varepsilon\varepsilon}) (\mathfrak{I}, -\beta')' \end{aligned}$$

$$\begin{aligned}
& + o_p(n^{-1}) \\
& = \Sigma_{\rho\rho}^{-1}(\underline{m}_{xv} + \underline{m}_{\xi v} - S_{\xi v}) + o_p(n^{-1}) \\
& = o_p(n^{-1/2}) \quad , \tag{3.95}
\end{aligned}$$

where

$$\begin{aligned}
\underline{m}_{xv} &= \underline{m}_{x\varepsilon}(\underline{I}, -\underline{\beta}')' \quad , \\
\underline{m}_{\xi v} &= \Sigma_{\rho\rho}(\underline{\beta}, \underline{I}) \Sigma_{\varepsilon\varepsilon}^{-1} \underline{m}_{\varepsilon\varepsilon}(\underline{I}, -\underline{\beta}')' \quad , \\
S_{\xi v} &= \Sigma_{\rho\rho}(\underline{\beta}, \underline{I}) \Sigma_{\varepsilon\varepsilon}^{-1} S_{\varepsilon\varepsilon}(\underline{I}, -\underline{\beta}')' \quad ,
\end{aligned}$$

and we have used Theorem 2.1. By (3.91), (3.92), and (3.95),

$$\begin{aligned}
\text{vec}(\hat{\underline{\beta}} - \underline{\beta}) &= [\underline{I}_{rxr} \otimes (\Sigma_{\rho\rho}^{-1} \Sigma_{xx})^{-1}] \text{vec}\{\Sigma_{\rho\rho}^{-1}(\underline{m}_{xv} + \underline{m}_{\xi v} - S_{\xi v})\} + o_p(n^{-1/2}) \\
&= (\underline{I}_{rxr} \otimes \Sigma_{xx}^{-1} \Sigma_{\rho\rho})(\underline{I}_{rxr} \otimes \Sigma_{\rho\rho}^{-1}) \text{vec}(\underline{m}_{xv} + \underline{m}_{\xi v} - S_{\xi v}) \\
&\quad + o_p(n^{-1/2}) \\
&= (\underline{I}_{rxr} \otimes \Sigma_{xx}^{-1}) \text{vec}(\underline{m}_{xv} + \underline{m}_{\xi v} - S_{\xi v}) + o_p(n^{-1/2}) \quad . \tag{3.96}
\end{aligned}$$

Observing that

$$\hat{S}_{\varepsilon\varepsilon}^{-1/2} \mathfrak{M}_{ZZ} \hat{S}_{\varepsilon\varepsilon}^{-1/2} \hat{Q}_{(2)} = \hat{Q}_{(2)} \hat{\Lambda}_2 ,$$

we have

$$\begin{aligned} \mathfrak{M}_{ZZ} (\hat{I}, -\hat{\beta}')' \hat{T}_{rr} &= \mathfrak{M}_{ZZ} \hat{T}_{(2)} \\ &= \mathfrak{M}_{ZZ} \hat{S}_{\varepsilon\varepsilon}^{-1/2} \hat{Q}_{(2)} \\ &= \hat{S}_{\varepsilon\varepsilon}^{1/2} \hat{Q}_{(2)} \hat{\Lambda}_2 \\ &= \hat{S}_{\varepsilon\varepsilon} (\hat{I}, -\hat{\beta}')' \hat{T}_{rr} \hat{\Lambda}_2 . \end{aligned}$$

Multiplying  $(\hat{I}, -\hat{\beta}')'$  from the left, we obtain

$$(\hat{I}, -\hat{\beta}') \mathfrak{M}_{ZZ} (\hat{I}, -\hat{\beta}')' \hat{T}_{rr} = (\hat{I}, -\hat{\beta}) \hat{S}_{\varepsilon\varepsilon} (\hat{I}, -\hat{\beta}')' \hat{T}_{rr} \hat{\Lambda}_2 . \quad (3.97)$$

Hence,

$$\begin{aligned} \hat{T}_{rr} (\hat{\Lambda}_2 - \hat{I}) \hat{T}_{rr}' &= \{ [(\hat{I}, -\hat{\beta}') \hat{S}_{\varepsilon\varepsilon} (\hat{I}, -\hat{\beta}')']^{-1} (\hat{I}, -\hat{\beta}') \mathfrak{M}_{ZZ} (\hat{I}, -\hat{\beta}')' \\ &\quad - \hat{I} \} \hat{T}_{rr} \hat{T}_{rr}' \\ &= [(\hat{I}, -\hat{\beta}') \hat{S}_{\varepsilon\varepsilon} (\hat{I}, -\hat{\beta}')']^{-1} (\hat{I}, -\hat{\beta}') (\mathfrak{M}_{ZZ} - \hat{S}_{\varepsilon\varepsilon}) \end{aligned}$$

$$(\underline{I}, -\hat{\underline{\beta}}')' \hat{\underline{T}}_{rr} \hat{\underline{T}}'_{rr} . \quad (3.98)$$

By (3.96),

$$\hat{\underline{\beta}} = \underline{\beta} + o_p(n^{-1/2}) . \quad (3.99)$$

Thus,

$$(\underline{I}, -\hat{\underline{\beta}}')' \underline{S}_{\epsilon\epsilon} (\underline{I}, -\hat{\underline{\beta}}')' = \underline{\Sigma}_{vv} + o_p(n^{-1/2}) . \quad (3.100)$$

Also, by Theorem 3.4,

$$\hat{\underline{T}}_{rr} \hat{\underline{T}}'_{rr} = \underline{\Sigma}_{vv}^{-1} + o_p(1) . \quad (3.101)$$

Using (3.93), (3.94), and (3.99), we have

$$\begin{aligned} (\underline{I}, -\hat{\underline{\beta}}')' (\underline{m}_{ZZ} - \underline{S}_{\epsilon\epsilon}) (\underline{I}, -\hat{\underline{\beta}}')' &= (\underline{I}, -\hat{\underline{\beta}}')' (\underline{m}_{\epsilon Z} + \underline{m}_{\epsilon\epsilon} - \underline{\Sigma}_{\epsilon\epsilon} + \underline{\Sigma}_{\epsilon\epsilon} - \underline{S}_{\epsilon\epsilon}) \\ &\quad [(\underline{I}, -\hat{\underline{\beta}}')' - (\underline{0}, \hat{\underline{\beta}}' - \underline{\beta}')'] \\ &= (\underline{I}, -\hat{\underline{\beta}}')' (\underline{m}_{\epsilon\epsilon} - \underline{S}_{\epsilon\epsilon}) (\underline{I}, -\hat{\underline{\beta}}')' + o_p(n^{-1}) \\ &= \underline{m}_{vv} - \underline{S}_{vv} + o_p(n^{-1}) \\ &= o_p(n^{-1/2}) , \end{aligned} \quad (3.102)$$

where

$$\underline{m}_{vv} = (n-1)^{-1} \sum_{t=1}^n (\underline{v}_t - \bar{\underline{v}})' (\underline{v}_t - \bar{\underline{v}}) ,$$

$$\underline{v}_t = \underline{e}_t - \underline{u}_t \underline{\beta} ,$$

$$\bar{\underline{v}} = n^{-1} \sum_{t=1}^n \underline{v}_t ,$$

$$\underline{S}_{vv} = (\underline{I}, -\underline{\beta}') \underline{S}_{\varepsilon\varepsilon} (\underline{I}, -\underline{\beta}')' .$$

By (3.98), (3.100), (3.101), and (3.102),

$$\begin{aligned} \hat{\underline{T}}_{rr} (\hat{\underline{A}}_2 - \underline{I}) \hat{\underline{T}}'_{rr} &= \underline{\Sigma}_{vv}^{-1} (\underline{m}_{vv} - \underline{S}_{vv}) \underline{\Sigma}_{vv}^{-1} + o_p(n^{-1/2}) \\ &= o_p(n^{-1/2}) . \end{aligned} \quad (3.103)$$

We have, from (3.9) and (3.89) ,

$$\begin{aligned} \underline{m}_{ZZ} &= (\hat{\underline{\beta}}, \underline{I})' (\hat{\underline{\Sigma}}_{xx} + \hat{\underline{P}}_{kk} \hat{\underline{P}}'_{kk}) (\hat{\underline{\beta}}, \underline{I}) + \underline{S}_{\varepsilon\varepsilon} (\underline{I}, -\hat{\underline{\beta}}')' \hat{\underline{D}}_{rr} (\underline{I}, -\hat{\underline{\beta}}') \underline{S}_{\varepsilon\varepsilon} \\ &= (\hat{\underline{\beta}}, \underline{I})' \hat{\underline{\Sigma}}_{xx} (\hat{\underline{\beta}}, \underline{I}) + \hat{\underline{P}}_{(1)} \hat{\underline{P}}'_{(1)} + \underline{S}_{\varepsilon\varepsilon} \hat{\underline{T}}_{(2)} \hat{\underline{A}}_2 \hat{\underline{T}}'_{(2)} \underline{S}_{\varepsilon\varepsilon} \\ &= (\hat{\underline{\beta}}, \underline{I})' \hat{\underline{\Sigma}}_{xx} (\hat{\underline{\beta}}, \underline{I}) + \underline{S}_{\varepsilon\varepsilon}^{1/2} (\underline{I} - \hat{\underline{Q}}_{(2)} \hat{\underline{Q}}'_{(2)}) \underline{S}_{\varepsilon\varepsilon}^{1/2} \\ &\quad + \underline{S}_{\varepsilon\varepsilon}^{1/2} \hat{\underline{Q}}_{(2)} \hat{\underline{A}}_2 \hat{\underline{Q}}'_{(2)} \underline{S}_{\varepsilon\varepsilon}^{1/2} \end{aligned}$$

$$\begin{aligned}
&= (\hat{\beta}, \hat{I})' \hat{\Sigma}_{XX}' (\hat{\beta}, \hat{I}) + S_{\varepsilon\varepsilon} + S_{\varepsilon\varepsilon}^{1/2} \hat{Q}_{(2)} (\hat{\Lambda}_2 - I) \hat{Q}_{(2)}' S_{\varepsilon\varepsilon}^{1/2} \\
&= (\hat{\beta}, \hat{I})' \hat{\Sigma}_{XX}' (\hat{\beta}, \hat{I}) + S_{\varepsilon\varepsilon} + S_{\varepsilon\varepsilon} (I, -\hat{\beta}')' \hat{T}_{rr} (\hat{\Lambda}_2 - I) \hat{T}_{rr}' (I, -\hat{\beta}') S_{\varepsilon\varepsilon} .
\end{aligned} \tag{3.104}$$

Therefore,

$$\hat{\Sigma}_{\varepsilon\varepsilon} = (n-1+d)^{-1} [d S_{\varepsilon\varepsilon} + (n-1) S_{\varepsilon\varepsilon} + S_{\varepsilon\varepsilon} (I, -\hat{\beta}')' \hat{T}_{rr} (\hat{\Lambda}_2 - I) \hat{T}_{rr}' (I, -\hat{\beta}') S_{\varepsilon\varepsilon}] .$$

Thus, using (3.94), (3.99), and (3.103), we have

$$\hat{\Sigma}_{\varepsilon\varepsilon} = S_{\varepsilon\varepsilon} + (n-1+d)^{-1} (n-1) \Sigma_{\varepsilon V} \Sigma_{VV}^{-1} (\Sigma_{VV} - S_{VV}) \Sigma_{VV}^{-1} \Sigma_{VV}^{-1} \Sigma_{VV\varepsilon} + o_p(n^{-1/2}) . \tag{3.105}$$

Since under either Assumption 3.1 or Assumption 3.2

$$\bar{X} = \mu_X + o_p(1) ,$$

it follows from (3.99) that

$$\begin{aligned}
\hat{\beta}_0 &= \bar{Y} - \bar{X} \hat{\beta} - \bar{X} (\hat{\beta} - \beta) \\
&= \bar{Y} - \mu_X (\hat{\beta} - \beta) + o_p(n^{-1/2}) \\
&= \bar{Y} - [\text{vec}(\hat{\beta} - \beta)]' (I_{r \times r} \otimes \mu_X') + o_p(n^{-1/2}) .
\end{aligned} \tag{3.106}$$

Using Theorem 2.13 if Assumption 3.1 holds, and using Theorem 2.9 if Assumption 3.2 holds, we have

$$\frac{1}{n^{1/2}} \begin{bmatrix} \bar{\tilde{y}}' \\ \text{vec } \tilde{m}_{xv} \\ \text{vec } \tilde{m}_{\xi v} \\ \text{vech}(\tilde{m}_{vv} - \tilde{\Sigma}_{vv}) \end{bmatrix} \xrightarrow{L} N(\tilde{Q}, \tilde{G}_{(1)}) \quad , \quad (3.107)$$

where

$$\tilde{G}_{(1)} = \text{block diag}\{\tilde{\Sigma}_{vv}, \tilde{\Sigma}_{vv} \boxtimes \tilde{\Sigma}_{xx}, \tilde{\Sigma}_{vv} \boxtimes \tilde{\Sigma}_{\rho\rho}, 2\tilde{\Psi}_{\Gamma}(\tilde{\Sigma}_{vv} \boxtimes \tilde{\Sigma}_{vv})\tilde{\Psi}_{\Gamma}'\} \quad .$$

Also, by Assumption 3.8,

$$\frac{1}{n^{1/2}} \begin{bmatrix} \text{vech}(\tilde{S}_{\varepsilon\varepsilon} - \tilde{\Sigma}_{\varepsilon\varepsilon}) \\ \text{vec } \tilde{S}_{\xi v} \\ \text{vech}(\tilde{S}_{vv} - \tilde{\Sigma}_{vv}) \end{bmatrix} = (nd^{-1})^{1/2} d^{1/2} \begin{bmatrix} \text{vech}(\tilde{S}_{\varepsilon\varepsilon} - \tilde{\Sigma}_{\varepsilon\varepsilon}) \\ \text{vec } \tilde{S}_{\xi v} \\ \text{vech}(\tilde{S}_{vv} - \tilde{\Sigma}_{vv}) \end{bmatrix} \xrightarrow{L} N(\tilde{Q}, \tilde{F}_{(1)}) \quad , \quad (3.108)$$

where

$$\tilde{F}_{(1)} = c \begin{bmatrix} 2\tilde{\Psi}_p(\tilde{\Sigma}_{\varepsilon\varepsilon} \boxtimes \tilde{\Sigma}_{\varepsilon\varepsilon})\tilde{\Psi}_p' & \tilde{F}_{12} & 2\tilde{\Psi}_p(\tilde{\Sigma}_{\varepsilon v} \boxtimes \tilde{\Sigma}_{\varepsilon v})\tilde{\Psi}_{\Gamma}' \\ & \tilde{F}_{12}' & \tilde{\Sigma}_{vv} \boxtimes \tilde{\Sigma}_{\rho\rho} & \tilde{Q} \\ 2\tilde{\Psi}_{\Gamma}(\tilde{\Sigma}_{v\varepsilon} \boxtimes \tilde{\Sigma}_{v\varepsilon})\tilde{\Psi}_p' & \tilde{Q} & 2\tilde{\Psi}_{\Gamma}(\tilde{\Sigma}_{vv} \boxtimes \tilde{\Sigma}_{vv})\tilde{\Psi}_{\Gamma}' \end{bmatrix} \quad ,$$

$$\begin{aligned}
F_{12} &= 2\psi_p(\Sigma_{\epsilon\epsilon} \otimes \Sigma_{\epsilon\epsilon})\psi_p'\phi_p' \{(\mathbf{I}, -\beta)'\} \otimes [\Sigma_{\epsilon\epsilon}^{-1}(\beta, \mathbf{I})'\Sigma_{\rho\rho}] \\
&= 2(\Sigma_{\epsilon v} \otimes \Sigma_{\epsilon\xi})\psi_p' .
\end{aligned}$$

By (3.96), (3.105), (3.106), (3.107), (3.108), and the independence of  $\Sigma_{\epsilon\epsilon}$  and  $Z_t$ ,  $n^{1/2}\{(\hat{\beta}_0 - \beta), [\text{vec}(\hat{\beta} - \beta)]', [\text{vech}(\hat{\Sigma}_{\epsilon\epsilon} - \Sigma_{\epsilon\epsilon})]'\}'$  converges in law to a multivariate normal random vector, and also

$$\begin{aligned}
V_{\beta\beta} &= (\mathbf{I} \otimes \Sigma_{xx}^{-1})[\Sigma_{vv} \otimes \Sigma_{xx} + (1+c)(\Sigma_{vv} \otimes \Sigma_{\rho\rho})](\mathbf{I} \otimes \Sigma_{xx}^{-1}) \\
&= \Sigma_{vv} \otimes \{\Sigma_{xx}^{-1}[\Sigma_{xx} + (1+c)\Sigma_{\rho\rho}]\Sigma_{xx}^{-1}\} ,
\end{aligned}$$

$$\begin{aligned}
V_{\beta\epsilon} &= -(\mathbf{I} \otimes \Sigma_{xx}^{-1}) 2c(\Sigma_{v\epsilon} \otimes \Sigma_{\xi\epsilon})\psi_p' \\
&= -2c[\Sigma_{v\epsilon} \otimes (\Sigma_{xx}^{-1}\Sigma_{\xi\epsilon})]\psi_p' ,
\end{aligned}$$

$$\begin{aligned}
V_{\epsilon\epsilon} &= 2c\psi_p(\Sigma_{\epsilon\epsilon} \otimes \Sigma_{\epsilon\epsilon})\psi_p' + (1+c^{-1})^{-2}\psi_p[(\Sigma_{\epsilon v}\Sigma_{vv}^{-1}) \\
&\quad \otimes (\Sigma_{\epsilon v}\Sigma_{vv}^{-1})\phi_r' 2(1+c)\psi_r(\Sigma_{vv} \otimes \Sigma_{vv})\psi_r'\phi_r'[(\Sigma_{vv}^{-1}\Sigma_{v\epsilon}) \otimes (\Sigma_{vv}^{-1}\Sigma_{v\epsilon})]\psi_p' \\
&\quad - 2c(1+c^{-1})^{-1}\psi_p[(\Sigma_{\epsilon v}\Sigma_{vv}^{-1}) \otimes (\Sigma_{\epsilon v}\Sigma_{vv}^{-1})]\phi_r'\psi_r(\Sigma_{v\epsilon} \otimes \Sigma_{v\epsilon})\psi_p' \\
&\quad - 2c(1+c^{-1})^{-1}\psi_p(\Sigma_{\epsilon v} \otimes \Sigma_{\epsilon v})\psi_r'\phi_r'[(\Sigma_{vv}^{-1}\Sigma_{v\epsilon}) \otimes (\Sigma_{vv}^{-1}\Sigma_{v\epsilon})]\psi_p'
\end{aligned}$$

$$\begin{aligned}
&= 2c\psi_p(\Sigma_{\varepsilon\varepsilon} \boxtimes \Sigma_{\varepsilon\varepsilon})\psi_p' + 2(1+c)(1+c^{-1})^{-2}\psi_p[(\Sigma_{\varepsilon v} \Sigma_{vv}^{-1} \Sigma_{v\varepsilon}) \boxtimes (\Sigma_{\varepsilon v} \Sigma_{vv}^{-1} \Sigma_{v\varepsilon})]\psi_p' \\
&\quad - 4c(1+c^{-1})^{-1}\psi_p[(\Sigma_{\varepsilon v} \Sigma_{vv}^{-1} \Sigma_{v\varepsilon}) \boxtimes (\Sigma_{\varepsilon v} \Sigma_{vv}^{-1} \Sigma_{v\varepsilon})]\psi_p' \\
&= 2c\psi_p[(\Sigma_{\varepsilon\varepsilon} \boxtimes \Sigma_{\varepsilon\varepsilon}) - (1+c^{-1})^{-1}(\Sigma_{\varepsilon v} \Sigma_{vv}^{-1} \Sigma_{v\varepsilon}) \boxtimes (\Sigma_{\varepsilon v} \Sigma_{vv}^{-1} \Sigma_{v\varepsilon})]\psi_p',
\end{aligned}$$

where

$$\Sigma_{\xi\varepsilon} = \Sigma_{\rho\rho}(\beta, \mathbb{I})$$

and we have used for any  $(a \times b)$  matrix  $\underline{A}$

$$\begin{aligned}
\psi_a(\underline{A} \boxtimes \underline{A})\psi_b' &= \psi_a\psi_a'(\underline{A} \boxtimes \underline{A})\psi_b' \\
&= \psi_a(\underline{A} \boxtimes \underline{A})\psi_b'.
\end{aligned}$$

(b) Observe that

$$\begin{aligned}
\Sigma_{\varepsilon\varepsilon}(\mathbb{I}, -\hat{\beta}')' &= \Sigma_{\varepsilon\varepsilon}(\mathbb{I}, -\beta')' + o_p(n^{-1/2}) \\
&= \Sigma_{\varepsilon v} + o_p(n^{-1/2}).
\end{aligned} \tag{3.109}$$

Using (3.103) and (3.109), we write the lower  $k \times k$  corner of (3.104)

as

$$\hat{\Sigma}_{xx} = \underline{m}_{xx} - \underline{s}_{uu} - \underline{\Sigma}_{uv} \underline{\Sigma}^{-1}_{vv} (\underline{m}_{vv} - \underline{s}_{vv}) \underline{\Sigma}^{-1}_{vv} \underline{\Sigma}_{vu} + o_p(n^{-1/2}) .$$

Therefore,

$$\begin{aligned} \hat{\Sigma}_{xx} - \underline{\Sigma}_{xx} &= (\underline{m}_{xx} - \underline{\Sigma}_{xx}) - (\underline{m}^{**}_{uu} - \underline{\Sigma}^{**}_{uu}) - (\underline{s}_{uu} - \underline{\Sigma}_{uu}) \\ &\quad + (\underline{\Sigma}^{**}_{uu} - \underline{\Sigma}_{uu}) + o_p(n^{-1/2}) , \end{aligned} \quad (3.110)$$

where

$$\underline{m}^{**}_{uu} = \underline{\Sigma}_{uv} \underline{\Sigma}^{-1}_{vv} \underline{m}_{vv} \underline{\Sigma}^{-1}_{vv} \underline{\Sigma}_{vu} ,$$

$$\underline{s}^{**}_{uu} = \underline{\Sigma}_{uv} \underline{\Sigma}^{-1}_{vv} \underline{s}_{vv} \underline{\Sigma}^{-1}_{vv} \underline{\Sigma}_{vu} ,$$

$$\underline{\Sigma}^{**}_{uu} = \underline{\Sigma}_{uv} \underline{\Sigma}^{-1}_{vv} \underline{\Sigma}_{vu} ,$$

$$\underline{\Sigma}_{xx} = \underline{\Sigma}_{xx} + \underline{\Sigma}_{uu} .$$

Under Assumption 3.2 and Assumption 3.3,

$$n^{1/2} \begin{bmatrix} \tilde{\mathbf{v}}' \\ \text{vec } \tilde{\mathbf{m}}_{\mathbf{xv}} \\ \text{vec } \tilde{\mathbf{m}}_{\xi\mathbf{v}} \\ \text{vech}(\tilde{\mathbf{m}}_{\mathbf{vv}} - \tilde{\Sigma}_{\mathbf{vv}}) \\ \text{vech}(\tilde{\mathbf{m}}_{\mathbf{xx}} - \tilde{\Sigma}_{\mathbf{xx}}) \\ \text{vech}(\tilde{\mathbf{m}}_{\mathbf{uu}}^{\wedge} - \tilde{\Sigma}_{\mathbf{uu}}^{\wedge}) \end{bmatrix} \xrightarrow{\mathbf{L}} \mathbf{N} \left[ \begin{bmatrix} 0 \\ 0 \end{bmatrix}, \begin{bmatrix} \mathbf{G}_{(1)} & \mathbf{G}_{(12)} \\ \mathbf{G}_{(12)}' & \mathbf{G}_{(2)} \end{bmatrix} \right], \quad (3.111)$$

where

$$\mathbf{G}_{(12)}' = \begin{bmatrix} 0 & 2\psi_{\mathbf{k}}(\tilde{\Sigma}_{\mathbf{uv}} \boxtimes \tilde{\Sigma}_{\mathbf{xx}}) & \mathbf{G}_{53} & 2\psi_{\mathbf{r}}(\tilde{\Sigma}_{\mathbf{vu}} \boxtimes \tilde{\Sigma}_{\mathbf{vu}})\psi_{\mathbf{k}}' \\ 0 & 0 & 0 & 2\psi_{\mathbf{r}}(\tilde{\Sigma}_{\mathbf{vu}} \boxtimes \tilde{\Sigma}_{\mathbf{vu}})\psi_{\mathbf{k}}' \end{bmatrix},$$

$$\mathbf{G}_{(2)} = \begin{bmatrix} \Gamma_{\mathbf{x}} + 2\psi_{\mathbf{k}}(\tilde{\Sigma}_{\mathbf{xx}} \boxtimes \tilde{\Sigma}_{\mathbf{uu}} + \tilde{\Sigma}_{\mathbf{uu}} \boxtimes \tilde{\Sigma}_{\mathbf{xx}} + \tilde{\Sigma}_{\mathbf{uu}} \boxtimes \tilde{\Sigma}_{\mathbf{uu}})\psi_{\mathbf{k}}' & 2\psi_{\mathbf{k}}(\tilde{\Sigma}_{\mathbf{uu}}^{\wedge} \boxtimes \tilde{\Sigma}_{\mathbf{uu}}^{\wedge})\psi_{\mathbf{k}}' \\ 2\psi_{\mathbf{k}}(\tilde{\Sigma}_{\mathbf{uu}}^{\wedge} \boxtimes \tilde{\Sigma}_{\mathbf{uu}}^{\wedge})\psi_{\mathbf{k}}' & 2\psi_{\mathbf{k}}(\tilde{\Sigma}_{\mathbf{uu}}^{\wedge} \boxtimes \tilde{\Sigma}_{\mathbf{uu}}^{\wedge})\psi_{\mathbf{k}}' \end{bmatrix},$$

$$\mathbf{G}_{53} = 2\psi_{\mathbf{k}}\{\tilde{\Sigma}_{\mathbf{uv}} \boxtimes [(\mathbf{0}, \mathbf{I})\tilde{\Sigma}_{\mathbf{ee}}\tilde{\Sigma}_{\mathbf{ee}}^{-1}(\mathbf{g}, \mathbf{I})'\tilde{\Sigma}_{\mathbf{pp}}]\}$$

$$= 2\psi_{\mathbf{k}}[\tilde{\Sigma}_{\mathbf{uv}} \boxtimes \tilde{\Sigma}_{\mathbf{pp}}] \quad .$$

Also, by Assumption 3.8,

$$\frac{1}{n^{1/2}} \begin{bmatrix} \text{vech}(\underline{S}_{\varepsilon\varepsilon} - \underline{\Sigma}_{\varepsilon\varepsilon}) \\ \text{vec } \underline{S}_{\varepsilon v} \\ \text{vech}(\underline{S}_{vv} - \underline{\Sigma}_{vv}) \\ \text{vech}(\underline{S}_{uu} - \underline{\Sigma}_{uu}) \\ \text{vech}(\underline{\hat{S}}_{uu} - \underline{\hat{\Sigma}}_{uu}) \end{bmatrix} \xrightarrow{L} N \left[ \begin{bmatrix} 0 \\ 0 \end{bmatrix}, \begin{bmatrix} \underline{F}_{(1)} & \underline{F}_{(12)} \\ \underline{F}'_{(12)} & \underline{F}_{(2)} \end{bmatrix} \right], \quad (3.112)$$

where

$$\underline{F}_{(12)} = c \begin{bmatrix} 2\Psi_p(\underline{\Sigma}_{\varepsilon u} \boxtimes \underline{\Sigma}_{\varepsilon u})\psi'_k & 2\Psi_p(\underline{\hat{\Sigma}}_{\varepsilon u} \boxtimes \underline{\hat{\Sigma}}_{\varepsilon u})\psi'_k \\ 0 & 0 \\ 2\Psi_r(\underline{\Sigma}_{vu} \boxtimes \underline{\Sigma}_{vu})\psi'_k & 2\Psi_r(\underline{\Sigma}_{vu} \boxtimes \underline{\Sigma}_{vu})\psi'_k \end{bmatrix},$$

$$\underline{F}_{(2)} = c \begin{bmatrix} 2\Psi_k(\underline{\Sigma}_{uu} \boxtimes \underline{\Sigma}_{uu})\psi'_k & 2\Psi_k(\underline{\hat{\Sigma}}_{uu} \boxtimes \underline{\hat{\Sigma}}_{uu})\psi'_k \\ 2\Psi_k(\underline{\hat{\Sigma}}_{uu} \boxtimes \underline{\hat{\Sigma}}_{uu})\psi'_k & 2\Psi_k(\underline{\hat{\Sigma}}_{uu} \boxtimes \underline{\hat{\Sigma}}_{uu})\psi'_k \end{bmatrix}.$$

Hence, by (3.96), (3.105), (3.106), (3.110), (3.111), (3.112), and the independence of  $\underline{S}_{\varepsilon\varepsilon}$  and  $\underline{Z}_t$ , the result follows.  $\square$

Note that the existence of the fourth moment of  $\underline{x}_t$  is necessary only for the asymptotic normality of  $\hat{\underline{\Sigma}}_{xx}$ . If Assumption 3.4 holds, i.e., if  $\underline{x}_t$  is normal, then

$$\begin{aligned} \tilde{V}_{XX} = & 2\psi_k' I(\tilde{\Sigma}_{XX} \otimes \tilde{\Sigma}_{XX}) - (\tilde{\Sigma}_{uu}^{\wedge\wedge} \otimes \tilde{\Sigma}_{uu}^{\wedge\wedge}) + c(\tilde{\Sigma}_{uu} \otimes \tilde{\Sigma}_{uu}) \\ & - c(\tilde{\Sigma}_{uu}^{\wedge\wedge} \otimes \tilde{\Sigma}_{uu}^{\wedge\wedge}) \psi_k', \end{aligned} \quad (3.113)$$

where

$$\tilde{\Sigma}_{XX} = \tilde{\Sigma}_{xx} + \tilde{\Sigma}_{uu}.$$

Except for the part associated with  $\tilde{V}_{XX}$ , the limiting covariance matrix of the estimators is a function of the unknown parameters. Thus, a consistent estimator of the limiting covariance matrix can be obtained by replacing  $\beta$ ,  $\Sigma_{\varepsilon\varepsilon}$ ,  $\Sigma_{xx}$ ,  $\mu_x$ , and  $c$  with  $\hat{\beta}$ ,  $\hat{\Sigma}_{\varepsilon\varepsilon}$ ,  $\hat{\Sigma}_{xx}$ ,  $\bar{X}$ , and  $d^{-1}n$ , respectively. A consistent estimator of  $\tilde{V}_{XX}$  can be obtained in the same way if, in addition, the  $\tilde{x}_t$  are assumed to be normally distributed. In the next theorem, we choose slightly different degrees of freedom for the estimated covariance matrices by analogy to usual regression analysis.

**Theorem 3.6.** Let the model (3.1) hold, and let Assumptions 3.5 through 3.8 hold.

(a) If either Assumption 3.1 or Assumption 3.2 holds, then the covariance matrix of the limiting distribution of

$(\hat{\beta}_0, [\text{vec } \hat{\beta}]', [\text{vech } \hat{\Sigma}_{\varepsilon\varepsilon}]')'$  is consistently estimated by  $n \hat{\tilde{V}}_{(a)}$ ,

where

$$\hat{\tilde{V}}_{(a)} = \begin{bmatrix} \hat{\tilde{V}}_{oo} & \hat{\tilde{V}}_{o\beta} & \hat{\tilde{V}}_{o\epsilon} \\ \hat{\tilde{V}}_{o\beta}' & \hat{\tilde{V}}_{\beta\beta} & \hat{\tilde{V}}_{\beta\epsilon} \\ \hat{\tilde{V}}_{o\epsilon}' & \hat{\tilde{V}}_{\beta\epsilon}' & \hat{\tilde{V}}_{\epsilon\epsilon} \end{bmatrix},$$

$$\hat{\tilde{V}}_{oo} = n^{-1} \hat{\tilde{\Sigma}}_{vv} + (\hat{\mathbb{I}}_{rxr} \boxtimes \bar{\tilde{X}}) \hat{\tilde{V}}_{\beta\beta} (\hat{\mathbb{I}}_{rxr} \boxtimes \bar{\tilde{X}}') ,$$

$$\hat{\tilde{V}}_{o\beta} = - (\hat{\mathbb{I}}_{rxr} \boxtimes \bar{\tilde{X}}) \hat{\tilde{V}}_{\beta\beta} ,$$

$$\hat{\tilde{V}}_{o\epsilon} = - (\hat{\mathbb{I}}_{rxr} \boxtimes \bar{\tilde{X}}) \hat{\tilde{V}}_{\beta\epsilon} ,$$

$$\hat{\tilde{V}}_{\beta\beta} = (n-1)^{-1} \hat{\tilde{\Sigma}}_{vv} \boxtimes \hat{\tilde{\Sigma}}_{xx}^{-1} + [(n-1)^{-1} + d^{-1}] \hat{\tilde{\Sigma}}_{vv} \boxtimes (\hat{\tilde{\Sigma}}_{xx}^{-1} \hat{\tilde{\Sigma}}_{\rho\rho} \hat{\tilde{\Sigma}}_{xx}^{-1}) ,$$

$$\hat{\tilde{V}}_{\beta\epsilon} = - 2 d^{-1} \{ \hat{\tilde{\Sigma}}_{v\epsilon} \boxtimes [\hat{\tilde{\Sigma}}_{xx}^{-1} \hat{\tilde{\Sigma}}_{\rho\rho} (\hat{\beta}, \hat{\mathbb{I}})] \} \psi_p' ,$$

$$\hat{\tilde{V}}_{\epsilon\epsilon} = 2 \psi_p \{ d^{-1} (\hat{\tilde{\Sigma}}_{\epsilon\epsilon} \boxtimes \hat{\tilde{\Sigma}}_{\epsilon\epsilon}) - [d^{-1} - (n-1+d)^{-1}] (\hat{\tilde{\Sigma}}_{\epsilon v} \hat{\tilde{\Sigma}}_{vv}^{-1} \hat{\tilde{\Sigma}}_{v\epsilon})$$

$$\boxtimes (\hat{\tilde{\Sigma}}_{\epsilon v} \hat{\tilde{\Sigma}}_{vv}^{-1} \hat{\tilde{\Sigma}}_{v\epsilon}) \} \psi_p' ,$$

$$\hat{\tilde{\Sigma}}_{\rho\rho} = \hat{\tilde{\Sigma}}_{uu} - \hat{\tilde{\Sigma}}_{uv} \hat{\tilde{\Sigma}}_{vv}^{-1} \hat{\tilde{\Sigma}}_{vu}$$

$$\hat{\tilde{\Sigma}}_{vv} = (n-k-1+d)^{-1} (n-1+d) (\hat{\mathbb{I}}, -\hat{\beta}') \hat{\tilde{\Sigma}}_{\epsilon\epsilon} (\hat{\mathbb{I}}, -\hat{\beta}')' ,$$

$$\hat{\Sigma}_{v\varepsilon} = \hat{\Sigma}'_{\varepsilon v} = (\hat{\Sigma}_{v\varepsilon}, \hat{\Sigma}_{vu})$$

$$= (\mathbf{I}_v, -\hat{\beta}') \hat{\Sigma}_{\varepsilon\varepsilon}.$$

(b) If Assumption 3.2 and Assumption 3.4 hold, then the covariance matrix of the limiting of  $(\hat{\beta}_0, [\text{vec } \hat{\beta}]', [\text{vech } \hat{\Sigma}_{\varepsilon\varepsilon}]', [\text{vech } \hat{\Sigma}_{xx}]')$  is consistently estimated by  $n \hat{V}_{(b)}$ , where

$$\hat{V}_{(b)} = \begin{bmatrix} \hat{V}_{(a)} & \hat{C} \\ \hat{C}' & \hat{V}_{xx} \end{bmatrix},$$

$$\hat{C} = (\hat{V}'_{ox}, \hat{V}'_{\beta x}, \hat{V}'_{\varepsilon x})',$$

$$\hat{V}_{ox} = -(\mathbf{I}_{r \times r} \otimes \bar{X}) \hat{V}_{\beta x},$$

$$\hat{V}_{\beta x} = 2\{ (n-1)^{-1} \hat{\Sigma}_{vu} \otimes \mathbf{I}_{k \times k} + [(n-1)^{-1} + d^{-1}] \hat{\Sigma}_{vu} \otimes (\hat{\Sigma}_{xx}^{-1} \hat{\Sigma}_{\rho\rho}) \} \psi'_k,$$

$$\hat{V}_{\varepsilon x} = 2 d^{-1} \psi_p [(\hat{\Sigma}_{\varepsilon u} \hat{\Sigma}_{\varepsilon u} - (\hat{\Sigma}_{\varepsilon u} \otimes \hat{\Sigma}_{\varepsilon u})) \psi'_k],$$

$$\hat{V}_{xx} = 2 \psi_k \{ (n-1)^{-1} (\hat{\Sigma}_{xx} + \hat{\Sigma}_{uu}) \otimes (\hat{\Sigma}_{xx} + \hat{\Sigma}_{uu}) + d^{-1} (\hat{\Sigma}_{uu} \otimes \hat{\Sigma}_{uu})$$

$$- [(n-1)^{-1} + d^{-1}] (\hat{\Sigma}_{uu} \otimes \hat{\Sigma}_{uu}) \} \psi'_k,$$

$$\hat{\Sigma}_{uu}^{\wedge\wedge} = \hat{\Sigma}_{uv} \hat{\Sigma}_{vv}^{-1} \hat{\Sigma}_{vu} \quad ,$$

$$\hat{\Sigma}_{ue}^{\wedge} = \hat{\Sigma}_{\varepsilon u}^{\wedge'} = \hat{\Sigma}_{uv} \hat{\Sigma}_{vv}^{-1} \hat{\Sigma}_{ve} \quad .$$

Proof. The results follow from Theorem 3.4, Theorem 3.5, and (3.113).

□

#### D. Comments

In this section, we comment on the results in this chapter and the work of other authors. The maximum likelihood estimators for the functional model derived in Theorem 3.1 were first obtained by Anderson (1951), and were reproduced by Healy (1980). Our derivation, following that of Healy, clearly pointed out the nature of inconsistency in the maximum likelihood estimator of  $\Sigma_{\varepsilon\varepsilon}$ , and enabled us to obtain a simple adjustment to produce the consistent estimator  $\hat{\Sigma}_{\varepsilon\varepsilon}^{\wedge}$ .

The maximum likelihood estimation of the multivariate structural model considered in Theorem 3.2 and Theorem 3.3 has not been discussed in the literature.

In Theorem 3.4, we showed the strong consistency of the maximum likelihood estimators adjusted for degrees of freedom under the assumption that  $d^{-1}n \longrightarrow c$ . Under the same assumption on the rate of increase of  $d$ , Healy (1980) obtained the almost sure limits of the maximum likelihood estimators without adjustment for degrees of freedom for the functional model with normal errors.

For the same model as that of Healy, Anderson (1951) stated that the limiting distribution of the eigenvectors of a certain matrix can be used to obtain the limiting distribution of  $\hat{\beta}$  as  $d$  increases holding  $n$  fixed. But, he did not obtain an explicit form of the limiting distribution. In Theorem 3.5, we derived the limiting joint distribution of the maximum likelihood estimators adjusted for degrees of freedom of all the structural parameters. We showed in Theorem 3.5 and Theorem 3.6 that the limiting covariance matrix and its estimator have the same forms for a wide range of assumptions on the unobservable true values  $x_t$ . It can be shown that the limiting covariance matrix of the maximum likelihood estimators adjusted for degrees of freedom of  $\beta_0$ ,  $\beta$ , and  $\Sigma_{xx}$  for the model with known  $\Sigma_{\epsilon\epsilon}$  is given by setting  $c = 0$  in the corresponding portions of the covariance matrices in Theorem 3.5. Gleser (1981) gave the covariance matrix of the estimators for  $\Sigma_{\epsilon\epsilon} = I$  known.

#### IV. THE NONLINEAR MODEL WITH KNOWN ERROR COVARIANCE MATRIX

This chapter deals with estimation of the general nonlinear errors-in-variables model under the assumption that the error covariance matrix is known. We consider the functional model with a single relationship. Some results in this chapter are extensions of the work by Wolter and Fuller (1982b). Also, some of the techniques used in the proofs will be discussed thoroughly in Chapter V. Therefore, the presentation of the proofs in this chapter is not as detailed as those in other chapters.

##### A. Introduction

To define the model, let  $\{b_n\}_{n=1}^{\infty}$  and  $\{a_n\}_{n=1}^{\infty}$  be sequences of positive real numbers such that  $n = a_n b_n$  for  $n = 1, 2, \dots, \infty$ ,  $a_n = o(n)$ , and  $b_n = o(n)$ . We assume the existence of a sequence of experiments indexed by  $n$ . Let

$$y_t^0 = f(\underline{x}_t^0; \underline{\beta}^0) ,$$

$$Y_{nt} = y_t^0 + e_{nt} ,$$

$$\underline{X}_{nt} = \underline{x}_t^0 + \underline{u}_{nt} , \quad t = 1, 2, \dots, b_n, \quad (4.1)$$

where  $\underline{x}_t^0$  are  $1 \times q$  vectors of fixed constants belonging to a parameter space  $\underline{\Gamma}$ , a convex subset of  $q$ -dimensional Euclidean space,

$\beta^0$  is a  $k \times 1$  vector belonging to a parameter space  $\Omega$ , a convex compact subset of  $k$ -dimensional Euclidean space,  $f(z; \theta)$  is continuous on  $\Gamma \times \Omega$ ,  $(Y_{nt}, X_{nt})$  are observed in the  $n$ -th experiment, and  $\varepsilon_{nt} = (e_{nt}, u_{nt})$  denote errors of measurement. We assume that  $\varepsilon_{nt}$  are independently and identically distributed with mean zero and covariance matrix  $\Sigma_n$ , and that  $\Sigma_n$  is of order  $a_n^{-1}$ . One way of interpreting this assumption is to let  $a_n$  denote the number of observations made at each point  $(y_t^0, x_t^0)$ ,  $t = 1, 2, \dots, b_n$ . Under this interpretation, the total number of observations is  $b_n a_n = n$  and each of the vectors  $(Y_{nt}, X_{nt})$  used in the analysis is the mean of  $a_n$  observations. Another interpretation is that the asymptotic results obtained with  $n$  tending to infinity hold if the ratios of measurement error variances to the sums of squares of  $b_n$  true values are small. Thus, the asymptotic results can be used as approximations if either the error variances are small or the number of observations is large.

We assume  $f(x; \beta)$  possesses continuous first and second derivatives with respect to both arguments on  $\Gamma \times \Omega$ . Let  $\tilde{f}_x(z; \theta)$  denote the  $q$ -dimensional row vector of partial derivatives of  $f(x; \beta)$  with respect to the elements of  $x$  evaluated at  $(z; \theta)$ ; let  $\tilde{f}_\beta(z; \theta)$  denote the  $k$ -dimensional column vector of partial derivatives with respect to the elements of  $\beta$  evaluated at  $(z; \theta)$ ; let  $\tilde{f}_{\beta x}(z; \theta)$  denote the  $k \times q$  matrix of second partial derivatives with respect to the elements of  $\beta$  and  $x$  evaluated at  $(z; \theta)$ ; let  $\tilde{f}_{xx}(z; \theta)$  denote the  $q \times q$  matrix of second partial derivatives with respect to the elements of  $x$  evaluated at  $(z; \theta)$ ; and let

$f_{\beta\beta}(z; \varrho)$  be the  $k \times k$  matrix of second partial derivatives with respect to the elements of  $\beta$  evaluated at  $(z; \varrho)$ . We also let

$$f_t^0 = f(x_t^0; \beta^0) ,$$

$$\tilde{f}_{xt}^0 = \tilde{f}_x(x_t^0; \beta^0) ,$$

$$\tilde{f}_{\beta t}^0 = \tilde{f}_{\beta}(x_t^0; \beta^0) ,$$

$$\tilde{f}_{xxt}^0 = \tilde{f}_{xx}(x_t^0; \beta^0) ,$$

$$\tilde{f}_{\beta xt}^0 = \tilde{f}_{\beta x}(x_t^0; \beta^0) ,$$

$$\tilde{f}_{\beta\beta t}^0 = \tilde{f}_{\beta\beta}(x_t^0; \beta^0) .$$

We introduce assumptions associated with the model (4.1).

Assumption 4.1. The  $x_t^0$  are fixed constants and are interior points of  $\tilde{I}$ , and  $\tilde{I}$  is a convex subset of  $q$ -dimensional Euclidean space.

Assumption 4.2. The parameter vector  $\beta^0$  is an interior point of  $\tilde{\Omega}$ , and  $\tilde{\Omega}$  is a convex compact subset of  $k$ -dimensional Euclidean space.

Assumption 4.3. The random variables  $\varepsilon_{nt}$ ,  $t = 1, 2, \dots, b_n$ , are independently distributed with mean zero and covariance matrix  $\Sigma_n$ , where  $\Sigma_n$  are known positive definite matrices satisfying

$$\lim_{n \rightarrow \infty} a_n \sum_{\sim n} = \tilde{\Phi} \quad ,$$

for a nonsingular matrix  $\tilde{\Phi}$  .

Assumption 4.4. Let

$$\tilde{m}_{xx} = n^{-1} \sum_{t=1}^{b_n} \sigma_{vnt}^{-2} \tilde{f}_{\beta t}^0 \tilde{f}_{\beta t}^{0'} \quad ,$$

where

$$\sigma_{vnt}^2 = [1, - \tilde{f}_{xt}^0] \sum_{\sim n} [1, - \tilde{f}_{xt}^0]' \quad .$$

Then, for a positive definite matrix  $\tilde{m}_{xx}$  ,

$$\lim_{n \rightarrow \infty} \tilde{m}_{xx} = \tilde{m}_{xx} \quad .$$

Assumption 4.5. There is available a preliminary estimator  $\tilde{\beta}$  in  $\tilde{\Omega}$  satisfying

$$\tilde{\beta} - \beta^0 = o_p(\max[a_n^{-1}, n^{-1/2}]) \quad .$$

Assumption 4.6. The partial derivatives through order two of  $f(\underline{x}; \beta)$  are continuous and bounded on  $\tilde{\Gamma} \times \tilde{\Omega}$  .

Assumption 4.6a. The partial derivatives through order three of  $f(\underline{x}; \beta)$  are continuous and bounded on  $\tilde{\Gamma} \times \tilde{\Omega}$  .

Assumption 4.7. The  $2 + \delta$  moments of  $a_n^{1/2} \varepsilon_{nt}$  are bounded for some  $\delta > 0$ .

Assumption 4.7a. For some real  $L$  and all  $t$  and  $n$ ,

$$E\{|\varepsilon_{nt}|^4\} \leq L a_n^{-2},$$

where the norm is the Euclidean norm.

Assumption 4.8.  $a_n^{-1} = o(n^{-1/2})$ .

Assumption 4.8a.  $a_n^{-1} = o(n^{-1/3})$ .

#### B. Results of Wolter and Fuller

In this section, we summarize the work of Wolter and Fuller (1982b). Assuming a preliminary estimator  $\bar{\beta}$  satisfying Assumption 4.5 is available, they considered two estimators of  $\beta^0$  for the model (4.1).

Under Assumptions 4.1, 4.2, and 4.3 and the assumption of the normality of  $\varepsilon_{nt}$ , the maximum likelihood estimators of  $x_t^0$  and  $\beta^0$  are those values of  $x_t$  and  $\beta$  contained in  $\Gamma \times \Omega$  that minimize the sum of squares

$$\sum_{t=1}^{b_n} q(\beta, x_t; Y_{nt}, X_{nt})$$

$$= \sum_{t=1}^{b_n} [Y_{nt} - f(\tilde{x}_t; \beta), \tilde{x}_{nt} - \tilde{x}_t] \Sigma_n^{-1} [Y_{nt} - f(\tilde{x}_t; \beta), \tilde{x}_{nt} - \tilde{x}_t]' . \quad (4.2)$$

While an explicit expression for the maximum likelihood estimator of  $\beta^0$  has not been obtained, Wolter and Fuller considered an iterative procedure leading to an estimator of  $\beta^0$ . Given a preliminary estimator  $\bar{\beta}$  of  $\beta^0$ , let  $\bar{x}_t$  be the value of  $x_t$  contained in  $\bar{I}$  that minimizes  $q(\bar{\beta}, \bar{x}_t; Y_{nt}, \tilde{x}_{nt})$ . The local approximation to the sum of squares (4.2) is

$$\sum_{t=1}^{b_n} [\bar{e}_{nt} - \Delta y_t, \bar{u}_{nt} - \Delta x_t] \Sigma_n^{-1} [\bar{e}_{nt} - \Delta y_t, \bar{u}_{nt} - \Delta x_t]' , \quad (4.3)$$

where

$$\bar{e}_{nt} = Y_{nt} - f(\bar{x}_t; \bar{\beta}) ,$$

$$\bar{u}_{nt} = \tilde{x}_{nt} - \bar{x}_t ,$$

$$\Delta y_t = f(\tilde{x}_t; \beta) - f(\bar{x}_t; \bar{\beta}) ,$$

$$\Delta x_t = \tilde{x}_t - \bar{x}_t .$$

Retaining only linear terms in the Taylor expansion of  $f(\underline{x}_t, \underline{\beta})$ , we obtain

$$\Delta y_t = \underline{f}'_{\beta}(\bar{\underline{x}}_t; \bar{\underline{\beta}})(\Delta \underline{\beta}) + \underline{f}'_{\underline{x}}(\bar{\underline{x}}_t; \bar{\underline{\beta}})(\Delta \underline{x}_t)' , \quad (4.4)$$

where

$$\Delta \underline{\beta} = \underline{\beta} - \bar{\underline{\beta}} .$$

Minimizing (4.3) with respect to  $\Delta y_t$ ,  $\Delta \underline{x}_t$ , and  $\Delta \underline{\beta}$  subject to (4.4), Wolter and Fuller obtained an improved estimator of  $\underline{\beta}^0$  given by

$$\hat{\underline{\beta}} = \bar{\underline{\beta}} + \Delta \hat{\underline{\beta}} , \quad (4.5)$$

where  $\Delta \hat{\underline{\beta}}$  satisfies

$$\bar{\underline{M}}_{\underline{x}\underline{x}}(\Delta \hat{\underline{\beta}}) = n^{-1} \sum_{t=1}^{b_n} \bar{\sigma}_{vnt}^{-2} \underline{f}_{\beta}(\bar{\underline{x}}_t; \bar{\underline{\beta}}) [\bar{e}_{nt} - \bar{u}_{nt} \underline{f}'_{\underline{x}}(\bar{\underline{x}}_t; \bar{\underline{\beta}})] ,$$

$$\bar{\sigma}_{vnt}^2 = [1, - \underline{f}'_{\underline{x}}(\bar{\underline{x}}_t; \bar{\underline{\beta}})] \underline{\Sigma}_n [1, - \underline{f}'_{\underline{x}}(\bar{\underline{x}}_t; \bar{\underline{\beta}})]' ,$$

$$\bar{\underline{M}}_{\underline{x}\underline{x}} = n^{-1} \sum_{t=1}^{b_n} \bar{\sigma}_{vnt}^{-2} \underline{f}_{\beta}(\bar{\underline{x}}_t; \bar{\underline{\beta}}) \underline{f}'_{\beta}(\bar{\underline{x}}_t; \bar{\underline{\beta}}) .$$

Given the preliminary estimators  $\bar{\beta}$  and  $\bar{x}_t$ , Wolter and Fuller introduced a modified estimator of  $\beta^0$  given by

$$\tilde{\beta} = \bar{\beta} + \Delta\tilde{\beta}, \quad (4.6)$$

where  $\Delta\tilde{\beta}$  satisfies

$$\bar{M}_{xx}(\Delta\tilde{\beta}) = n^{-1} \sum_{t=1}^{b_n} \bar{\sigma}_{vnt}^{-2} f_{\beta}(\bar{x}_t; \bar{\beta}) \bar{v}_{nt},$$

$$\bar{v}_{nt} = \bar{e}_{nt} - \bar{u}_{nt} f'_x(\bar{x}_t; \bar{\beta}) - \frac{1}{2} \text{tr}[f_{xx}(\bar{x}_t; \bar{\beta})(\bar{u}'_{nt} \bar{u}_{nt} - \Sigma_{uun})],$$

$$\Sigma_n = \begin{bmatrix} \sigma_{een} & \Sigma_{eun} \\ \Sigma_{uen} & \Sigma_{uun} \end{bmatrix}.$$

Without proofs, we state the results and some extensions of Wolter and Fuller (1982b).

Result 4.1. Let the model (4.1) hold, and let Assumptions 4.1, 4.2, 4.3, 4.5, and 4.6 hold. Then,

$$\tilde{x}_t = x_t^0 + \delta_{nt} + O_p(\max[a_n^{-1}, n^{-1/2}]) ,$$

where

$$\delta_{nt} = \varepsilon_{nt} \Sigma_n^{-1} [f_{xt}^0, I]' \Lambda_{nt}^{-1},$$

$$\Lambda_{nt} = [f_{xt}^{0'}, \underline{I}] \Sigma_n^{-1} [f_{xt}^{0'}, \underline{I}]' .$$

Also, there is a constant  $K_0$  such that for  $t = 1, 2, \dots, b_n$ ,

$$E\{|\bar{x}_t - x_t^0|^2\} < K_0 a_n^{-1} .$$

If, in addition, Assumption 4.7a holds, there is a constant  $K$  such that for  $t = 1, 2, \dots, b_n$ ,

$$E\{|\bar{x}_t - x_t^0|^4\} < K a_n^{-2} .$$

Observe that the leading term in the error of estimation, denoted by  $\delta_{nt}$ , is of the form of the error in the generalized least squares estimator with  $[f_{xt}^{0'}, \underline{I}]$  the "X-matrix" and  $\Sigma_n$  the covariance matrix of the error vector  $\varepsilon_{nt}$ .

Result 4.2. Let the model (4.1) hold, and let Assumptions 4.1, 4.2, 4.3, 4.5, and 4.6 hold. Then,

$$a_n \bar{\sigma}_{vnt}^2 = a_n \sigma_{vnt}^2 + a_n \Delta_{vnt} + O_p(\max[a_n^{-1}, n^{-1/2}]) ,$$

where

$$\Delta_{vnt} = 2 \delta_{nt} f_{xxt}^0 [\Sigma_{uun} f_{xt}^{0'} - \Sigma_{uen}] .$$

Result 4.3. Let the model (4.1) hold, and let Assumptions 4.1 through 4.6 hold. Then,

$$\bar{\tilde{M}}_{xx} = \tilde{m}_{xx} + o_p(\max[a_n^{-1}, n^{-1/2}]) ,$$

$$\begin{aligned} \hat{\beta} - \beta^0 &= \tilde{m}_{xx}^{-1} (\delta_1 \hat{\beta}) + o_p(a_n^{-1}) \\ &= o_p(\max[a_n^{-1}, n^{-1/2}]) , \end{aligned}$$

where

$$\delta_1 \hat{\beta} = n^{-1} \sum_{t=1}^{b_n} \sigma_{vnt}^{-2} \tilde{f}_{\beta t}^0 v_{nt}^* ,$$

$$v_{nt}^* = e_{nt} - u_{nt} \tilde{f}_{xt}^0 .$$

If, in addition, Assumptions 4.7 and 4.8 hold, then

$$n^{1/2} (\hat{\beta} - \beta^0) \xrightarrow{L} N(0, \bar{\tilde{m}}_{xx}^{-1}) .$$

Result 4.4. Let the model (4.1) hold, and let Assumptions 4.1 through 4.5, 4.6a, and 4.7a hold. Then,

$$\tilde{\beta} - \beta^0 = \tilde{m}_{xx}^{-1} (\delta_1 \hat{\beta}) + o_p(\max[a_n^{-3/2}, a_n^{-1/2} n^{-1/2}])$$

$$= O_p(\max[a_n^{-3/2}, n^{-1/2}]) .$$

If, in addition, Assumption 4.8a holds, then

$$n^{1/2}(\tilde{\beta} - \beta^0) \xrightarrow{L} N(0, \bar{m}_{xx}^{-1}) .$$

The procedure for obtaining  $\hat{\beta}$  can be iterated using  $\hat{\beta}$  as the preliminary estimator in the second round of calculation. Since the  $\hat{\beta}$  was obtained through the linearization of the sum of squares (4.2), the maximum likelihood estimators of  $x_t$  and  $\beta^0$  can be obtained as the limits of the iteration by including a modification in the procedure to guarantee convergence. By Result 4.3, the  $\hat{\beta}$  satisfies the same property as the  $\tilde{\beta}$  ;

$$\hat{\beta} - \beta^0 = O_p(\max[a_n^{-1}, n^{-1/2}]) .$$

Thus, Result 4.3 holds for the final estimator obtained by any finite number of iterations. The procedure for obtaining  $\tilde{\beta}$  can also be iterated, and for any finite number of iterations the asymptotic properties of the final estimator are given in Result 4.4.

We call  $\hat{\beta}$  the pseudo-maximum likelihood estimator, and  $\tilde{\beta}$  the modified maximum likelihood estimator.

Recall that we assumed  $a_n = o(n)$  and  $b_n = o(n)$  . If  $b_n$  is a constant and  $a_n = O(n)$  , then  $\hat{\beta}$  and  $\tilde{\beta}$  are still consistent. But,

if  $a_n = O(n)$ , the limiting distributions of  $n^{1/2}(\hat{\beta} - \beta^0)$  and  $n^{1/2}(\tilde{\beta} - \beta^0)$  are not normal unless  $\varepsilon_{nt}$  are normally distributed.

Wolter and Fuller suggested certain modifications to improve the small sample properties of  $\tilde{\beta}$ . Their Monte Carlo study showed that such modified estimators were superior to  $\tilde{\beta}$ . They also showed that under some assumptions the ordinary least squares estimator  $\hat{\beta}_L$  of  $\beta^0$  obtained by minimizing

$$\sum_{t=1}^{b_n} [Y_{nt} - f(\tilde{x}_{nt}; \beta)]^2$$

satisfies Assumption 4.5, and thus can be used as the preliminary estimator  $\bar{\beta}$ .

#### C. Bias of the Pseudo-Maximum Likelihood Estimator

In this section, we investigate further the properties of  $\hat{\beta}$  and of an estimator of  $\tilde{x}_t^0$  based on  $\hat{\beta}$ . By Result 4.3, the error in  $\hat{\beta}$  as an estimator of  $\beta^0$  is  $O_p(\max[a_n^{-1}, n^{-1/2}])$ . The term of  $O_p(a_n^{-1})$  is due to the nonlinearity of the function  $f(z; \theta)$ . That is, the nonzero second partial derivative  $f_{xx}(z; \theta)$  introduces a bias in the estimator  $\hat{\beta}$ . We observe that  $\bar{x}_t$  was obtained by projecting the point  $(Y_{nt}, \tilde{x}_{nt})$  onto a nonlinear surface  $f(x; \bar{\beta})$  in the metric  $\Sigma_n^{-1}$ . Thus,  $\bar{x}_t$  is biased, even if  $\bar{\beta}$  is close to  $\beta^0$ . Also,  $\hat{\beta}$

minimizes the sum of distances between  $(Y_{nt}, X_{nt})$  and  $(y_t, x_t)$  by a linearization based on  $[f(\bar{x}_t; \bar{\beta}), \bar{x}_t]$ . Hence, the bias in  $\bar{x}_t$  leads to a bias in  $\hat{\beta}$ . In this section, we obtain an exact expression for the portion of the bias in  $\hat{\beta}$  that is  $O(a_n^{-1})$ . In order to simplify our presentation in the following theorem, we use a stronger set of assumptions than that in Result 4.3.

**Theorem 4.1.** Let the model (4.1) hold, and let Assumptions 4.1 through 4.5, 4.6a, and 4.7a hold. Then,

$$\hat{\beta} - \beta^0 = \bar{\Sigma}_{xx}^{-1} [\delta_1 \hat{\beta} + \delta_2 \hat{\beta}] + O_p(\max[a_n^{-3/2}, n^{-1/2} a_n^{-1/2}]),$$

where

$$\delta_1 \hat{\beta} = n^{-1} \sum_{t=1}^{b_n} \sigma_{vnt}^{-2} f_{\beta t}^0 v_{nt}^* = O_p(n^{-1/2}),$$

$$\delta_2 \hat{\beta} = n^{-1} \sum_{t=1}^{b_n} \sigma_{vnt}^{-2} f_{\beta t}^0 \hat{\varepsilon}_{nt} f_{xxt}^0 \left( \frac{1}{2} \hat{\varepsilon}_{nt}' - u_{nt}' \right)$$

$$= O_p(a_n^{-1}).$$

$$\hat{\varepsilon}_{nt} = \varepsilon_{nt} \Sigma_n^{-1} [f_{xt}^0, \mathbf{I}]' \Lambda_{nt}^{-1},$$

$$\Lambda_{nt} = [f_{xt}^0, \mathbf{I}] \Sigma_n^{-1} [f_{xt}^0, \mathbf{I}]',$$

and  $\tilde{f}_{\beta t}^0$ ,  $\tilde{f}_{xt}^0$ , and  $\tilde{f}_{xxt}^0$  were defined following (4.1). Also,

$$E\{\delta_1 \hat{\beta}\} = 0,$$

$$\begin{aligned} E\{\delta_2 \hat{\beta}\} &= -\frac{1}{2} b_n^{-1} \sum_{t=1}^{b_n} (a_n \sigma_{vnt}^2)^{-1} \tilde{f}_{\beta t}^0 \operatorname{tr}[\tilde{f}_{xxt}^0 \Lambda_{nt}^{-1}] \\ &= O(a_n^{-1}). \end{aligned}$$

Proof. By Result 4.3,

$$\bar{M}_{xx} = m_{xx} + O_p(\max[a_n^{-1}, n^{-1/2}]) \quad (4.7)$$

where  $\bar{M}_{xx}$  is defined following (4.5). By Result 4.1, for  $t = 1, 2, \dots, b_n$ ,

$$E\{|\bar{z}_t - z_t^0|^4\} \leq K a_n^{-2},$$

and it follows that for a constant  $K_1$  and for  $t = 1, 2, \dots, b_n$ ,

$$E\{|\bar{z}_t - z_t^0|\} \leq K_1 a_n^{-1/2}, \quad (4.8)$$

$$E\{|\bar{z}_t - z_t^0|^2\} \leq K_1 a_n^{-1}, \quad (4.9)$$

$$E\{|\bar{x}_t - \bar{x}_t^0|^3\} \leq K_1 a_n^{-3/2} . \quad (4.10)$$

Also, it can be shown that

$$b_n^{-1} \sum_{t=1}^{b_n} [(\bar{\sigma}_{vnt}^2)^{-1} - (a_n \sigma_{vnt}^2)^{-1} + (a_n \sigma_{vnt}^2)^{-2} (a_n \Delta_{vnt})]^2 = o_p(a_n^{-2}) . \quad (4.11)$$

Using Assumptions 4.4 and 4.6a, (4.8), (4.9), (4.10), and (4.11), we obtain

$$\begin{aligned} & n^{-1} \sum_{t=1}^{b_n} (\bar{\sigma}_{vnt}^2)^{-1} \bar{f}_\beta(\bar{x}_t; \bar{\beta}) [\bar{e}_{nt} - \bar{u}_{nt} \bar{f}'_x(\bar{x}_t; \bar{\beta})] \\ &= \bar{M}_{xx}(\beta^0 - \bar{\beta}) + b_n^{-1} \sum_{t=1}^{b_n} [(a_n \sigma_{vnt}^2)^{-1} - (a_n \sigma_{vnt}^2)^{-2} a_n \Delta_{vnt}] \\ & \quad [\bar{f}_{\beta t}^0 + \bar{f}_{\beta xt}^0 \frac{\delta'_t}{\bar{\sigma}_{vnt}}] [\bar{v}_{nt}^* + \bar{\delta}_{nt} \bar{f}_{xxt}^0 (\frac{1}{2} \frac{\delta'_t}{\bar{\sigma}_{nt}} - \bar{u}_{nt}')] \\ & \quad + o_p(\max[a_n^{-3/2}, n^{-1/2} a_n^{-1/2}]) . \end{aligned} \quad (4.12)$$

Therefore, by (4.7) and (4.12),

$$\begin{aligned}
\mathbf{m}_{xx}(\hat{\beta} - \beta^0) &= \delta_1 \hat{\beta} + \delta_2 \hat{\beta} + b_n^{-1} \sum_{t=1}^{b_n} (a_n \sigma_{vnt}^2)^{-1} [\mathbf{f}_{\beta xt}^0 \delta' - \sigma_{vnt}^{-2} \mathbf{f}_{\beta t}^0 \Delta_{vnt}] v_{nt}^* \\
&+ o_p(\max[a_n^{-3/2}, n^{-1/2} a_n^{-1/2}]) \quad . \quad (4.13)
\end{aligned}$$

We observe that

$$E\{\hat{\delta}_{nt}' v_{nt}^*\} = \mathbf{A}_{nt}^{-1} [\mathbf{f}_{xt}^0, \mathbf{I}] \Sigma_n^{-1} E\{\varepsilon_{nt}' \varepsilon_{nt}\} [1, -\mathbf{f}_{xt}^0]' = 0 \quad (4.14)$$

Also, by Assumption 4.7a,

$$E\{\hat{\delta}_{nt}' \hat{\delta}_{nt} v_{nt}^{*2}\} = o(a_n^{-2}) \quad . \quad (4.15)$$

Since  $\Delta_{vnt}$  is a linear combination of  $\hat{\delta}_{nt}$ , it follows from (4.14) and (4.15) that the third term in the right hand side of (4.13) is  $o_p(n^{-1/2} a_n^{-1/2})$ . Since  $\delta_1 \hat{\beta}$  is a linear combination of  $v_{nt}^*$ , by Assumption 4.3 and 4.4,

$$E\{\delta_1 \hat{\beta}\} = 0 \quad ,$$

$$V\{\delta_1 \hat{\beta}\} = o(b_n^{-1} a_n^{-1})$$

$$= o(n^{-1}) \quad ,$$

and thus,

$$\delta_{1\hat{\beta}} = o_p(n^{-1/2}) \quad .$$

We observe that by Assumptions 4.3, 4.4, and 4.6a, for some constant  $L_1$ ,

$$\begin{aligned} E\{|\delta_{2\hat{\beta}}|\} &= L_1 b_n^{-1} \sum_{t=1}^{b_n} E\{|\hat{\delta}_{nt}|^4 + |\hat{\delta}_{nt}|^2 |\underline{u}_{nt}|^2\} \\ &= o(a_n^{-1}) \quad . \end{aligned}$$

Hence,

$$\delta_{2\hat{\beta}} = o_p(a_n^{-1}) \quad .$$

Finally,

$$E\{\text{tr}[\hat{\delta}_{nt}' \hat{\delta}_{nt}]\} = \Lambda_{nt}^{-1} \quad ,$$

$$\begin{aligned} E\{\text{tr}[\underline{u}_{nt}' \hat{\delta}_{nt}]\} &= \Lambda_{nt}^{-1} \\ &= o(a_n^{-1}) \quad . \end{aligned}$$

Therefore, the results follow.  $\square$

Hence, taking the expectation through terms of  $o_p(a_n^{-1})$ , the bias

in  $\hat{\beta}$  as an estimator of  $\beta^0$  is

$$-\frac{1}{2} \mathbf{m}_{xx}^{-1} \mathbf{b}_n^{-1} \sum_{t=1}^{b_n} (\mathbf{a}_n \sigma_{vnt}^2)^{-1} \mathbf{f}_{\beta t}^0 \operatorname{tr}[\mathbf{f}_{xxt}^0 \mathbf{\Lambda}_{nt}^{-1}] , \quad (4.16)$$

which is of  $O(\mathbf{a}_n^{-1})$ . The bias (4.16) is due to the nonlinearity of the function  $f(\mathbf{x}; \beta)$ , and vanishes if the model is linear in  $\mathbf{x}_t$ , i.e., if  $\mathbf{f}_{xxt}^0 = \mathbf{0}$  for all  $t$ . Note that by Theorem 2.1

$$\mathbf{\Lambda}_{nt}^{-1} = \Sigma_{uun} - \Sigma_{uvnt} \sigma_{vnt}^{-2} \Sigma_{vunt} ,$$

where

$$\begin{aligned} \Sigma_{uvnt} &= \Sigma'_{vunt} \\ &= \Sigma_{uen} - \Sigma_{uun} \mathbf{f}_{xt}^{0'} . \end{aligned}$$

We observe that the bias (4.16) is a weighted average of

$$-\frac{1}{2} \operatorname{tr}[\mathbf{f}_{xxt}^0 \mathbf{\Lambda}_{nt}^{-1}] , \quad t = 1, 2, \dots, b_n, \quad (4.17)$$

where the weights are

$$\mathbf{m}_{xx}^{-1} \mathbf{b}_n^{-1} (\mathbf{a}_n \sigma_{vnt}^2)^{-1} \mathbf{f}_{\beta t}^0 , \quad t = 1, 2, \dots, b_n .$$

Thus, the quantity in (4.17) can be considered as a contribution of each true value to the bias of  $\hat{\beta}$ . Such a contribution is small for an  $x_t^0$  at which either  $f_{\text{xx}t}^0$  or  $\Lambda_{\text{nt}}^{-1}$  is small. Note that  $f_{\text{xx}t}^0$  is the nonlinearity at the point  $x_t^0$ , and that  $\Lambda_{\text{nt}}^{-1}$  is the approximate covariance matrix of  $\bar{x}_t$  as an estimator of  $x_t^0$ .

We consider a simple quadratic model as an example. Let

$$y_t^0 = \beta_0 + \beta_1 x_t^0 + \beta_2 (x_t^0)^2,$$

$$\Sigma_n = \sigma^2 I_{2 \times 2},$$

where

$$\sigma^2 = O(a_n^{-1}).$$

Then, the quantity in (4.17) becomes

$$-\sigma^2 \beta_2 [1 + (\beta_1 + 2\beta_2 x_t^0)^2]^{-1} = -\sigma^2 \beta_2 + \sigma^2 \beta_2 [1 + (\beta_1 + 2\beta_2 x_t^0)^2]^{-1} (\beta_1 + 2\beta_2 x_t^0)^2.$$

Observing that

$$m_{\text{xx}}^{-1} b_n^{-1} \sum_{t=1}^{b_n} (a_n \sigma_{\text{vnt}}^2)^{-1} (1, x_t^0, x_t^{02})' = (1, 0, 0)',$$

the bias expression in (4.16) becomes

$$(-\beta_2 \sigma^2, 0, 0)' + (\beta_2 \sigma^2) \mathbb{M}_{xx}^{*-1} \sum_{t=1}^{b_n} [1 + (\beta_1 + 2\beta_2 x_t^0)^2]^{-2} (\beta_1 + 2\beta_2 x_t^0)^2 [1, x_t^0, (x_t^0)^2]',$$

where

$$\mathbb{M}_{xx}^* = \sum_{t=1}^{b_n} [1 + (\beta_1 + \beta_2 x_t^0)^2]^{-1} [1, x_t^0, (x_t^0)^2]' [1, x_t^0, (x_t^0)^2].$$

We observe that at the point  $x_t^0 = -2^{-1} \beta_2^{-1} \beta_1$  the bias contribution is concentrated in the intercept term and is given by  $-\beta_2 \sigma^2$ . Also, as  $x_t^0$  moves away from  $-2^{-1} \beta_2^{-1} \beta_1$ , the bias contribution decreases for  $\beta_0$  and increases for  $\beta_1$  and  $\beta_2$ .

By Theorem 4.1, the bias in  $\hat{\beta}$  due to the curvature is given by  $\mathbb{M}_{xx}^{-1}(\delta_2 \hat{\beta})$ , which is a function of

$$\delta_{nt} f_{xxt}^0 \left( \frac{1}{2} \delta_{nt}' - u_{nt}' \right) = \frac{1}{2} (u_{nt} - \delta_{nt}) f_{xxt}^0 (u_{nt} - \delta_{nt})' - \frac{1}{2} u_{nt} f_{xxt}^0 u_{nt}'.$$

Since the leading term for the error in  $\bar{x}_t$  is  $\delta_{nt}$ ,

$$\bar{u}_{nt} = \bar{x}_{nt} - \bar{x}_t$$

$$= \tilde{u}_{nt} - (\tilde{x}_t - x_t^0)$$

estimates  $(\tilde{u}_{nt} - \delta_{nt})$ . Also,

$$E\{\tilde{u}_{nt} f_{xxt}^0 u'_{nt}\} = \text{tr}[f_{xxt}^0 \Sigma_{uun}] .$$

Hence, the modification in  $\tilde{\beta}$  suggested by Wolter and Fuller (1982b) is the subtraction of an estimate of the bias contribution from each

$$[\tilde{e}_{nt} - \tilde{u}_{nt} f'_x(\tilde{x}_t; \tilde{\beta})] .$$

Using  $\hat{\beta}$  we can obtain an estimator of  $x_t^0$  which is an improvement of the estimator  $\tilde{x}_t$ . Let  $\hat{x}_t$  be the value of  $x_t$  contained in  $\Gamma$  that minimizes  $q(\hat{\beta}, x_t; Y_{nt}, X_{nt})$ . Since

$$\hat{\beta} - \beta^0 = O_p(\max[a_n^{-1}, n^{-1/2}]) ,$$

Result 4.1 also holds for  $\hat{x}_t$ . Now, we investigate the bias in  $\hat{x}_t$  as an estimator of  $x_t^0$ . The effect of the bias in  $\hat{\beta}$  on  $\hat{x}_t$  is given in the next theorem.

Theorem 4.2. Let the model (4.1) hold, and let Assumptions 4.1 through 4.5, 4.6a, and 4.7a hold. Then,

$$\begin{aligned} \hat{x}_t - x_t^0 &= \delta_{nt} + [\delta_1 \hat{x}_t + \delta_2 \hat{x}_t + \delta_3 \hat{x}_t](a_n \Lambda_{nt}^{-1}) \\ &\quad + O_p(\max[a_n^{-3/2}, a_n^{-1/2} n^{-1/2}]) , \end{aligned}$$

where

$$\begin{aligned}\tilde{\delta}_{nt} &= \varepsilon_{nt} \Sigma_n^{-1} [\tilde{f}_{xt}^0, \tilde{I}]' \Lambda_{nt}^{-1} \\ &= O_p(a_n^{-1/2}),\end{aligned}$$

$$\begin{aligned}\delta_{1\tilde{x}_t} &= - [\tilde{f}_{\beta t}^0 m_{xx}^{-1}(\delta_1 \hat{\beta}), 0] (a_n \Sigma_n)^{-1} [\tilde{f}_{xt}^0, \tilde{I}]' \\ &= O_p(n^{-1/2}),\end{aligned}$$

$$\begin{aligned}\delta_{2\tilde{x}_t} &= \varepsilon_{nt}^* (a_n \Sigma_n)^{-1} (\delta_{nt}', 0)' \tilde{f}_{xxt}^0 \\ &= O_p(a_n^{-1}),\end{aligned}$$

$$\begin{aligned}\delta_{3\tilde{x}_t} &= - [\frac{1}{2} \delta_{nt} \tilde{f}_{xxt}^0 \delta_{nt}' + \tilde{f}_{\beta t}^0 m_{xx}^{-1}(\delta_2 \hat{\beta}), 0] (a_n \Sigma_n)^{-1} [\tilde{f}_{xt}^0, \tilde{I}]' \\ &= O_p(a_n^{-1}),\end{aligned}$$

$$\varepsilon_{nt}^* = \varepsilon_{nt} - \delta_{nt}(\tilde{f}_x^0, \tilde{I}).$$

Also,

$$E\{\tilde{\delta}_{nt}\} = E\{\delta_{1\tilde{x}_t}\} = E\{\delta_{2\tilde{x}_t}\} = 0,$$

$$\begin{aligned}
E\{\delta_{3\hat{x}_t}\} &= \frac{1}{2} \{f_{\beta t}^0, m_{xx}^{-1} n^{-1} \sum_{s=1}^{b_n} \sigma_{vns}^2 f_{\beta s}^0 \text{tr}[f_{xxs}^0 \Lambda_{ns}^{-1}] \\
&\quad - \text{tr}[f_{xxt}^0 \Lambda_{nt}^{-1}]\} a_n^{-1} [\sigma_n^{ee} f_{xt}^0 + \Sigma_n^{eu}] \\
&= o(a_n^{-1})
\end{aligned}$$

where

$$\Sigma_n^{-1} = \begin{bmatrix} \sigma_n^{ee} & \Sigma_n^{eu} \\ \Sigma_n^{ue} & \Sigma_n^{uu} \end{bmatrix} .$$

Proof. Since  $\hat{x}_t^0$  is an interior point of  $\Gamma$ , and since  $\hat{x}_t$  is consistent for  $\hat{x}_t^0$  by Result 4.1,  $\hat{x}_t$  satisfies

$$[Y_{nt} - f(\hat{x}_t; \hat{\beta}), X_{nt} - \hat{x}_t](a_n \Sigma_n)^{-1} [f'(\hat{x}_t; \hat{\beta}), I]' = o_p(\max[a_n^{-3/2}, n^{-1}]) . \quad (4.18)$$

By Result 4.1,

$$\hat{x}_t - \hat{x}_t^0 = \delta_{nt} + o_p(\max[a_n^{-1}, n^{-1/2}]) . \quad (4.19)$$

Using (4.19) and Theorem 4.1, we obtain that

$$\begin{aligned}
y_{nt} - f(\hat{x}_t; \hat{\beta}) &= e_{nt} - f_{xt}^0(\hat{x}_t - x_t^0)' - \frac{1}{2} \delta_{nt}' f_{xxt}^0 \delta_{nt}' \\
&\quad - f_{\beta t}^0 m_{xx}^{-1}(\delta_1 \hat{\beta} + \delta_2 \hat{\beta}) + o_p(\max[a_n^{-3/2}, a_n^{-1/2} n^{-1/2}]) \\
&= o_p(a_n^{-1/2}) , \tag{4.20}
\end{aligned}$$

and that

$$f_x(\hat{x}_t; \hat{\beta}) = f_{xt}^0 + \delta_{nt}' f_{xxt}^0 + o_p(\max[a_n^{-1}, n^{-1/2}]) . \tag{4.21}$$

Substituting (4.20) and (4.21) into (4.18), we have

$$\begin{aligned}
&[e_{nt} - f_{xt}^0(\hat{x}_t - x_t^0)' - \frac{1}{2} \delta_{nt}' f_{xxt}^0 \delta_{nt}' - f_{\beta t}^0 m_{xx}^{-1}(\delta_1 \hat{\beta} + \delta_2 \hat{\beta}), u_{nt}'(\hat{x}_t - x_t^0)] \\
&(a_n \Sigma_n)^{-1} [f_{xt}^0 + f_{xxt}^0 \delta_{nt}', I]' = o_p(\max[a_n^{-3/2}, a_n^{-1/2} n^{-1/2}]) .
\end{aligned}$$

Hence, the expansion of  $(\hat{x}_t - x_t^0)$  follows. We note that

$$\begin{aligned}
E\{\delta_{nt}\} &= 0 , \\
E\{\delta_1 \hat{\beta}\} &= 0 .
\end{aligned}$$

Also,

$$E\{\delta_{nt}' \varepsilon_{nt}^*\} = \Lambda_{nt}^{-1}(f_{xt}^0, I) [I - \Sigma_n^{-1}(f_{xt}^0, I)' \Lambda_{nt}^{-1}(f_{xt}^0, I)]$$

$$= 0 \quad ,$$

and thus, each element of  $\hat{\delta}_{nt}$  is uncorrelated with each element of  $\hat{\varepsilon}_{nt}^*$ . Therefore,  $E\{\hat{\delta}_{2\hat{x}_t}\} = 0$ . Finally, using Theorem 4.1 and taking the expectation of  $\hat{\delta}_{3\hat{x}_t}$ , we obtain the results.  $\square$

As in  $\hat{\beta}$ , there is a bias of  $O(a_n^{-1})$  in  $\hat{x}_t$  as an estimator of  $x_t^0$ . Also, the bias is again a function of the quantity in (4.17). But, in  $\hat{x}_t$ , the bias is due to the deviation of  $\text{tr}[f_{\hat{x}xt}^0 \Lambda_{nt}^{-1}]$  from the weighted average

$$\sum_{s=1}^b \sigma_{vns}^{-2} f_{\beta t}^0 \left\{ \sum_{r=1}^b \sigma_{vnr}^{-2} f_{\beta r}^0 f_{\beta r}^{0'} \right\}^{-1} f_{\beta s}^0 \text{tr}[f_{\hat{x}xs}^0 \Lambda_{ns}^{-1}] \quad . \quad (4.22)$$

Hence, the bias in  $\hat{x}_t$  is small for  $x_t^0$  such that  $\text{tr}[f_{\hat{x}xt}^0 \Lambda_{nt}^{-1}]$  is close to the weighted average (4.22). The term (4.22) is the contribution of the bias  $\hat{\delta}_{2\hat{\beta}}$  in  $\hat{\beta}$ . Thus, the bias of  $\hat{\beta}$  actually helps to reduce the bias of  $\hat{x}_t$  for the values of  $t$  for which  $\text{tr}[f_{\hat{x}xt}^0 \Lambda_{nt}^{-1}]$  is close to (4.22). To investigate this point in more detail, we consider the maximum likelihood estimator  $\hat{x}_t^*$  under the assumption that  $\beta^0$  is known. The estimator  $\hat{x}_t^*$  is the value of  $\hat{x}_t$  contained in  $\hat{\Gamma}$  that minimizes  $q(\beta^0, \hat{x}_t; Y_{nt}, X_{nt})$ . The asymptotic properties of  $\hat{x}_t^*$  are given in the following theorem.

Theorem 4.3. Let the model (4.1) hold, and let Assumptions 4.1, 4.2, 4.3, 4.6a, and 4.7a hold. Then,

$$\tilde{x}_t^* - \tilde{x}_t^0 = \hat{\delta}_{2\tilde{x}_t} + [\delta_{2\tilde{x}_t} + \delta_{\tilde{x}_t}^*](a_{n\tilde{nt}}^{-1}) + o_p(\max[a_n^{-3/2}, a_n^{-1/2}n^{-1/2}]) ,$$

where  $\hat{\delta}_{2\tilde{x}_t}$  is given in Theorem 4.2, and

$$\begin{aligned}\delta_{\tilde{x}_t}^* &= -\frac{1}{2} [\delta_{\tilde{nt}\tilde{xx}\tilde{nt}} f_{\tilde{xx}\tilde{nt}}^0 \delta_{\tilde{nt}}', 0] (a_{n\tilde{nn}} \Sigma_n)^{-1} [f_{\tilde{xt}}^0, \underline{I}]' \\ &= o_p(a_n^{-1}) .\end{aligned}$$

Also,

$$\begin{aligned}E\{\delta_{\tilde{x}_t}^*\} &= -\frac{1}{2} \text{tr}[f_{\tilde{xx}\tilde{nt}}^0 \Lambda_{\tilde{nt}}^{-1}] a_n^{-1} [\sigma_n^{ee} f_{\tilde{xx}}^0 + \Sigma_n^{eu}] \\ &= o(a_n^{-1}) .\end{aligned}$$

Proof. It is easy to show that

$$\begin{aligned}\tilde{x}_t^* - \tilde{x}_t^0 &= \hat{\delta}_{\tilde{nt}} + o_p(a_n^{-1}) \\ &= o_p(a_n^{-1/2}) .\end{aligned}$$

Then,  $\tilde{x}_t^*$  satisfies

$$[Y_{nt} - f(\tilde{x}_t^*; \beta), X_{nt} - \tilde{x}_t^*](a_{n\tilde{nn}} \Sigma_n)^{-1} [f_x'(\tilde{x}_t^*; \beta^0), \underline{I}]' = o_p(a_n^{-3/2}) . \quad (4.23)$$

Also,

$$Y_{nt} - f(\tilde{x}_t^*; \beta^0) = e_{nt} - \tilde{f}_{xt}^0 (\tilde{x}_t^* - \tilde{x}_t^0)' - \frac{1}{2} \delta_{nt} \tilde{f}_{xxt}^0 \delta_{nt}' + O_p(a_n^{-3/2}) , \quad (4.24)$$

and

$$\tilde{f}_x(\tilde{x}_t^*; \beta^0) = \tilde{f}_{xt}^0 + \delta_{nt} \tilde{f}_{xx}^0 + O_p(a_n^{-1}) . \quad (4.25)$$

Substituting (4.24) and (4.25) into (4.23), we obtain the results.  $\square$

Note that  $\tilde{x}_t^*$  was obtained by projecting  $(Y_{nt}, \tilde{X}_{nt})$  onto the known nonlinear surface  $y = f(\tilde{x}; \beta^0)$ . The bias given in Theorem 4.3 is due to the nonlinearity of the function  $f(\tilde{x}; \beta^0)$ . Comparing the results in Theorem 4.2 and Theorem 4.3, we obtain an interesting result. The estimator  $\hat{\tilde{x}}_t$ , the projection of  $(Y_{nt}, \tilde{X}_{nt})$  onto the estimated surface  $y = f(\tilde{x}; \hat{\beta})$ , has a smaller bias than the estimator  $\tilde{x}_t^*$ , the projection of  $(Y_{nt}, \tilde{X}_{nt})$  onto the true surface  $y = f(\tilde{x}; \beta^0)$ , provided that  $\text{tr}[\tilde{f}_{xxt}^0 \Lambda_{nt}^{-1}]$  has the same sign as the weighted average (4.22). That is, the bias in  $\hat{\beta}$  due to the curvature of  $f(\tilde{x}; \beta)$  partially cancels the bias due to the projection onto a nonlinear surface.

The iterative procedure for successively obtaining  $\hat{\beta}$  and  $\hat{\tilde{x}}_t$  leads to the maximum likelihood estimator. The asymptotic properties of estimators of  $\beta^0$  and  $\tilde{x}_t^0$  obtained by any finite number of iterations are given by Theorem 4.1 and Theorem 4.2. We conjecture that the same

asymptotic properties apply to the maximum likelihood estimator.

#### D. Bias of the Modified Maximum Likelihood Estimator

In this section, we investigate biases of the modified maximum likelihood estimator  $\tilde{\beta}$  and an estimator of  $x_t^0$  based on  $\tilde{\beta}$ , and compare the results to those in the previous section.

The following theorem is a restatement of Result 4.4.

Theorem 4.4. Let the model (4.1) hold, and let Assumptions 4.1 through 4.5, 4.6a, and 4.7a hold. Let  $\tilde{\beta}$  be defined by (4.6). Then,

$$\tilde{\beta} - \beta^0 = \underline{m}_{xx}^{-1}(\delta_1 \hat{\beta}) + O_p(\max[a_n^{-3/2}, a_n^{-1/2} n^{-1/2}]) ,$$

where

$$\delta_1 \hat{\beta} = n^{-1} \sum_{t=1}^{b_n} \sigma_{vnt}^{-2} \tilde{f}_{\beta t}^0 v_{nt}^* = O_p(n^{-1/2}) ,$$

$$v_{nt}^* = e_{nt} - \tilde{f}_{xt}^0 u_{nt} ,$$

$$E\{\delta_1 \hat{\beta}\} = 0 .$$

Proof. We observe that

$$\bar{v}_{nt} = e_{nt} + \tilde{f}_t^0 - f(x_{nt}; \tilde{\beta}) + \frac{1}{2} \text{tr}[\tilde{f}_{xxt}^0 \Sigma_{uun}] + O_p(a_n^{-3/2})$$

$$\begin{aligned}
&= \tilde{f}'_p(\tilde{x}_t; \tilde{\beta})(\beta^0 - \tilde{\beta}) + v_{nt}^* - \frac{1}{2} \text{tr}[\tilde{f}_{xxt}^0(u'_{nt}u_{nt} - \Sigma_{uun})] \\
&\quad + o_p(\max[a_n^{-3/2}, a_n^{-1/2}n^{-1/2}]) .
\end{aligned}$$

By Assumption 4.7a,

$$E\{|\text{vech}(u'_{nt}u_{nt} - \Sigma_{uun})|^2\} = o(a_n^{-2}) . \quad (4.26)$$

By (4.26) and Assumption 4.6a,

$$b_n^{-1} \sum_{t=1}^{b_n} \text{tr}[\tilde{f}_{xxt}^0(u'_{nt}u_{nt} - \Sigma_{uun})] = o_p(a_n^{-1/2}n^{-1/2}) . \quad (4.27)$$

The results follow by (4.27) and the argument in the proof of Theorem 4.1. □

Thus, the estimator  $\tilde{\beta}$  does not have the bias that is  $o(a_n^{-1})$  and is due to the curvature of  $f(x; \beta)$ . Therefore,  $\tilde{\beta}$  is superior to  $\hat{\beta}$  not only because the limiting distribution of  $n^{1/2}(\tilde{\beta} - \beta^0)$  can be obtained under weaker assumptions than those for  $n^{1/2}(\hat{\beta} - \beta^0)$ , but also because  $\tilde{\beta}$  has bias of smaller order.

We define the estimator  $\tilde{x}_t$  of  $x_t^0$  to be the value of  $x_t$  contained in  $\tilde{\Gamma}$  that minimizes  $q(\tilde{\beta}, x_t; Y_{nt}, X_{nt})$ . The next theorem presents the bias of  $\tilde{x}_t$ .

Theorem 4.5. Let the model (4.1) hold, and let Assumptions 4.1 through 4.5, 4.6a, and 4.7a hold. Then,

$$\tilde{x}_t - x_t^0 = \delta_{nt} + [\delta_1 \hat{x}_t + \delta_2 \hat{x}_t + \delta x_t^*] (a_n^{-1}) + o_p(\max(a_n^{-3/2}, a_n^{-1/2} n^{-1/2})) ,$$

where  $\delta_1 \hat{x}_t$  and  $\delta_2 \hat{x}_t$  are given in Theorem 4.2, and  $\delta x_t^*$  is given in Theorem 4.3.

Proof. Since

$$\tilde{\beta} - \beta^0 = o_p(\max[a_n^{-3/2}, n^{-1/2}]) ,$$

Result 4.1 holds for  $\tilde{x}_t$ . Thus,

$$\tilde{x}_t - x_t^0 = \delta_{nt} + o_p(\max[a_n^{-1}, n^{-1/2}]) ,$$

$$[Y_{nt} - f(\tilde{x}_t; \tilde{\beta}), \tilde{x}_{nt} - \tilde{x}_t] (a_n \Sigma_n)^{-1} [f'_x(\tilde{x}_t; \tilde{\beta}), I]' = o_p(\max[a_n^{-3/2}, n^{-1}]) .$$

Also,

$$\begin{aligned} Y_{nt} - f(\tilde{x}_t; \tilde{\beta}) &= e_{nt} - f_{xt}^0 (\tilde{x}_t - x_t^0)' - \frac{1}{2} \delta_{nt} f_{xxt}^0 \delta_{nt}' \\ &\quad - f_{\beta t}^0 \Sigma_n^{-1} (\delta_1 \hat{\beta}) + o_p(\max[a_n^{-3/2}, a_n^{-1/2} n^{-1/2}]) , \end{aligned}$$

$$f_x(\tilde{x}_t; \tilde{\beta}) = f_{xt}^0 + \delta_{nt} f_{xxt}^0 + o_p(\max[a_n^{-1}, n^{-1/2}]) .$$

Hence, the results follow.  $\square$

Taking the expectation through terms of  $O_p(a_n^{-1})$ , the bias of  $\tilde{x}_t$  is the same as that of  $x_t^*$ , the estimator assuming known  $\beta^0$ . Since  $\tilde{\beta}$  does not possess the bias due to the curvature, the cancellation of the nonlinearity biases in  $\hat{x}_t$  does not occur for  $\tilde{x}_t$ . Therefore, for certain values of  $t$  the estimator  $\tilde{x}_t$  based on  $\tilde{\beta}$  has a larger bias than the estimator  $\hat{x}_t$  based on  $\hat{\beta}$ .

#### E. A Class of Estimators Adjusted for Bias

In this section, we introduce a class of estimators  $(\tilde{\beta}_A, \tilde{x}_{tA})$  of  $(\beta^0, x^0)$  for the model (4.1). For each member of the class,  $\tilde{\beta}_A$  has the same asymptotic properties as the modified maximum likelihood estimator  $\tilde{\beta}$ . That is, each  $\tilde{\beta}_A$  in the class has no bias due to the curvature, and has a limiting distribution under Assumption 4.8a. Also, we can choose certain members of the class for which the  $\tilde{x}_{tA}$  have smaller biases than the estimators  $\hat{x}_t$ ,  $\tilde{x}_t$ , and  $x_t^*$ .

To define the class of estimators, we assume that there are available preliminary estimators  $\bar{\beta}$  satisfying Assumption 4.5 and  $\bar{x}_t$  satisfying the conclusions of Result 4.1. For example, the maximum likelihood estimators of  $\beta^0$  and  $x_t^0$ , i.e., the estimators  $\hat{\beta}$  and  $\hat{x}_t$  iterated to convergence, can be used as preliminary estimators. Let  $dY_{nt}$  and  $d\tilde{x}_{nt}$ ,  $t = 1, 2, \dots, b_n$ , be functions of  $\bar{\beta}$  and  $\bar{x}_t$ , where  $dY_{nt}$  is a scalar and  $d\tilde{x}_{nt}$  is a  $1 \times q$  vector. The quantities  $dY_{nt}$  and  $d\tilde{x}_{nt}$  are modifications to observations  $Y_{nt}$  and  $X_{nt}$  and are introduced to reduce the bias due to the nonlinearity. Given  $dY_{nt}$

and  $d\tilde{x}_{nt}$ , we define the adjusted estimators  $\tilde{\beta}_A$  and  $\tilde{x}_{tA}$  to be the maximum likelihood estimators of  $\beta^0$  and  $x_t^0$  based on the observations  $(Y_{nt}^*, X_{nt}^*)$ , where

$$Y_{nt}^* = Y_{nt} + dY_{nt},$$

$$X_{nt}^* = X_{nt} + dX_{nt}. \quad (4.28)$$

That is,  $\tilde{\beta}_A$  and  $\tilde{x}_{tA}$  are the estimators  $\hat{\beta}$  and  $\hat{x}_t$  iterated to convergence using  $(Y_{nt}^*, X_{nt}^*)$  as observations. Depending on the choice of  $dY_{nt}$  and  $dX_{nt}$ , we have different adjusted estimators  $\tilde{\beta}_A$  and  $\tilde{x}_{tA}$ . Of course, the adjustment term  $(dY_{nt}, dX_{nt})$  has to satisfy some condition to reduce the biases of the resulting adjusted estimators  $\tilde{\beta}_A$  and  $\tilde{x}_{tA}$ . The following Assumption 4.9 provides a condition for  $(dY_{nt}, dX_{nt})$  to produce an adjusted estimator  $\tilde{\beta}_A$  with a small bias. We later introduce Assumption 4.10 which provides a condition for  $(dY_{nt}, dX_{nt})$  to produce an adjusted estimator  $\tilde{x}_{tA}$  with a small bias. We define the class of adjusted estimators to be estimators  $\tilde{\beta}_A$  and  $\tilde{x}_{tA}$  which are obtained using  $(dY_{nt}, dX_{nt})$  satisfying Assumptions 4.9 and 4.10. Specific examples of the adjustment term  $(dY_{nt}, dX_{nt})$  are given in Section F. We first investigate theoretical properties.

Assumption 4.9. There exists a constant  $K_2$  such that for all  $t = 1, 2, \dots, b_n$ , and all  $n$ ,

$$E\{ |(dy_{nt}, dx_{nt})|^4 \} < K_2 a_n^{-4} .$$

Also,

$$\begin{aligned} n^{-1} \sum_{t=1}^{b_n} \sigma_{vnt}^{-2} f_{\beta t}^0 \{ dy_{nt} - dx_{nt} f_{xt}^0 - \frac{1}{2} \text{tr}[f_{xxt}^0 \Lambda_{nt}^{-1}] \} \\ = O_p(\max[a_n^{-3/2}, n^{-1/2} a_n^{-1/2}]) . \end{aligned} \quad (4.29)$$

Since under Assumptions 4.4, 4.6a, and 4.7a,

$$\begin{aligned} n^{-1} \sum_{t=1}^{b_n} \sigma_{vnt}^{-2} f_{\beta t}^0 \{ \delta_{nt} f_{xxt}^0 (\frac{1}{2} \delta_{nt} - u_{nt})' + \frac{1}{2} \text{tr}[f_{xxt}^0 \Lambda_{nt}^{-1}] \} \\ = O_p(\max[a_n^{-3/2}, n^{-1/2} a_n^{-1/2}]) , \end{aligned}$$

the condition (4.29) is equivalent to the condition

$$n^{-1} \sum_{t=1}^{b_n} \sigma_{vnt}^{-2} f_{\beta t}^0 \{ dy_{nt} - dx_{nt} f_{xt}^0 + \delta_{nt} f_{xxt}^0 (\frac{1}{2} \delta_{nt} - u_{nt})' \}$$

$$= O_p(\max[a_n^{-3/2}, n^{-1/2} a_n^{-1/2}]) \quad (4.30)$$

Also, the condition (4.29) implies that  $(dY_{nt} - d\tilde{x}_{nt} f_{xt}^{0'})$  is estimating the bias contribution given in (4.17).

To investigate the asymptotic properties of  $\tilde{\beta}_A$ , we consider the following one-step estimator  $\hat{\beta}_A$  using a preliminary estimator  $\bar{\beta}$ .

Let  $\bar{x}_{tA}$  be the value of  $x_t$  contained in  $I$  that minimizes  $q(\bar{\beta}, x_t; Y_{nt}^*, \tilde{x}_{nt}^*)$ , and let

$$\hat{\beta}_A = \bar{\beta} + (\Delta \hat{\beta}_A),$$

where  $\Delta \hat{\beta}_A$  satisfies

$$\hat{M}_{xx}(\Delta \hat{\beta}_A) = n^{-1} \sum_{t=1}^{b_n} \sigma_{vnt}^{-2} f_{\beta}(\bar{x}_{tA}; \bar{\beta}) [\hat{e}_{nt} - \hat{u}_{nt} f'_x(\bar{x}_{tA}; \bar{\beta})],$$

$$\hat{\sigma}_{vnt}^2 = [1, -f_x(\bar{x}_{tA}; \bar{\beta})] \Sigma_n [1, -f_x(\bar{x}_{tA}; \bar{\beta})]',$$

$$\hat{M}_{xx} = n^{-1} \sum_{t=1}^{b_n} \sigma_{vnt}^{-2} f_{\beta}(\bar{x}_{tA}; \bar{\beta}) f'_{\beta}(\bar{x}_{tA}; \bar{\beta}),$$

$$\hat{e}_{nt} = Y_{nt}^* - f(\bar{x}_{tA}; \bar{\beta}),$$

$$\hat{u}_{nt} = \bar{x}_{nt}^* - \bar{x}_{tA}$$

The properties of the one step estimator  $\hat{\beta}_A$  are given in the following theorem.

Theorem 4.6. Let the model (4.1) hold, and let Assumptions 4.1 through 4.5, 4.6a, 4.7a, and 4.9 hold. Then,

$$\hat{\beta}_A - \beta^0 = \bar{m}_{xx}^{-1} (\delta_1 \hat{\beta}) + o_p(\max[a_n^{-3/2}, a_n^{-1/2} n^{-1/2}])$$

If, in addition, Assumption 4.8a holds, then

$$n^{1/2} (\hat{\beta}_A - \beta^0) \xrightarrow{L} N(0, \bar{m}_{xx}^{-1})$$

Proof. By Assumption 4.9,

$$(dY_{nt}, dX_{nt}) = o_p(a_n^{-1})$$

Thus, it follows that

$$\bar{x}_{tA} - x_t^0 = \delta_{nt} + o_p(\max[a_n^{-1}, n^{-1/2}]) \quad (4.31)$$

We observe that

$$q(\bar{\beta}, \bar{x}_{tA}; Y_{nt}^*, X_{nt}^*) < q(\bar{\beta}, x_t^0; Y_{nt}^*, X_{nt}^*) \quad (4.32)$$

By the compactness of  $\Omega$  and Assumptions 4.6a, 4.7a, and 4.9, it can be shown that for a constant  $K_3$  and for  $t = 1, 2, \dots, b_n$ ,

$$E\{q^2(\bar{\beta}, \bar{x}_t^0; y_{nt}^*, \bar{x}_{nt}^*)\} \leq K_3. \quad (4.33)$$

It follows from (4.32) and (4.33) that for a constant  $K_4$  and for  $t = 1, 2, \dots, b_n$ ,

$$E\{|\bar{x}_{nt}^* - \bar{x}_{tA}|^4\} \leq K_4 a_n^{-2}. \quad (4.34)$$

By (4.34) and Assumptions 4.7a and 4.9, for a constant  $K_5$  and for  $t = 1, 2, \dots, b_n$ ,

$$E\{|\bar{x}_{tA} - \bar{x}_t^0|^4\} \leq K_5 a_n^{-2}. \quad (4.35)$$

By (4.35) and Assumptions 4.3 and 4.6a, it can be shown that

$$b_n^{-1} \sum_{t=1}^{b_n} [(\hat{a}_n \sigma_{vnt}^2)^{-1} - (a_n \sigma_{vnt}^2)^{-1} + (a_n \sigma_{vnt}^2)^{-2} (a_n \Delta_{vnt})]^2 = o_p(a_n^{-2}). \quad (4.36)$$

The expansion (4.35) implies that the order of an average over  $b_n$  values of  $(\bar{x}_{tM} - \bar{x}_t^0)^i$  is of  $o_p(a_n^{-i/2})$  for  $i \leq 4$ . Thus, using this fact, Assumption 4.6a, and (4.36), we obtain

$$\hat{\tilde{M}}_{xx} = \tilde{m}_{xx} + o_p(\max[a_n^{-1}, n^{-1/2}]) \quad (4.37)$$

We observe that

$$\begin{aligned} \hat{e}_{nt} - \hat{\tilde{u}}_{nt} f'_x(\bar{x}_{tA}; \bar{\beta}) \\ = f_{\beta t}^0(\beta^0 - \bar{\beta}) + v_{nt}^* + dY_{nt} - dX_{nt} f_{xt}^0 + \delta_{nt} f_{xxt}^0 \left( \frac{1}{2} \delta'_{nt} - u'_{nt} \right) \\ + o_p(\max[a_n^{-3/2}, n^{-1/2} a_n^{-1/2}]) \quad (4.38) \end{aligned}$$

where the remainder is a function of the powers of  $(\bar{x}_{tA} - x_t^0)$ ,  $(\bar{\beta} - \beta^0)$ , and  $dX_{nt}$ . Thus, it follows from (4.14), (4.15), (4.30), (4.35), (4.36), (4.38), and Assumptions 4.5, 4.6a, and 4.9 that

$$\begin{aligned} n^{-i} \sum_{t=1}^{b_n} \hat{\sigma}_{vnt}^{-2} \tilde{f}_{\beta}(\bar{x}_{tA}; \bar{\beta}) [\hat{e}_{nt} - \hat{\tilde{u}}_{nt} f'_x(\bar{x}_{tA}; \bar{\beta})] \\ = n^{-1} \sum_{t=1}^{b_n} \hat{\sigma}_{vnt}^{-2} \tilde{f}_{\beta}(\bar{x}_{tA}; \bar{\beta}) f_{\beta t}^0(\beta^0 - \bar{\beta}) + (\delta_1 \hat{\beta}) \\ + o_p(\max[a_n^{-3/2}, n^{-1/2} a_n^{-1/2}]) \quad (4.39) \end{aligned}$$

Therefore, the result for the expansion of  $(\hat{\beta}_A - \beta^0)$  follows from (4.37) and (4.39). The limiting distribution result holds by Result 4.4 □

To consider the asymptotic properties of the adjusted estimator  $\tilde{x}_{tA}$ , we define  $\hat{x}_{tA}$  to be the value of  $x_t$  contained in  $\mathcal{I}$  that minimizes  $q(\hat{\beta}_A, x_t; Y_{nt}^*, X_{nt}^*)$ . We introduce a further assumption on  $(dY_{nt}, dX_{nt})$ .

Assumption 4.10. There exist  $\Delta Y_{nt}$  and  $\Delta X_{nt}$  such that

$$dY_{nt} = \Delta Y_{nt} + o_p(\max[a_n^{-3/2}, n^{-1/2} a_n^{-1/2}]),$$

$$dX_{nt} = \Delta X_{nt} + o_p(\max[a_n^{-3/2}, n^{-1/2} a_n^{-1/2}]),$$

$$E\{\Delta Y_{nt}\} = \frac{1}{2} \text{tr}[f_{xxt}^0 \Lambda^{-1}] + E\{\Delta X_{nt}\} f_{xt}^0.$$

The next theorem gives the properties of  $\hat{x}_{tA}$ .

Theorem 4.7. Let the model (4.1) hold, and let Assumptions 4.1 through 4.5, 4.6a, 4.7a, 4.9, and 4.10 hold. Then,

$$\hat{x}_{tA} - x_t^0 = \hat{\Delta x}_{tA} + o_p(\max[a_n^{-3/2}, a_n^{-1/2} n^{-1/2}]),$$

where

$$\begin{aligned}\hat{\Delta}_{\tilde{x}_{tA}}^x &= \delta_{nt} + [\delta_1 \hat{x}_t + \delta_2 \hat{x}_t + \delta_{\tilde{x}_t}^*](a_n \Lambda_{nt}^{-1}) \\ &\quad + (\Delta Y_{nt}, \Delta \tilde{x}_{nt})(a_n \Sigma_n)^{-1} [\tilde{f}_{xt}^0, \tilde{I}]' (a_n \Lambda_{nt}^{-1}) .\end{aligned}$$

Also,

$$E\{\hat{\Delta}_{\tilde{x}_{tA}}^x\} = E\{\Delta \tilde{x}_{nt}\} .$$

Proof. We observe that the modification (4.28) is of  $O_p(a_n^{-1})$ , and that

$$\hat{\beta}_A - \beta^0 = O_p(\max[a_n^{-3/2}, n^{-1/2}]) .$$

Thus, by Result 4.1,

$$\hat{x}_{tA} - x_t^0 = \delta_{nt} + O_p(\max[a_n^{-1}, n^{-1/2}]) . \quad (4.40)$$

Using (4.40) and Theorem 4.6, we obtain that

$$\begin{aligned}y_{nt}^* - f(\hat{x}_{tA}; \hat{\beta}_M) &= e_{nt} + \Delta Y_{nt} - \tilde{f}_{xt}^0 (\hat{x}_{tA} - x_t^0)' - \frac{1}{2} \delta_{nt} \tilde{f}_{xx}^0 \delta_{nt}' \\ &\quad - \tilde{f}_{\beta t}^0 \tilde{m}^{-1} \delta_1 \hat{\beta} + O_p(\max[a_n^{-3/2}, n^{-1/2} a_n^{-1/2}]) ,\end{aligned} \quad (4.41)$$

and that

$$\tilde{f}_x(\hat{x}_{tA}; \hat{\beta}_A) = \tilde{f}_{xt}^0 + \delta_{nt} \tilde{f}_{xxt}^0 + o_p(\max[a_n^{-1}, n^{-1/2}]) \quad (4.42)$$

It follows from (4.41) and (4.42) that

$$[e_{nt} + \Delta Y_{nt} \tilde{f}_{xt}^0 (\hat{x}_{tA} - x_t^0)' - \frac{1}{2} \delta_{nt} \tilde{f}_{xxt}^0 \delta_{nt}' \tilde{f}_{\beta t}^0 m^{-1} \delta_{nt} \hat{\beta}, \tilde{u}_{nt} + \Delta X_{nt} - (x_{tA} - x_t^0)]$$

$$(a_n \Sigma_n)^{-1} [\tilde{f}_{xt}^0 + \tilde{f}_{xxt}^0 \delta_{nt}', \tilde{I}] = o_p(\max[a_n^{-3/2}, a_n^{-1/2} n^{-1/2}])$$

Therefore, the expansion for  $(\hat{x}_{tA} - x_t^0)$  follows. If Assumption 4.10 holds, then

$$\begin{aligned} & E\{(\Delta Y_{nt}, \Delta X_{nt}) \Sigma_n^{-1} [\tilde{f}_{xt}^0, \tilde{I}]' \Lambda_{nt}^{-1}\} \\ &= \left(\frac{1}{2} \text{tr}[\tilde{f}_{xxt}^0 \Lambda_{nt}^{-1}], 0\right) \Sigma_n^{-1} [\tilde{f}_{xt}^0, \tilde{I}]' \Lambda_{nt}^{-1} + E\{\Delta X_{nt}\} \end{aligned}$$

Also, by Theorem 4.2 and Theorem 4.3,

$$E\{\delta_{nt}\} = E\{\delta_{1\tilde{x}_t}\} = E\{\delta_{2\tilde{x}_t}\} = 0,$$

$$E\{\delta_{\tilde{x}_t}^*\} = -\left(\frac{1}{2} \text{tr}[\tilde{f}_{xxt}^0 \Lambda_{nt}^{-1}], 0\right) (a_n \Sigma_n)^{-1} [\tilde{f}_{xt}^0, \tilde{I}]$$

Thus, under Assumption 4.10,

$$E\{\Delta \hat{x}_{tA}\} = E\{\Delta X_{nt}\} \quad \square$$

The iterative procedure for successively obtaining  $\hat{\beta}_A$  and  $\hat{x}_{tA}$  leads to the adjusted estimators  $\tilde{\beta}_A$  and  $\tilde{x}_{tA}$ . The estimators of  $\beta^0$  and  $x_t^0$  obtained by any finite number of such iterations have asymptotic properties given in Theorem 4.6 and Theorem 4.7. We might expect that the same asymptotic properties would apply to the adjusted estimators  $\tilde{\beta}_A$  and  $\tilde{x}_{tA}$  which maximize the likelihood function of  $(Y_{nt}^*, X_{nt}^*)$ . But, we emphasize that this is an unproven conjecture.

The adjusted estimators  $\tilde{\beta}_A$  and  $\tilde{x}_{tA}$  can be improved by updating the function  $(dY_{nt}, dX_{nt})$  at each stage of iteration based on the most recent estimates of  $\beta^0$  and  $x_t^0$ , or using  $\tilde{\beta}_A$  and  $\tilde{x}_{tA}$  to estimate a new adjustment term and repeating the maximization of the likelihood. Because of the computational simplicity we chose to define our class of adjusted estimators to be the maximum likelihood estimator based on the observations with fixed modifications that do not vary by iteration. If the maximum likelihood estimators without adjustment are used as the preliminary estimators  $\bar{\beta}$  and  $\bar{x}$ , then two runs of a standard nonlinear regression algorithm can produce any member of our class of adjusted estimators.

#### F. Examples of the Adjusted Estimators

As we have seen in the previous section, any adjusted estimator  $\tilde{\beta}_A$  satisfying Assumption 4.9 has no bias due to the curvature and has the same asymptotic properties as the estimator  $\tilde{\beta}$ . Also, certain choices of  $dY_{nt}$  and  $dX_{nt}$  reduce errors in the adjusted estimator

$\tilde{x}_{tA}$ . In this section, we present three examples of the adjusted estimators, and make comparisons among them. In Section G, we present the results from a Monte Carlo experiment carried out to study the small sample properties of various estimators of  $\beta^0$  including three estimators introduced in this section.

For the three adjusted estimators considered in this section, we use the maximum likelihood estimators  $\hat{\beta}_M$  and  $\hat{x}_{tM}$  as the preliminary estimators  $\tilde{\beta}$  and  $\tilde{x}_t$ . In this section, we act as if  $\hat{\beta}_M, \hat{x}_{tM}$ , and the three adjusted estimators have the same asymptotic properties as their corresponding one step estimators.

#### 1. Adjustment based on the local quadratic approximation

A simple adjustment satisfying Assumptions 4.9 and 4.10 is obtained by setting

$$d_1 Y_{nt} = \frac{1}{2} \text{tr}[\tilde{f}_{xx}(\tilde{x}_{tM}; \hat{\beta}_M) \hat{\Lambda}_{nt}^{-1}] ,$$

$$d_1 \tilde{x}_{nt} = 0 ,$$

where

$$\hat{\Lambda}_{nt}^{-1} = \Sigma_{uun} - \hat{\Sigma}_{uvnt} \hat{\sigma}_{vnt}^{-2} \hat{\Sigma}_{vunt} ,$$

$$\hat{\Sigma}_{uvnt} = \hat{\Sigma}'_{vunt} = \Sigma_{uen} - \Sigma_{uun} \tilde{f}'_x(\tilde{x}_{tM}; \hat{\beta}_M) ,$$

$$\hat{\sigma}_{vnt}^2 = [1, -\tilde{f}_x(\tilde{x}_{tM}; \hat{\beta}_M)] \Sigma_n [1, -\tilde{f}_x(\tilde{x}_{tM}; \hat{\beta}_M)]' .$$

Denote the resulting adjusted estimators by  $\tilde{\beta}_1$  and  $\tilde{x}_{t1}$ . Since  $\hat{x}_{tM}$  and  $\hat{\beta}_M$  are the maximum likelihood estimators, under Assumptions 4.1 through 4.5, 4.6a, and 4.7a,  $(d_1 Y_{nt}, d_1 X_{nt})$  satisfies Assumptions 4.9 and 4.10 with

$$\Delta_1 Y_{nt} = \frac{1}{2} \text{tr}[\tilde{f}_{xxt}^0 \Lambda_{nt}^{-1}] ,$$

$$\Delta_1 X_{nt} = 0 .$$

Thus, the results of Theorem 4.6 and Theorem 4.7 hold for  $\tilde{\beta}_1$  and  $\tilde{x}_{t1}$ . Since  $\Delta_1 X_{nt} = 0$ , the estimator  $\tilde{x}_{t1}$  is unbiased for  $x_t^0$  through terms of  $O_p(a_n^{-1})$ . One interpretation of the adjusted estimators  $\tilde{\beta}_1$  and  $\tilde{x}_{t1}$  can be obtained through the following observation.

Suppose the model (4.1) is quadratic in  $x_t^0$  and is given by

$$y_t^0 = \theta_0 + x_t^0 \theta_1 + \frac{1}{2} x_t^0 \theta_2 x_t^0 , \quad (4.43)$$

where  $\theta_2$  is a symmetric  $q \times q$  matrix. The model (4.43) is linear in the parameters, and we can write

$$y_t^0 = z_t^0 \beta^0 ,$$

where

$$z_t^0 = (1, x_t^0, \frac{1}{2} [x_t^0 \otimes x_t^0] \phi_q) ,$$

$$\beta^0 = (\theta_0, \theta_1', [\text{vech } \theta_2]') .$$

We observe that

$$f_{xt}^0 = \theta_1' + x_t^0 \theta_2 ,$$

$$f_{xxt}^0 = \theta_2 ,$$

$$f_{\beta t}^0 = z_t^{0'} .$$

Thus, the term  $\hat{m}_{xx}^{-1}(\delta_2 \hat{\beta})$  in the expansion of the maximum likelihood estimator  $\hat{\beta}_M$  of  $\beta^0$  for this model becomes

$$\hat{m}_{xx}^{-1} \delta_2 \hat{\beta} = \left\{ \sum_{t=1}^n \sigma_{vnt}^{-2} z_t^0 z_t^{0'} \right\}^{-1} \sum_{t=1}^n \sigma_{vnt}^{-2} z_t^0 \delta_{nt} \theta_2 \left( \frac{1}{2} \delta_{nt}' - u_{nt}' \right) , \quad (4.44)$$

where

$$\sigma_{vnt}^2 = [1, -\theta_1' - x_t^0 \theta_2] \Sigma_n [1, -\theta_1' - x_t^0 \theta_2]'$$

$$\delta_{nt} = \varepsilon_{nt} \Sigma_n^{-1} [\theta_1 + \theta_2 x_t^{0'}, I] \Lambda_{nt}^{-1}$$

$$\Lambda_{nt}^{-1} = \Sigma_{uun} - \Sigma_{uvn} \sigma_{vnt}^{-2} \Sigma_{vun} ,$$

$$\Sigma_{uvn} = \Sigma'_{vun} = \Sigma_{eun} - \Sigma_{uun} (\theta_1 + \theta_2 \tilde{x}_t^0)' .$$

Also,

$$E\{\tilde{m}_{xx}^{-1} \delta_2 \hat{\beta}\} = -\frac{1}{2} \left\{ \sum_{t=1}^{b_n} \sigma_{vnt}^{-2} z_t^0 z_t^{0'} \right\}^{-1} \sum_{t=1}^{b_n} \sigma_{vnt}^{-2} z_t^0 \text{tr}[\theta_2 \Lambda_{nt}^{-1}] . \quad (4.45)$$

For a general nonlinear model (4.1), we approximate the function  $f(\tilde{x}; \beta^0)$  around  $\tilde{x} = \tilde{x}_s^0$  by the first three terms of the Taylor expansion

$$f(\tilde{x}; \beta^0) \doteq f_s^0 + (\tilde{x} - \tilde{x}_s^0) f_{xs}^{0'} + \frac{1}{2} (\tilde{x} - \tilde{x}_s^0) f_{xxs}^0 (\tilde{x} - \tilde{x}_s^0)' . \quad (4.46)$$

Suppose that we estimate  $f_s^0$ ,  $f_{xs}^0$ , and  $\text{vech } f_{xxs}^0$  instead of  $\beta^0$  based on the observations on  $\tilde{x}_t^0$ ,  $t = 1, 2, \dots, b_n$ , which are close to  $\tilde{x}_s^0$ . Then, the bias of the maximum likelihood estimator is given by (4.45) with  $\tilde{f}_s^0$ ,  $\tilde{f}_{xs}^{0'}$ , and  $\tilde{f}_{xxs}^0$  replacing  $\theta_0$ ,  $\theta_1$ , and  $\theta_2$ , respectively. If  $\Lambda_{nt}$  is approximated by  $\Lambda_{ns}$  for  $\tilde{x}_t^0$  close to  $\tilde{x}_s^0$ , then the bias becomes

$$-\frac{1}{2} \text{tr}[\tilde{f}_{xxs}^0 \Lambda_{ns}^{-1}] (1, 0, 0)' , \quad (4.47)$$

where we have used

$$\left\{ \sum_{t=1}^b \sigma_{vnt}^{-2} \begin{bmatrix} z_t^0 & z_t^0 \end{bmatrix} \right\}^{-1} \left\{ \sum_{t=1}^b \sigma_{vnt}^{-2} z_t^0 \right\} = (1, 0, 0)' . \quad (4.48)$$

Thus, the maximum likelihood estimators for  $f_{xs}^0$  and  $\text{vech } f_{xxs}^0$  are unbiased, while the maximum likelihood estimator of the intercept term has the bias

$$- \frac{1}{2} \text{tr} [f_{xxs}^0 \hat{\Lambda}_{ns}^{-1}] . \quad (4.49)$$

Hence, the bias may be removed by subtracting (4.49) from  $Y_{nt}$  corresponding to  $\tilde{x}_t^0$  close to  $\tilde{x}_s^0$ . If we apply the same argument to each  $\tilde{x}_s^0$ , and if we estimate (4.48) by a natural estimator

$$- \frac{1}{2} \text{tr} [f_{xx}(\hat{x}_{sM}; \hat{\beta}_M) \hat{\Lambda}_{ns}^{-1}] ,$$

then we arrive at the adjustment given by  $(d_1 Y_{ns}, d_1 \tilde{x}_{ns})$ . Therefore, the estimators  $\tilde{\beta}_1$  and  $\tilde{x}_{t1}$  are derived based on the bias results for the local quadratic approximation to the function  $f(\tilde{x}; \beta^0)$ .

## 2. Adjustment based on the curvature relative to the tangent

The bias result (4.45) for the quadratic model (4.43) motivates another type of adjustment. Assume that

$$\Sigma_n = \sigma^2 I , \quad (4.50)$$

where  $\sigma^2 = O(a_n^{-1})$ . Then, using (4.48), we rewrite (4.45) to obtain

$$E\{\mathbf{m}_{xx}^{-1} \delta_2 \hat{\beta}\} = -\frac{1}{2} \sigma^2 \text{tr}[\theta_2](1, \underline{0}, \underline{0})' + \frac{1}{2} \sigma^2 \left\{ \sum_{t=1}^{b_n} \sigma_{vnt}^2 \mathbf{z}_t^0 \mathbf{z}_t^{0'} \right\}^{-1}$$

$$\sum_{t=1}^{b_n} \sigma_{vnt}^{-4} \mathbf{z}_t^0 \text{tr}[(\theta_1' + \mathbf{x}_t^0 \theta_2) \theta_2 (\theta_1 + \theta_2 \mathbf{x}_t^0)] .$$

Thus, the bias contribution from the point  $\mathbf{x}_t^0 = -\theta_1' \theta_2^{-1}$  is  $(-\frac{1}{2} \sigma^2 \text{tr}[\theta_2], \underline{0}, \underline{0})'$ . This suggests the following adjustment. First, for each  $\mathbf{x}_t^0$ , we rotate the coordinate system from  $(y, \mathbf{x})$  to  $(z, \mathbf{w})$  so that the function  $z = y(\mathbf{w}; \beta^0)$  transformed from  $y = f(\mathbf{x}; \beta^0)$  satisfies  $g_{\mathbf{w}}(\mathbf{w}_t^0; \beta^0) = \underline{0}$ , where  $(\mathbf{z}_t^0, \mathbf{w}_t^0)$  is the rotated  $(y_t^0, \mathbf{x}_t^0)$ . Then, the adjustment in the new coordinate system is obtained by adding  $\frac{1}{2} \sigma^2 \text{tr}[g_{\mathbf{ww}}(\mathbf{w}_t^0; \beta^0)]$  to the observation on  $\mathbf{z}_t^0$ . Then, we transform back to the original axes to obtain an adjustment to  $(y_{nt}, \mathbf{x}_{nt})$ . Note that  $g_{\mathbf{ww}}(\mathbf{w}_t^0; \beta^0)$  is the curvature of the function  $f(\mathbf{x}; \beta^0)$  relative to the tangent at  $\mathbf{x} = \mathbf{x}_t^0$ .

The desired rotation of axes can be obtained by an orthogonal transformation

$$(y, \mathbf{x}) = (z, \mathbf{w}) \mathbf{P} , \quad (4.51)$$

where

$$\underline{P} = \begin{bmatrix} P_{11} & P_{12} \\ P_{21} & P_{22} \end{bmatrix},$$

$$P_{11} = (1 + \underline{f}_{xt}^0 \underline{f}_{xt}^{0'})^{-1/2},$$

$$P_{12} = - (1 + \underline{f}_{xt}^0 \underline{f}_{xt}^{0'})^{-1/2} \underline{f}_{xt}^0,$$

$$P_{21} = (\underline{I} + \underline{f}_{xt}^{0'} \underline{f}_{xt}^0)^{-1/2} \underline{f}_{xt}^{0'},$$

$$P_{22} = (\underline{I} + \underline{f}_{xt}^{0'} \underline{f}_{xt}^0)^{-1/2},$$

Since  $y = f(\underline{x}; \underline{\beta}^0)$ , we have

$$F(z, \underline{w}) = z P_{11} + \underline{w} P_{21} - f(z P_{12} + \underline{w} P_{22}; \underline{\beta}^0) = 0. \quad (4.52)$$

We observe that

$$\begin{aligned} F_z(z, \underline{w}) &= \frac{\partial}{\partial z} F(z, \underline{w}) \\ &= P_{11} - \underline{f}_x(z P_{12} + \underline{w} P_{22}; \underline{\beta}^0) P'_{12}, \end{aligned}$$

$$F_w(z, \underline{w}) = \frac{\partial}{\partial \underline{w}} F(z, \underline{w})$$

$$= P'_{21} - f_x(z P_{12} + w P_{22}; \beta^0) P_{22} .$$

Thus, by (4.52) and the implicit function theorem, for  $(z, w)$  close to  $(z_t^0, w_t^0)$ , there exists a function  $z = g(w; \beta^0)$  such that

$$g_w(w; \beta^0) = - \{F_z[g(w; \beta^0), w]\}^{-1} F_w[g(w; \beta^0), w] .$$

Hence,

$$g_{ww}(w; \beta^0) = (F_z^{-2} F'_w P_{12} - F_z^{-1} P_{22})(-f_{xx})(P_{22} + g'_w P_{12}) ,$$

where

$$F_z = F_z[g(w; \beta^0), w] ,$$

$$F_w = F_w[g(w; \beta^0), w] ,$$

$$f_{xx} = f_{xx}[g(w; \beta^0) P_{12} + w P_{22}; \beta^0] ,$$

$$g_w = g_w(w; \beta^0) .$$

Note that

$$f_x[g(w_t^0; \beta^0) P_{12} + w_t^0 P_{22}; \beta^0] = f_{xt}^0 ,$$

and thus,

$$F_z[g(\tilde{w}_t^0; \beta^0), \tilde{w}_t^0] = (1 + \tilde{f}_{xt}^0 \tilde{f}_{xt}^{0'})^{1/2},$$

$$\tilde{F}_w[g(\tilde{w}_t^0; \beta^0), \tilde{w}_t^0] = 0,$$

$$g_w(\tilde{w}_t^0; \beta^0) = 0.$$

Therefore,

$$g_{ww}(\tilde{w}_t^0; \beta^0) = (1 + \tilde{f}_{xt}^0 \tilde{f}_{xt}^{0'})^{-1/2} [\tilde{I} + \tilde{f}_{xt}^{0'} \tilde{f}_{xt}^0]^{-1/2} \tilde{f}_{xxt}^0 [\tilde{I} + \tilde{f}_{xt}^{0'} \tilde{f}_{xt}^0]^{-1/2}.$$

The additive adjustment term for the observation on  $(z_t^0, \tilde{w}_t^0)$  is  $(dz_{nt}, 0)$ , where

$$dz_{nt} = \frac{1}{2} \sigma^2 \text{tr}[g_{ww}(\tilde{w}_t^0; \beta^0)] \quad .$$

Using (4.51), we transform back to the original coordinate  $(y, x)$ .

Then, the additive adjustment term for  $(y_{nt}, x_{nt})$  is

$$dz_{nt}(p_{11}, p_{12}) = \frac{1}{2} \sigma^2 (1 + \tilde{f}_{xt}^0 \tilde{f}_{xt}^{0'})^{-1} \text{tr}\{\tilde{f}_{xxt}^0 [\tilde{I} + \tilde{f}_{xt}^{0'} \tilde{f}_{xt}^0]^{-1}\} (1, -\tilde{f}_{xt}^0) \quad . \quad (4.53)$$

Note that under the assumption (4.50)

$$\sigma^2 (1 + \tilde{f}_{xt}^0 \tilde{f}_{xt}^{0'}) = \sigma_{vnt}^2,$$

$$\sigma^2 (1, - \underline{f}_{xt}) = \underline{\Sigma}_{vent} = (\sigma_{vent}^2, \underline{\Sigma}_{vunt}) ,$$

$$\sigma^2 [\underline{I} + \underline{f}_{xt}^0 \underline{f}_{xt}^0]^{-1} = \underline{\Lambda}_{nt}^{-1} .$$

Thus, the adjustment (4.53) can be written as

$$\frac{1}{2} \text{tr} \{ \underline{f}_{xxt}^0 \underline{\Lambda}_{nt}^{-1} \} \sigma_{vnt}^{-2} \underline{\Sigma}_{vent} . \quad (4.54)$$

If the error covariance matrix  $\underline{\Sigma}_n$  is not of the form (4.50), then we consider the tangent and the curvature relative to the tangent in the metric  $\underline{\Sigma}_n^{-1}$ . Instead of the orthogonal transformation (4.51), we obtain the desired rotation of axes by a transformation

$$(y, \underline{x}) = (z, \underline{w})Q , \quad (4.55)$$

where

$$Q = (Q_1', Q_2')' ,$$

$$Q_1 = (\sigma_{vnt}^2)^{-1/2} (1, - \underline{f}_{xt}^0) \underline{\Sigma}_n ,$$

$$Q_2 = \underline{\Lambda}_{nt}^{-1/2} [\underline{f}_{xt}^0, \underline{I}] .$$

We observe that

$$Q \Sigma_n^{-1} Q' = I .$$

Following the above steps with  $Q$  replacing  $P$ , we find that the adjustment term for  $(Y_{nt}, X_{nt})$  with a general  $\Sigma_n$  is also given by the expression (4.54). Replacing unknown parameters  $x_t$  and  $\beta$  by  $\hat{x}_{tM}$  and  $\hat{\beta}_M$ , we define the adjustment based on the curvature relative to the tangent to be

$$(d_2 Y_{nt}, d_2 X_{nt}) = \frac{1}{2} \text{tr} \{ f_{xx}(\hat{x}_{tM}; \hat{\beta}_M) \hat{\Lambda}_{nt}^{-1} \} \hat{\sigma}_{vnt}^{-2} \hat{\Sigma}_{vent} , \quad (4.56)$$

where

$$\hat{\Sigma}_{vent} = (\hat{\sigma}_{vent}, \hat{\Sigma}_{vunt}) ,$$

$$\hat{\sigma}_{vent} = \sigma_{een} - f_x(\hat{x}_{tM}; \hat{\beta}_M) \Sigma_{uen} .$$

By noting that

$$\Sigma_{vent} (1, -f_{xt}^0)' = \sigma_{vnt}^2 ,$$

it immediately follows that the adjustment (4.56) satisfies Assumptions 4.9 and 4.10 with  $(\Delta_2 Y_{nt}, \Delta_2 X_{nt})$  being the quantity in (4.54). Therefore, the estimators  $\tilde{\beta}_2$  and  $\tilde{x}_{t2}$  obtained using  $(d_2 Y_{nt}, d_2 X_{nt})$  have the properties stated in Theorem 4.6 and Theorem 4.7. Since

$$\Sigma_{vent} \Sigma_n^{-1} [f_{xt}^{0'}, I] = 0 ,$$

the expansion of  $\tilde{x}_{t2}$  is given by

$$\begin{aligned} \tilde{x}_{t2} - x_t^0 &= \delta_{nt} + [\delta_1 \hat{x}_t + \delta_2 \hat{x}_t + \delta_{x_t^*}] (a_n \Lambda_{nt}^{-1}) \\ &+ O_p(\max[a_n^{-3/2}, a_n^{-1/2} n^{-1/2}]) . \end{aligned}$$

Hence, up to the terms of  $O_p(a_n^{-1})$ , the adjusted estimators  $\tilde{\beta}_2$  and  $\tilde{x}_{t2}$  are equivalent to the modified maximum likelihood estimators  $\tilde{\beta}$  and  $\tilde{x}_t$ . The relation that  $E[\delta_{x_t^*}] (a_n \Lambda_{nt}^{-1}) = \Delta_2 x_{nt}$  could have been derived directly by Theorem 2.1.

### 3. Adjustment based on the expansion terms

The two previous adjusted estimators are obtained by subtracting estimates of the bias from the observations. Instead of estimating the bias, we can estimate a term in the expansion of  $\hat{\beta}_M$  which introduces the bias in  $\hat{\beta}_M$ . From the expression (4.38), we observe that an adjustment can be obtained by letting  $(dy_{nt} - dx_{nt} f_{xt}^{0'})$  estimate the term  $-\delta_{nt} f_{xxt}^0 (\frac{1}{2} \delta_{nt}' - u_{nt}')$ .

Let

$$d_3 y_{nt} = -\frac{1}{2} \text{tr}[f_{xx}(\hat{x}_{tM}; \hat{\beta}_M)(\hat{u}_{nt}' \hat{u}_{nt} - \Sigma_{uun})] ,$$

$$d_3 x_{nt} = 0 ,$$

where

$$\hat{\underline{u}}_{nt} = \underline{X}_{nt} - \hat{\underline{x}}_{tM} .$$

We note that  $d_3 Y_{nt}$  has the same form as the modification used by Wolter and Fuller (1982b). It follows that

$$d_3 Y_{nt} = \Delta_3 Y_{nt} + O_p(\max[a_n^{-3/2}, n^{-1/2} a_n^{-1/2}]) ,$$

where

$$\Delta_3 Y_{nt} = -\frac{1}{2} \text{tr}\{\underline{f}_{\text{xxxt}}^0 [(\underline{u}_{nt} - \underline{\delta}_{nt})'(\underline{u}_{nt} - \underline{\delta}_{nt}) - \underline{\Sigma}_{\text{uun}}]\} ,$$

$$E\{\Delta_3 Y_{nt}\} = \frac{1}{2} \text{tr}[\underline{f}_{\text{xxxt}}^0 \underline{\Lambda}_{nt}^{-1}] .$$

Therefore,  $(d_3 Y_{nt}, d_3 X_{nt})$  satisfies Assumption 4.10 with  $\Delta_3 X_{nt} = 0$ .

Since

$$\underline{\delta}_{nt} \underline{f}_{\text{xxxt}}^0 \left( \frac{1}{2} \underline{\delta}_{nt}' - \underline{u}_{nt}' \right) = \frac{1}{2} \text{tr}\{\underline{f}_{\text{xxxt}}^0 [(\underline{u}_{nt} - \underline{\delta}_{nt})'(\underline{u}_{nt} - \underline{\delta}_{nt}) - \underline{u}_{nt}' \underline{u}_{nt}]\} ,$$

we expect the adjustment  $(d_3 Y_{nt}, d_3 X_{nt})$  to cancel the term  $\delta_2 \hat{\underline{\beta}}$  in the expansion of  $\hat{\underline{\beta}}_M$ . However, to derive the asymptotic properties of the resulting estimators  $\tilde{\underline{\beta}}_3$  and  $\tilde{\underline{x}}_{t3}$ , we need to introduce an additional assumption.

Assumption 4.11. For some real  $L^0$  and all  $t$  and  $n$ ,

$$E\{|\varepsilon_{nt}|^8\} < L^0 a_n^{-4}.$$

Under Assumption 4.11 and the assumptions in Theorem 4.6, it follows from (4.29) that  $(d_3 Y_{nt}, d_3 X_{nt})$  satisfies Assumption 4.9. Thus, with Assumption 4.11, Theorem 4.6 and Theorem 4.7 hold for the adjusted estimators  $\tilde{\beta}_3$  and  $\tilde{x}_{t3}$ . Hence,  $\tilde{\beta}_3$  has the same properties as  $\tilde{\beta}_1$  and  $\tilde{\beta}_2$ , and  $\tilde{x}_{t3}$  is unbiased through terms of  $O_p(a_n^{-1})$ . Note that the adjustment  $d_3 Y_{nt}$  is similar to the modification used in the modified maximum likelihood estimators  $\tilde{\beta}$  and  $\tilde{x}_t$ . While  $\tilde{\beta}_3$  and  $\tilde{\beta}$  are asymptotically equivalent, the almost unbiased estimator  $\tilde{x}_{t3}$  is clearly superior to the estimator  $\tilde{x}_t$  which has the bias due to the curvature.

#### 4. Comparison among the adjusted estimators

While all three adjusted estimators  $\tilde{\beta}_1$ ,  $\tilde{\beta}_2$ , and  $\tilde{\beta}_3$  have the same expansion form, the estimators  $\tilde{x}_{t1}$ ,  $\tilde{x}_{t2}$ , and  $\tilde{x}_{t3}$  have slightly different expansions. By Theorem 2.1 and Theorem 4.7, we obtain

$$\tilde{x}_{t1} - x_t^0 = \Delta \tilde{x}_t + \frac{1}{2} \text{tr}[\tilde{f}_{xxt}^0 (\delta'_{nt} \delta_{nt} - \Lambda_{nt}^{-1})] \sigma_{vnt}^{-2} \Sigma_{vunt}$$

$$+ O_p(\max[a_n^{-3/2}, n^{-1/2} a_n^{-1/2}]) ,$$

$$\tilde{x}_{t2} - x_t^0 = \Delta \tilde{x}_t + \frac{1}{2} \text{tr}[\tilde{f}_{xxt}^0 \delta'_{nt} \delta_{nt}] \sigma_{vnt}^{-2} \Sigma_{vunt}$$

$$+ O_p(\max[a_n^{-3/2}, n^{-1/2} a_n^{-1/2}]) ,$$

$$\begin{aligned} \tilde{x}_{t3} - x_t^0 &= \Delta \tilde{x}_t + \frac{1}{2} \text{tr} [f_{xxt}^0 (\delta_{nt}' \delta_{nt} - \Lambda_{nt}^{-1} + \tilde{u}_{nt}' \tilde{u}_{nt} - \Sigma_{uun}^{\sim})] \sigma_{vnt}^{-2} \Sigma_{vunt} \\ &+ O_p(\max[a_n^{-3/2}, n^{-1/2} a_n^{-1/2}]) , \end{aligned} \quad (4.57)$$

where

$$\Delta \tilde{x}_t = \delta_{nt} + [\delta_1 \hat{x}_t + \delta_2 \hat{x}_t] (a_n \Lambda_{nt}^{-1}) ,$$

$$\tilde{u}_{nt} = u_{nt} - \delta_{nt} ,$$

$$\Sigma_{uun}^{\sim} = \Sigma_{uun} - \Lambda_{nt}^{-1} = \Sigma_{uvnt} \sigma_{vnt}^{-2} \Sigma_{vunt} .$$

It is obvious from (4.57) that  $\tilde{x}_{t1}$  is superior to  $\tilde{x}_{t2}$  as an estimator of  $x_t^0$ . If  $\varepsilon_{nt}$  are normally distributed, then  $\delta_{nt}$  is independent of  $v_{nt}^*$ ,  $\tilde{u}_{nt}$ , and  $\varepsilon_{nt}^*$ . Also, the central third moments of the normal distribution are zero. Thus, it follows that the covariance between the first and second terms in each of the three expansions is zero. Hence, for the normal error model we need to compare the mean squared errors of the second terms in (4.57). Under the normality of  $\varepsilon_{nt}$ , we obtain

$$\begin{aligned} V\{\text{vech}(\delta_{nt}' \delta_{nt} - \Lambda_{nt}^{-1} + \tilde{u}_{nt}' \tilde{u}_{nt} - \Sigma_{uun}^{\sim})\} &= 2\psi_q(\Lambda_{nt}^{-1} \otimes \Lambda_{nt}^{-1}) \psi_q' \\ &+ 2\psi_q(\Sigma_{uun}^{\sim} \otimes \Sigma_{uun}^{\sim}) \psi_q' . \end{aligned}$$

Therefore, in terms of asymptotic mean squared error,  $\tilde{x}_{t1}$  is preferred over  $\tilde{x}_{t2}$  and  $\tilde{x}_{t3}$  for the normal model.

For estimation of  $\beta^0$ , the choice of adjusted estimators is left undetermined. All three estimators  $\tilde{\beta}_1$ ,  $\tilde{\beta}_2$ , and  $\tilde{\beta}_3$  were derived based on the asymptotic results for the maximum likelihood type estimator of  $\beta^0$ . The estimator  $\tilde{\beta}_1$  is associated with  $\tilde{x}_1$  which is superior to the estimators of  $x_t^0$  associated with  $\tilde{\beta}_1$  and  $\tilde{\beta}_2$ . The estimator  $\tilde{\beta}_2$  has an intuitive appeal in the sense that the adjustment is distributed over  $Y_{nt}$  and  $x_{nt}$  proportioned to  $(\sigma_{vent}, \Sigma_{vunt})$ . On the other hand, the adjustment in  $\tilde{\beta}_3$  gives a cancellation of a term, which may be preferred over the adjustment in  $\tilde{\beta}_1$  and  $\tilde{\beta}_2$  using the expectation of the term. But, all three estimators have the same asymptotic properties up to the level of approximation we considered. In this section, we do not attempt to give any further comparison among the adjusted estimators of  $\beta^0$ . In the next section, using the Monte Carlo method, we investigate the small sample properties of various estimators of  $\beta^0$  including the three adjusted estimators.

#### G. A Monte Carlo Study

In this section, we present the results from a Monte Carlo study carried out to compare the small sample behavior of the estimators of  $\beta^0$ . Due to an extensive amount of iterative computations required to obtain the estimates, we conducted a relatively small Monte Carlo experiment.

### 1. Generating the data

The model considered is

$$y_t = \beta_0^0 + \beta_1^0 x_t + \beta_2^0 x_t^2$$

$$= (1, x_t, x_t^2) \beta^0, \quad t = 1, 2, \dots, N,$$

where

$$\beta^0 = (\beta_0^0, \beta_1^0, \beta_2^0)',$$

and we observe

$$Y_t = y_t + e_t,$$

$$X_t = x_t + u_t.$$

In this study, the sample size is  $N = 33$ , the vector of  $x$ -values is

$$(x_1, x_2, \dots, x_{33}) = (1, 1, 1) \oplus (-0.5, -0.4, -0.3, \dots, 0.4, 0.5),$$

the true parameter vector is  $\beta^0 = (0, 0, 1)'$ , and the measurement errors  $(e_t, u_t)$  were generated as a random sample from a bivariate normal  $(0, 0.0324 I)$ . Note that

$$(33)^{-1} \sum_{t=1}^{33} (x_t - \bar{x})^2 = 0.1.$$

Thus, the ratio of the standard error  $\sigma_u$  to the root mean square for  $x$  is 0.5692. For this parameter set, 100 samples were generated.

## 2. Estimators compared

For each 100 samples, we computed six estimators. These are the pseudo-maximum likelihood estimator  $\hat{\beta}_M$ , the modified maximum likelihood estimator  $\tilde{\beta}_M$ , three adjusted estimators  $\hat{\beta}_{A1}$ ,  $\hat{\beta}_{A2}$ , and  $\hat{\beta}_{A3}$  introduced in Section F, and the ordinary least squares estimator  $\hat{\beta}_L$ . Computations of the first five estimators require a preliminary estimator  $\bar{\beta}$  of  $\beta^0$ . Wolter and Fuller (1982b) reported a Monte Carlo mean 1.37 for  $\hat{\beta}_{M,2}$ , where

$$\hat{\beta}_M = (\hat{\beta}_{M,0}, \hat{\beta}_{M,1}, \hat{\beta}_{M,2})' .$$

We used

$$\bar{\beta} = (0, 0, 1.37)$$

as the preliminary estimator of  $\beta^0$  for  $\hat{\beta}_M$ ,  $\tilde{\beta}_M$ ,  $\hat{\beta}_{A1}$ ,  $\hat{\beta}_{A2}$ , and  $\hat{\beta}_{A3}$ .

Wolter and Fuller stated that the small sample distribution function of the estimator  $\hat{\beta}_M$  has thick tails, and introduced a modification to improve the small sample behavior of  $\hat{\beta}_M$  and  $\tilde{\beta}_M$ . We applied their modification to  $\hat{\beta}_M$ ,  $\tilde{\beta}_M$ ,  $\hat{\beta}_{A1}$ ,  $\hat{\beta}_{A2}$ , and  $\hat{\beta}_{A3}$ . Given the preliminary estimator  $\bar{\beta} = (\bar{\beta}_0, \bar{\beta}_1, \bar{\beta}_2)$ , let

$$\bar{x}_t = X_t - (1 - N^{-1}\alpha)\bar{u}_t, \quad (4.58)$$

where

$$\begin{aligned} \bar{u}_t &= \bar{v}_t \bar{\sigma}_{vt}^{-2} \bar{\sigma}_{vut}, \\ \bar{v}_t &= Y_t - (1, X_t, X_t^2) \bar{\beta}, \\ \bar{\sigma}_{vt}^2 &= (1, -\bar{f}_{xt}) \bar{\Sigma} (1, -\bar{f}_{xt})' \\ &= 0.0324 (1 + \bar{f}_{xt}^2)', \\ \bar{\sigma}_{vut} &= (1, -\bar{f}_{xt}) \bar{\Sigma} (0, 1)' \\ &= -0.0324 \bar{f}_{xt}, \\ \bar{f}_{xt} &= \bar{\beta}_1 + 2 \bar{\beta}_2 X_t, \end{aligned} \quad (4.59)$$

$$\bar{\Sigma} = 0.0324 \bar{I},$$

$$N = 33.$$

The modification associated with the  $\alpha$ -value is similar to that studied by Fuller (1980) for the linear errors-in-variables model. In our study, we used the value  $\alpha=4$ .

Substituting (4.58) into the expression (4.5) defining the pseudo-maximum likelihood estimator, we obtained the first round value of

$\hat{\beta}_M$ . Then, replacing  $X_t$  in (4.59) by  $\bar{x}_t$  and  $\bar{\beta}$  by the first round  $\hat{\beta}_M$ , a new estimate of  $x_t^0$  was obtained by (4.58). The final estimate  $\hat{\beta}_M$  was computed by four iterations of the procedure, successively applying (4.58) and (4.5).

The modified maximum likelihood estimator  $\tilde{\beta}_M$  was computed by four iterations of the procedure, successively obtaining  $\bar{x}_t$  and  $\hat{\beta}_M$  by (4.58) and (4.6).

For the three adjusted estimators  $\hat{\beta}_{A1}$ ,  $\hat{\beta}_{A2}$ , and  $\hat{\beta}_{A3}$ , we let

$$d_1 Y_t = \bar{\beta}_2 \bar{\Lambda}_t^{-1},$$

$$(d_2 Y_t, d_3 X_t) = \bar{\beta}_2 \bar{\Lambda}_t^{-1} \bar{\sigma}_{vt}^{-2} (\bar{\sigma}_{vet}, \bar{\sigma}_{vut}),$$

$$d_3 Y_t = -\bar{\beta}_2 (\hat{u}_t^2 - 0.0324),$$

where

$$\bar{\Lambda}_t^{-1} = 0.0324 - \bar{\sigma}_{vut}^2 \bar{\sigma}_{vt}^{-2},$$

$$\bar{\sigma}_{vet} = (1, -\bar{f}_{xt}) \Sigma (1, 0)'$$

$$= 0.0324,$$

$$\hat{u}_t = X_t - \bar{x}_t.$$

The estimate  $\hat{\beta}_{A1}$  was computed by four iterations of the procedure, successively applying (4.58) and (4.5), where at each iteration observations  $(Y_t, X_t)$  were updated by adding  $(d_1 Y_t, 0)$  based on the most recent  $\bar{x}_t$  and  $\hat{\beta}_{A1}$ .

The computational procedures for  $\hat{\beta}_{A2}$  and  $\hat{\beta}_{A3}$  are the same as that for  $\hat{\beta}_{A1}$ , except that the term added to observations  $(Y_t, X_t)$  at each iteration is  $(d_2 Y_t, d_2 X_t)$  for  $\hat{\beta}_{A2}$  and is  $(d_3 Y_t, 0)$  for  $\hat{\beta}_{A3}$ .

Finally, the ordinary least squares estimator  $\hat{\beta}_L$  was computed by regressing  $Y_t$  on  $(1, X_t, X_t^2)$ .

### 3. Results and conclusions

Our discussion concentrates on the results for estimators of  $\beta_2^0$ , the coefficient of the quadratic term. Table 1 contains the Monte Carlo properties of the six estimators of  $\beta_2^0$ .

Table 1. Monte Carlo Properties of Estimators of  $\beta_2^0$   
(100 samples)

Estimator	Mean	Var.	MSE	Percentile		
				25%	50%	75%
$\hat{\beta}_{M,2}$	1.152	0.633	0.706	0.506	0.951	1.561
$\tilde{\beta}_{M,2}$	1.078	0.779	0.785	0.473	0.874	1.425
$\hat{\beta}_{A1,2}$	0.976	0.352	0.353	0.521	0.888	1.291
$\hat{\beta}_{A2,2}$	0.932	0.311	0.316	0.511	0.858	1.246
$\hat{\beta}_{A3,2}$	1.152	0.716	0.739	0.522	0.928	1.490
$\hat{\beta}_{L,2}$	0.386	0.063	0.440	0.231	0.367	0.529

The ordinary least squares estimator  $\hat{\beta}_{L,2}$  has the smallest variance, but severely underestimates  $\beta_2^0 = 1$ . The results for  $\tilde{\beta}_{M,2}$  agree reasonably well with those reported for the same estimator in Wolter and Fuller (1982b).

The estimators  $\hat{\beta}_{A1,2}$  and  $\hat{\beta}_{A2,2}$  have smaller biases and smaller variances than the estimators  $\hat{\beta}_{M,2}$ ,  $\tilde{\beta}_{M,2}$ , and  $\hat{\beta}_{A3,2}$ . The mean squared errors of  $\hat{\beta}_{M,2}$ ,  $\tilde{\beta}_{M,2}$ , and  $\hat{\beta}_{A3,2}$  are at least twice as large as those of  $\hat{\beta}_{A1,2}$  and  $\hat{\beta}_{A2,2}$ .

It is somewhat surprising that  $\hat{\beta}_{M,2}$  has a smaller Monte Carlo mean squared error than  $\tilde{\beta}_{M,2}$ . The modification associated with  $\alpha = 4$  resulted in considerably smaller Monte Carlo bias and mean squared error for  $\hat{\beta}_{M,2}$  than those given in Wolter and Fuller (1982b) for the maximum likelihood estimator without the  $\alpha$ -modification.

We recall that the estimators  $\tilde{\beta}_M$  and  $\hat{\beta}_{A3}$  use a similar type of bias adjustment. The small sample behavior of  $\tilde{\beta}_M$  and  $\hat{\beta}_{A3}$  are also similar.

The distribution of the five iterative estimators of  $\beta_2^0$  are all positively skewed. The estimators  $\hat{\beta}_{A1,2}$  and  $\hat{\beta}_{A2,2}$  have a more symmetric distribution than the estimators  $\hat{\beta}_{M,2}$ ,  $\tilde{\beta}_{M,2}$  and  $\hat{\beta}_{A3,2}$ .

The variance of the asymptotic distribution is 0.1948 for  $\hat{\beta}_{M,2}$ ,  $\tilde{\beta}_{M,2}$ ,  $\hat{\beta}_{A1,2}$ ,  $\hat{\beta}_{A2,2}$ , and  $\hat{\beta}_{A3,2}$ . The Monte Carlo variances of all five iterative estimators are considerably larger than the asymptotic variance.

Statistics analogous to Student's  $t$  were also computed for  $\hat{\beta}_{M,2}$ ,  $\tilde{\beta}_{M,2}$ ,  $\hat{\beta}_{A1,2}$ ,  $\hat{\beta}_{A2,2}$ , and  $\hat{\beta}_{A3,2}$ . The covariance matrix of each

iterative estimator was estimated by the inverse of

$$\sum_{t=1}^{33} \hat{\sigma}_{vt}^{-2} (1, x_t, x_t^2)' (1, x_t, x_t^2), \quad (4.60)$$

where  $\hat{\sigma}_{vt}^2$  and  $\hat{x}_t$  are the estimators of  $\sigma_{vt}^2$  and  $x_t$  at the fourth iteration in the computation of each estimator. The t-statistics for all five iterative estimators of  $\beta_2^0$  had negatively skewed distributions with heavy tails in comparison to the  $N(0,1)$  distribution. The sample percentiles of the statistics deviated considerably from percentiles of the  $N(0,1)$  distribution. Wolter and Fuller (1982b) introduced a modified estimator of the covariance matrix. Their Monte Carlo study showed that the modified covariance estimator produced t-statistics whose distributions agree reasonably well with the  $N(0,1)$  distribution. We conjecture that a modification similar to that of Wolter and Fuller would improve the small sample behavior of the t-statistics for  $\hat{\beta}_{A1}$ ,  $\hat{\beta}_{A2}$ , and  $\hat{\beta}_{A3}$ .

Three conclusions are possible from the results of our Monte Carlo experiment. These are:

- 1) The ordinary least squares estimator  $\hat{\beta}_L$  is not practical because of its large bias.
- 2) The adjusted estimators  $\hat{\beta}_{A1}$  and  $\hat{\beta}_{A2}$  with the  $\alpha$ -modification are recommended for practical use.
- 3) The inverse of the matrix (4.60) is not recommended as an estimator of the covariance matrix of  $\hat{\beta}_M$ ,  $\tilde{\beta}_M$ ,  $\hat{\beta}_{A1}$ ,  $\hat{\beta}_{A1}$ ,  $\hat{\beta}_{A2}$ , or  $\hat{\beta}_{A3}$ .

## V. INSTRUMENTAL VARIABLE ESTIMATION OF THE NONLINEAR MODEL

In this chapter, we consider the estimation of the errors-in-variables model with a nonlinear functional relationship, when there are available observations on variables outside the relationship of interest. Such an additional information enables us to estimate the unknown parameters in the relationship without knowledge of the error covariance matrix. However, as in Chapter IV, the assumption of decreasing errors in the explanatory variables seems to be necessary to obtain consistent estimators for the general nonlinear model.

### A. Introduction

Let  $\{b_n\}_{n=1}^{\infty}$  and  $\{a_n\}_{n=1}^{\infty}$  be sequences of positive real numbers such that  $n = b_n a_n$  for  $n = 1, 2, \dots, \infty$ ,  $a_n = o(n)$ , and  $b_n = o(n)$ . We assume the existence of a sequence of experiments indexed by  $n$ . The functional relation of interest is

$$y_t^0 = f(\underline{x}_t^0; \underline{\beta}^0) + e_t^0, \quad t = 1, 2, \dots, b_n, \quad (5.1)$$

where  $\underline{x}_t^0$  are  $1 \times q$  vectors of fixed constants belonging to a parameter space  $\underline{\Gamma}$ ; a convex subset of  $q$ -dimensional Euclidean space,  $\underline{\beta}^0$  is a  $k \times 1$  vector belonging to a parameter space  $\underline{\Omega}$ ; a convex

compact subset of  $k$ -dimensional Euclidean space,  $f(z; \theta)$  is continuous on  $\Gamma \times \Omega$ , and  $e_t^0$  are errors in the equation. Our observations are  $(Y_{nt}, X_{nt})$ , where

$$\begin{aligned} Y_{nt} &= y_t^0 + e_{nt}^* , \\ X_{nt} &= x_t^0 + u_{nt} , \quad t = 1, 2, \dots, b_n , \end{aligned} \quad (5.2)$$

and  $e_{nt}^*$  and  $u_{nt}$  are the measurement errors. In addition, it is assumed that there are available observations

$$W_{nt} = w_t^0 + r_{nt} , \quad t = 1, 2, \dots, b_n , \quad (5.3)$$

where  $w_t^0$  are  $1 \times p$  vectors of constants,  $p \geq k$  and  $r_{nt}$  are random errors. The model is said to be just identified by the counting rule if  $p = k$ , and to be over identified by the counting rule if  $p > k$ .

The nonlinear errors-in-variables model defined above differs from the nonlinear simultaneous equation system in econometrics. If the errors-in-variables model is linear in two variables and is given by

$$\begin{aligned} Y_t &= x_t \beta + e_t , \\ X_t &= x_t + u_t , \quad t = 1, 2, \dots, n, \end{aligned}$$

then substituting the second equation into the first we have

$$Y_t = X_t\beta + v_t, \quad t = 1, 2, \dots, n,$$

where the  $v_t = e_t - u_t\beta$  are independently and identically distributed with zero mean. Note that  $v_t$  is correlated with  $X_t$ . It is assumed that the additional variable  $W_t$  is correlated with  $x_t$  but is uncorrelated with  $(e_t, u_t)$ , and with this information  $\beta$  can be estimated consistently. The linear errors-in-variables model is equivalent to the simultaneous equation model with  $Y_t$  and  $X_t$  as endogenous variables and  $W_t$  as an exogenous variable outside the equation of interest. For another type of correspondence between the linear errors-in-variables model and the linear simultaneous equation model, obtained by treating the estimated reduced form coefficients as observations in the errors-in-variables model, see Fuller (1977) and Anderson (1976). When the model is nonlinear in the variables, the equivalence of the two models no longer holds. An equation in the nonlinear simultaneous equation system has the form

$$Y_t = g(X_t; \beta) + v_t, \quad t = 1, 2, \dots, n, \quad (5.4)$$

where  $X_t$  is an endogenous variable, and we assume for simplicity that there is no exogenous variable in the equation. The nonlinear errors-in-variables model with  $b_n = n$  and  $a_n = 1$  is of the form

$$\begin{aligned} Y_t &= f(x_t; \beta) + e_t \\ X_t &= x_t + u_t \quad . \quad t = 1, 2, \dots, n \quad . \end{aligned} \quad (5.5)$$

In the model (5.4), for the two observed variables  $Y_t$  and  $X_t$ , the variable  $Y_t - g(X_t; \beta)$  is independently and identically distributed with zero mean. But, in (5.5), for the observations  $Y_t$  and  $X_t$ ,

$$Y_t - f(X_t; \beta) = e_t - \frac{\partial f}{\partial x}(X_t^*; \beta)u_t, \quad (5.6)$$

where  $X_t^*$  is a random variable lying on the segment joining  $x_t$  and  $X_t$ , and we have assumed  $f$  is continuously differentiable. Since  $X_t^*$  and, thus,  $\frac{\partial f}{\partial x}(X_t^*; \beta)$  depend on  $u_t$  the right-hand side of (5.6) does not have zero mean unless  $\frac{\partial f}{\partial x}(\cdot; \beta)$  is constant. But if  $\frac{\partial f}{\partial x}(\cdot; \beta)$  is constant, then the model (5.5) is linear in  $x_t$ .

Therefore, the nonlinear errors-in-variables model differs from the nonlinear simultaneous equation system in the sense that the relation to be estimated is the one between the unobserved true values in the former and is the one between observed values in the latter. To avoid the trivial case, we assume, from now on, that the model (5.1) is nonlinear in  $x_t^0$ .

In the model specified by (5.1), (5.2), and (5.3), we assume that  $(e_t^0, e_{nt}^*, u_{nt}, r_{nt})$  are independently and identically distributed with mean zero and covariance matrix

$$\begin{bmatrix} \sigma_{ee}^0 & 0 & 0 & 0 \\ 0 & \sigma_{een}^* & \Sigma_{eun} & \Sigma_{ern} \\ 0 & \Sigma_{uen} & \Sigma_{uun} & \Sigma_{urn} \\ 0 & \Sigma_{ren} & \Sigma_{run} & \Sigma_{rrn} \end{bmatrix} ,$$

where

$$\begin{bmatrix} \sigma_{een}^* & \Sigma_{eun} \\ \Sigma_{uen} & \Sigma_{uun} \end{bmatrix} = O(a_n^{-1}) \quad . \quad (5.7)$$

Interpretations of the Assumption (5.7) are the same as those given for the model in Chapter IV. Note that we are interested in the relationship between  $y_t^0$  and  $x_t^0$ , and that  $w_{nt}$  does not appear in the relationship. The  $w_{nt}$  may be measurements of  $x_t^0$  by a different measuring procedure from that used for  $x_{nt}$ . Each  $w_{nt}$  is not necessarily the mean of  $a_n$  observations. Also, the term  $r_{nt}$  does not entirely consist of measurement error.

Letting

$$e_{nt} = e_t^0 + e_{nt}^* ,$$

$$\sigma_{ee} = \sigma_{ee}^0 + \sigma_{ee}^* ,$$

we can write our model as

$$Y_{nt} = f(x_t^0; \beta^0) + e_{nt} ,$$

$$\tilde{x}_{nt} = x_t^0 + u_{nt} ,$$

$$\tilde{w}_{nt} = w_t^0 + r_{nt} , \quad t = 1, 2, \dots, b_n ,$$

$$\begin{bmatrix} e_{nt} \\ u'_{nt} \\ r'_{nt} \end{bmatrix} \sim II \begin{bmatrix} 0 \\ 0 \\ 0 \end{bmatrix} , \quad \begin{bmatrix} \Sigma_n & \Sigma_{ern} \\ \Sigma_{ren} & \Sigma_{rrn} \end{bmatrix} , \quad (5.8)$$

where

$$\Sigma_n = \begin{bmatrix} \sigma_{een} & \Sigma_{eun} \\ \Sigma_{uen} & \Sigma_{uun} \end{bmatrix} ,$$

$$\Sigma_{ren} = (\Sigma_{ren}, \Sigma_{run}) .$$

We assume that  $b_n$  is known, but that  $a_n$  is unobservable. That is, we can use the value  $b_n$  in our inference, but not the actual value of  $a_n$ .

If the error in the relationship exists, then  $\sigma_{ee}^2 > 0$  , and thus  $\sigma_{een} = O(1)$  . If the relationship is perfect, then  $\sigma_{een} = O(a_n^{-1})$  .

This difference introduces some difficulty in deriving asymptotic results. When the errors  $e_t^0$ ,  $t = 1, 2, \dots, b_n$ , are present,  $b_n^{1/2}$  is an obvious choice for the normalizing constant to obtain a limiting distribution of an estimator. However, the contribution of the  $b_n$  measurement errors each with variance of  $O(a_n^{-1})$  is  $n^{-1/2}$ . Hence, the measurement error contribution would disappear in the limiting results, if the normalizing constant  $b_n^{1/2}$  was used. But, it is desirable to keep the effect of measurement errors in the limiting results. Also, we are interested in the asymptotic results with the normalizing constant  $n^{1/2}$ . Therefore, for the purpose of investigating the limiting behavior of estimators, the model with errors in the relationship is of little interest. Hence, from now on, we assume that there are no errors in the relationship. That is, we assume  $\sigma_{\text{een}} = O(a_n^{-1})$ . This assumption does not necessarily restrict the range of applications of our asymptotic results. As we discussed in Chapter IV, the asymptotic results with normalizing constant  $n^{1/2}$  are applicable when either measurement errors are small or the number  $b_n$  of data points is large. Thus, we may apply our results to the case where the error in the relationship and the measurement errors are considered to be small and the number of observations is large. Hence, we assume that the term  $e_{nt}$  includes the error in the relationship and has a variance of  $O(a_n^{-1})$ . This is an unpleasing aspect of the nonlinear model. We recall that in the linear model the instrumental variable estimation is useful regardless of the existence of errors in the equation. The instrumental variable estimator of  $\beta^0$  which will be defined in the

next section can be shown to be consistent for  $\beta^0$  under appropriate assumptions, even when  $\sigma_{een} = O(1)$ , i.e., when an error in the relationship exists. However, the properties other than consistency do not hold if  $\sigma_{een} = O(1)$ , since the orders of the error in the equation and the measurement errors are different.

Some care is required in presenting the assumptions on  $\Sigma_{rrn}$ . As we noted earlier, the variable  $\tilde{W}_{nt}$  does not appear in the equation of interest. Thus, multiplying  $\tilde{W}_{nt}$  by a constant, say  $b_n^{-1}$ , does not change the relationship. Also, as we will see, our estimation procedure is invariant under this type of transformation of  $\tilde{W}_{nt}$ . Hence, there is some arbitrariness associated with the order of  $\tilde{W}_{nt}$ . We note that the true values  $x_t^0$  are fixed. As in the linear case, for an instrumental variable  $\tilde{W}_{nt}$  to produce a consistent estimator, the random part  $\tilde{\varepsilon}_{nt}$  with zero mean cannot dominate the fixed part which helps the estimation procedure. This condition may be written as

$$\tilde{T}_n' \Sigma_{rrn} \tilde{T}_n = O(1), \quad (5.9)$$

where  $\tilde{T}_n$  satisfies

$$\tilde{T}_n' (b_n^{-1} \sum_{t=1}^{b_n} w_t^0 w_t^0) \tilde{T}_n = I.$$

Under the Assumption (5.9), the order of  $b_n^{-1} \sum_{t=1}^{b_n} \tilde{W}_{nt}' \tilde{W}_{nt}$  is the same as

the order of

$$b_n^{-1} \sum_{t=1}^{b_n} \tilde{w}_t^0 \tilde{w}_t^0 .$$

To avoid the ambiguity due to the arbitrariness of the order of  $\tilde{w}_{nt}$  and to simplify our discussion, we assume

$$\lim_{n \rightarrow \infty} b_n^{-1} \sum_{t=1}^{b_n} \tilde{w}_t^0 \tilde{w}_t^0 = \bar{m}_{ww} , \quad (5.10)$$

where  $\bar{m}_{ww}$  is positive definite. With the Assumption (5.10), the condition (5.9) is equivalent to assuming that the order of  $\Sigma_{rrn}$  is at most one. If  $\tilde{w}_{nt}$  can be considered as a mean of  $a_n$  observations, or if  $\tilde{r}_{nt}$  are measurement errors, then  $\Sigma_{rrn} = O(a_n^{-1})$ . Also, if  $\tilde{w}_{nt}$  have no errors, then  $\Sigma_{rrn} = 0$ , and the condition  $\Sigma_{rrn} = O(1)$  still holds. Thus, the result derived using the order  $\Sigma_{rrn} = O(1)$  is also valid for the case where either  $\Sigma_{rrn} = O(a_n^{-1})$  or  $\Sigma_{rrn} = 0$ . Note also that if  $\Sigma_n = O(a_n^{-1})$  and  $\Sigma_{rrn} = O(1)$ , then  $\Sigma_{ren} = O(a_n^{-1/2})$ .

In the analysis of the linear instrumental variable model, it has often been assumed, without loss of generality, that

$$b_n^{-1} \sum_{t=1}^{b_n} \tilde{w}_{nt}' \tilde{w}_{nt} = I . \quad (5.11)$$

If  $\tilde{w}_{nt}$  are independent of  $(e_{nt}, u_{nt})$ , we can always make a transformation to satisfy (5.11) without altering the model assumption. But, the assumption of zero correlation between  $(e_{nt}, u_{nt})$  and  $\tilde{r}_{nt}$

without independence may be destroyed by a transformation used to obtain (5.11). It is still possible to use the form (5.11) to simplify the computations. We do not assume that the condition (5.11) holds.

We will use the same notations as those in Chapter IV to denote the partial derivatives of the function  $f(\underline{x}; \underline{\beta})$ . Thus, for example,  $\underline{f}_{\beta\beta}(\underline{z}; \underline{\theta})$  is the matrix of second partial derivatives with respect to the elements of  $\underline{\beta}$  evaluated at  $(\underline{z}; \underline{\theta})$ , and  $\underline{f}_{xt}^0$  is the partial derivative with respect to the elements of  $\underline{x}$  evaluated at  $(\underline{x}_t^0; \underline{\beta}^0)$ .

In this chapter, the order, or the order in probability statements are taken as  $n$  tends to infinity, unless specified otherwise.

## B. The Instrumental Variable Estimator

### 1. The estimator

We introduce an estimator of  $\underline{\beta}^0$  in the model (5.8) which does not require the knowledge of the error covariance matrix. Our estimator can be considered as a version of the two-stage least squares estimator. In the context of the nonlinear simultaneous equation model, T. Amemiya (1974) introduced the nonlinear two-stage least squares estimator. His model is

$$y_t = f(\underline{z}_t; \underline{\beta}^0) + e_t \quad t = 1, 2, \dots, n,$$

where  $\underline{z}_t$  consists partly of endogenous variables and partly of exogenous variables. Given the vector  $\underline{x}_t$  of exogenous variables in

the system T. Amemiya defined the nonlinear two-stage least squares estimator of  $\beta^0$  to be the value of  $\beta$  that minimizes

$$\Phi(\beta) = \left\{ \sum_{t=1}^n [Y_t - f(z_t; \beta)] X_t \right\} \left\{ \sum_{t=1}^n X_t' X_t \right\}^{-1} \left\{ \sum_{t=1}^n X_t' [Y_t - f(z_t; \beta)] \right\}.$$

It is easy to see that for the linear simultaneous equation model the estimator minimizing  $\Phi(\beta)$  reduces to the usual two-stage least squares estimator.

As we discussed earlier, the nonlinear errors-in-variables model is different from the nonlinear simultaneous equation model. In our nonlinear functional relationship model,  $Y_{nt} - f(X_{nt}; \beta^0)$  is not an identically distributed random variable with zero mean. However, we minimize a function analogous to  $\Phi(\beta)$  to obtain an estimator of  $\beta^0$  in the model (5.8).

We define the instrumental variable estimator  $\hat{\beta}_n$  of  $\beta^0$  to be the value of  $\beta$  in  $\Omega$  that minimizes

$$Q_n(\beta) = \left\{ \sum_{t=1}^{b_n} [Y_{nt} - f(X_{nt}; \beta)] W_{nt} \right\} \left\{ \sum_{t=1}^{b_n} W_{nt}' W_{nt} \right\}^{-1} \left\{ \sum_{t=1}^{b_n} W_{nt}' [Y_{nt} - f(X_{nt}; \beta)] \right\}. \quad (5.12)$$

It is to be understood that in (5.12) and in the ensuing development  $X_{nt}$  is replaced by its projection onto the space  $\Gamma$  whenever  $X_{nt}$  is outside of  $\Gamma$ .

## 2. Consistency

In this section, we derive the consistency of the estimator  $\hat{\beta}_n$  of  $\beta^0$  in the model (5.8). We introduce the following assumptions.

Assumption 5.1. The model (5.8) holds with  $\Sigma_n = O(a_n^{-1})$ .

Assumption 5.2.  $a_n = o(n)$  and  $b_n = o(n)$ .

Assumption 5.3.  $\lim_{n \rightarrow \infty} \Sigma_{rrn} = \Sigma_{rr}$ .

Assumption 5.4.  $\lim_{n \rightarrow \infty} b_n^{-1} \sum_{t=1}^{b_n} w_t^0 w_t^0 = \bar{m}_{ww}$ , where  $\bar{m}_{ww}$  is positive definite.

Assumption 5.5. The  $\beta$ -parameter space  $\Omega$  is a compact convex subset of  $k$ -dimensional Euclidean space.

Assumption 5.6. The  $x$ -parameter space  $\Gamma$  is a convex subset of  $q$ -dimensional Euclidean space.

Assumption 5.7. The partial derivatives of order one of  $f(x; \beta)$  with respect to  $x$  and  $\beta$  exist and are continuous on  $\Gamma \times \Omega$ .

Assumption 5.8. There exists a constant  $K_1$  such that for all  $z$  in  $\Gamma$  and all  $\theta$  in  $\Omega$  the absolute value of each element of  $f_x(z; \theta)$  is bounded by  $K_1$ .

Assumption 5.9 Uniformly for all  $\theta$  in  $\Omega$ ,

$$\lim_{n \rightarrow \infty} b_n^{-1} \sum_{t=1}^{b_n} f_{\beta}(x_t^0; \theta) [f_{\beta}(x_t^0; \theta)]' = m_{\beta\beta}(\theta).$$

Assumption 5.10. Uniformly for all  $\theta$  in  $\Omega$ ,

$$\lim_{n \rightarrow \infty} b_n^{-1} \sum_{t=1}^{b_n} f_{\beta}(x_t^0; \theta) w_t^0 = h(\theta),$$

where the rank of the  $k \times p$  matrix  $h(\theta)$  is  $k$  for all  $\theta$  in  $\Omega$ .

Note that by Lemma 2 of Jennrich (1969), under Assumptions 5.5 and 5.7, a measurable function  $\hat{\beta}_n$  of  $Y_{nt}$ ,  $X_{nt}$  and  $W_{nt}$  which minimizes (5.12) always exists. The following theorem shows the consistency of  $\hat{\beta}_n$ .

Theorem 5.1. Let Assumptions 5.1 through 5.10 hold. Then, as  $n \rightarrow \infty$ ,

$$\hat{\beta}_n \xrightarrow{P} \beta^0.$$

Proof. By Assumption 5.7 and Lemma 3 of Jennrich (1969), for any  $\beta$  in  $\Omega$ ,

$$\begin{aligned} Y_{nt} - f(X_{nt}; \beta) &= e_{nt} + f(x_t^0; \beta^0) - f(X_{nt}; \beta) \\ &= e_{nt} + f(x_t^0; \beta^0) - f(x_t^0; \beta) - f_x(x_{nt}^*; \beta) u'_{nt} \\ &= e_{nt} - f'_\beta(x_t^0; \beta^*)(\beta - \beta^0) - f_x(x_{nt}^*; \beta) u'_{nt}, \end{aligned}$$

where  $x_{nt}^*$  is on the segment joining  $x_t^0$  and  $X_{nt}$ , and  $\beta^*$  is on the segment joining  $\beta^0$  and  $\beta$ . Note that by Assumptions 5.5 and 5.6,  $\Gamma$  and  $\Omega$  are convex, and thus  $x_{nt}^*$  is in  $\Gamma$  and  $\beta^*$  is in  $\Omega$ . We have

$$b_n^{-1} \sum_{t=1}^{b_n} [Y_{nt} - f(X_{nt}; \beta)] W_{nt} = b_n^{-1} \sum_{t=1}^{b_n} e_{nt} (w_t^0 + r_{nt})$$

$$\begin{aligned}
& - b_n^{-1} \sum_{t=1}^{b_n} \tilde{f}_{\tilde{x}}(\tilde{x}_{nt}^*; \tilde{\beta}) \tilde{u}_{nt}' (\tilde{w}_t^0 + \tilde{r}_{nt}) \\
& - (\tilde{\beta} - \tilde{\beta}^0)' b_n^{-1} \sum_{t=1}^{b_n} \tilde{f}_{\tilde{\beta}}(\tilde{x}_t^0; \tilde{\beta}^*) (\tilde{w}_t^0 + \tilde{r}_{nt}) .
\end{aligned} \tag{5.13}$$

By Assumptions 5.1 and 5.4, and by the Chebyshev's inequality, as

$n \rightarrow \infty$  ,

$$b_n^{-1} \sum_{t=1}^{b_n} e_{nt} \tilde{w}_t^0 \xrightarrow{P} 0 . \tag{5.14}$$

Also, by Assumptions 5.1 and 5.3, and by the Cauchy-Schwarz inequality,

$$\begin{aligned}
E\left\{ \left| b_n^{-1} \sum_{t=1}^{b_n} e_{nt} \tilde{r}_{nt} \right| \right\} & < b_n^{-1} \sum_{t=1}^{b_n} [E\{e_{nt}^2\} E\{|\tilde{r}_{nt}|^2\}]^{1/2} \\
& = o(a_n^{-1/2}) ,
\end{aligned}$$

where the norm is the Euclidean norm. Thus, by Assumption 5.2, as

$n \rightarrow \infty$  ,

$$b_n^{-1} \sum_{t=1}^{b_n} e_{nt} \tilde{r}_{nt} \xrightarrow{P} 0. \quad (5.15)$$

Let  $f_{xi}(\tilde{x}_{nt}^*; \beta)$ ,  $u_{nti}$ ,  $w_{ti}^0$  and  $r_{nti}$  be the  $i$ -th elements of  $\tilde{f}_x(\tilde{x}_{nt}^*; \beta)$ ,  $\tilde{u}_{nt}$ ,  $\tilde{w}_t^0$  and  $\tilde{r}_{nt}$ , respectively. Also, let  $\sigma_{uunii} = V\{u_{nti}\}$ , and  $\sigma_{rrnii} = V\{r_{nti}\}$ . Then, by Assumptions 5.1, 5.3, 5.4, and 5.8, and the Cauchy-Schwarz inequality, for  $i = 1, 2, \dots, p$ ,

$$\begin{aligned} & E\left\{ \left| b_n^{-1} \sum_{t=1}^{b_n} \tilde{f}_{xi}(\tilde{x}_{nt}^*; \beta) \tilde{u}_{nt}' (w_{ti}^0 + r_{nti}) \right| \right\} \\ & < b_n^{-1} \sum_{t=1}^{b_n} \sum_{i=1}^q E\left\{ \left| \tilde{f}_{xi}(\tilde{x}_{nt}^*; \beta) u_{nti} w_{ti}^0 \right| + \left| \tilde{f}_{xi}(\tilde{x}_{nt}^*; \beta) u_{nti} r_{nti} \right| \right\} \\ & < b_n^{-1} \sum_{t=1}^{b_n} \sum_{i=1}^q K_1^{1/2} [E\{u_{nti}^2\}]^{1/2} (|w_{ti}^0| + [E\{r_{nti}^2\}]^{1/2}) \\ & = K_1^{1/2} \left( \sum_{i=1}^q \sigma_{uunii}^{1/2} \right) (\sigma_{rrnii}^{1/2} + b_n^{-1} \sum_{t=1}^{b_n} |w_{ti}^0|) \\ & < K_1^{1/2} \left( \sum_{i=1}^q \sigma_{uunii}^{1/2} \right) \{ \sigma_{rrnii}^{1/2} + (b_n^{-1} \sum_{t=1}^{b_n} w_{ti}^{02})^{1/2} \}. \end{aligned}$$

$$= M_1 a_n^{-1/2}, \quad (5.16)$$

where  $M_1$  is some constant. Therefore, by Assumption 5.2, for every  $\varepsilon > 0$ , there exists an  $N_1(\varepsilon)$  such that if  $n > N_1(\varepsilon)$  for all  $\beta$  in  $\Omega$ ,

$$P\left\{\left|b_n^{-1} \sum_{t=1}^{b_n} f_x(X_{nt}^*; \beta) u'_{nt} W_{nt}\right| > \varepsilon\right\} < \varepsilon, \quad (5.17)$$

where the norm is the Euclidean norm. By Assumption 5.10, uniformly for all  $\beta$  in  $\Omega$ ,

$$b_n^{-1} \sum_{t=1}^{b_n} f_\beta(x_t^0; \beta^*) w_t^0 \longrightarrow h(\beta^*) . \quad (5.18)$$

Let  $m_{\beta\beta ij}(\theta)$  be the  $(i,j)$ -th element of  $m_{\beta\beta}(\theta)$ , and let  $f_{\beta i}(x_t^0; \beta^*)$  be the  $i$ -th element of  $f_\beta(x_t^0; \beta^*)$ . By Assumptions 5.7 and 5.9,  $m_{\beta\beta ij}(\theta)$  are continuous on  $\Omega$ , since  $m_{\beta\beta ij}(\theta)$  are the uniform limits of continuous functions. Thus, by Assumption 5.5 there exists an  $M_2$  such that for  $i, j = 1, 2, \dots, k$ , and for all  $\theta$  in  $\Omega$ ,

$$m_{\beta\beta ij}(\theta) < M_2 < \infty . \quad (5.19)$$

It follows from (5.19) and Assumption 5.3 that for  $i = 1, 2, \dots, k$ ,  
 $j = 1, 2, \dots, p$  for all  $\xi$  in  $\Omega$ , and for some constant  $M_3 < \infty$ ,

$$E\left\{b_n^{-1} \sum_{t=1}^{b_n} f_{\beta i}(x_t^0; \xi^*) r_{ntj}\right\} = 0,$$

$$V\left\{b_n^{-1} \sum_{t=1}^{b_n} f_{\beta i}(x_t^0; \xi^*) r_{ntj}\right\} < b_n^{-1} M_3. \quad (5.20)$$

Thus, there exists an  $N_2(\epsilon)$  such that for all  $\xi$  in  $\Omega$ ,

$$P\left\{\left|b_n^{-1} \sum_{t=1}^{b_n} \tilde{f}_{\beta}(x_t^0; \xi^*) r_{nt}\right| > \epsilon\right\} < \epsilon. \quad (5.21)$$

Also, by Assumption 5.5,

$$|\xi - \xi^0| < M_4 \quad (5.22)$$

for all  $\xi$  in  $\Omega$  and some constant  $M_4 < \infty$ . It follows from Assumptions 5.2, 5.3, and 5.4 that

$$b_n^{-1} \sum_{t=1}^{b_n} W'_{nt} W_{nt} = \bar{m}_{WW} + o_p(b_n^{-1}) \quad , \quad (5.23)$$

where  $\bar{m}_{WW} = \bar{m}_{WW} + \bar{m}_{RR}$  is positive definite.

Therefore, by (5.13), (5.14), (5.15), (5.17), (5.18), (5.21), (5.22) and (5.23), for every  $\varepsilon > 0$  there exists an  $N_3(\varepsilon)$  such that if  $n > N_3(\varepsilon)$ , then for all  $\beta$  in  $\Omega$

$$P\{|Q_n(\beta) - (\beta - \beta^0)' h(\beta^*) \bar{m}_{WW}^{-1} [h(\beta^*)]' (\beta - \beta^0)| > \varepsilon\} < \varepsilon \quad , \quad (5.24)$$

where  $\beta^*$  is on the segment joining  $\beta^0$  and  $\beta$ . Since  $\hat{\beta}_n$  is in  $\Omega$  for all realizations, it follows from (5.24) that if  $n > N_3(\varepsilon)$  then

$$P\{|Q_n(\hat{\beta}_n) - (\hat{\beta}_n - \beta^0)' h(\beta_n^*) \bar{m}_{WW}^{-1} [h(\beta_n^*)]' (\hat{\beta}_n - \beta^0)| > \varepsilon\} < \varepsilon \quad , \quad (5.25)$$

where  $\beta_n^*$  is on the segment joining  $\beta^0$  and  $\hat{\beta}_n$ . By Assumptions 5.4, 5.5, 5.7, and 5.10, and by the argument used to obtain (5.19), there exists an  $M_5 > 0$  such that for all  $\theta$  in  $\Omega$  the smallest eigenvalue  $\lambda_m(\theta)$  of  $h(\theta) \bar{m}_{WW}^{-1} [h(\theta)]'$  satisfies

$$\lambda_m(\theta) > M_5 \quad . \quad (5.26)$$

By (5.24), as  $n \rightarrow \infty$ ,

$$Q_n(\beta^0) \xrightarrow{P} 0 . \quad (5.27)$$

Since  $\hat{\beta}_n$  minimizes  $Q_n(\beta)$  over  $\Omega$  ,

$$Q_n(\hat{\beta}_n) \leq Q_n(\beta^0) . \quad (5.28)$$

It follows from (5.27) and (5.28) that there is an  $N_4(\varepsilon)$  such that if  $n > N_4(\varepsilon)$  , then

$$P\{|Q_n(\hat{\beta}_n)| > \varepsilon\} \leq P\{|Q_n(\beta^0)| > \varepsilon\} < \varepsilon . \quad (5.29)$$

By (5.25), (5.26), and (5.29), if  $n > \max\{N_3(\varepsilon), N_4(\varepsilon)\}$  , then

$$P\{|\hat{\beta}_n - \beta^0|^2 > 2M_5^{-1}\varepsilon\} \leq P\{(\hat{\beta}_n - \beta^0)' h(\beta^*) \bar{m}_{WW}^{-1} [h(\beta^*)]' (\hat{\beta}_n - \beta) > 2\varepsilon\}$$

$$\leq P\{|(\hat{\beta}_n - \beta^0)' h(\beta^*) \bar{m}_{WW}^{-1} [h(\beta^*)]' (\hat{\beta}_n - \beta^0) - Q_n(\hat{\beta}_n)| > \varepsilon\}$$

$$+ P\{|Q_n(\hat{\beta}_n)| > \varepsilon\}$$

$$\leq 2\varepsilon .$$

Thus, the result holds.  $\square$

Note that in Theorem 5.1 we did not use any relation between  $(e_{nt}, u_{nt})$  and  $x_{nt}$  such as zero correlation or independence. We

recall that in the linear model instrumental variables have to be uncorrelated with errors in the variables contained in the equation to produce a consistent estimator. However, in our model, the errors  $(e_{nt}, u_{nt})$  are decreasing, and thus the correlation between  $(e_{nt}, u_{nt})$  and  $\xi_{nt}$  also tends to zero. This is why we were able to obtain the consistency of  $\hat{\beta}_n$  without an explicit assumption of the zero correlation.

Also, Theorem 5.1 is applicable to both the case with  $\Sigma_{rrn} = O(1)$  and the case with  $\Sigma_{rrn} = O(a_n^{-1})$ , as long as Assumption 5.3 holds.

As we mentioned earlier, the consistency of  $\hat{\beta}_n$  can be proved for the case with  $\sigma_{een} = O(1)$  under an additional assumption of zero correlation between  $e_{nt}$  and  $\xi_{nt}$ . The zero correlation is used only to obtain (5.15). The rest of the proof is exactly the same as the proof of Theorem 5.1.

Under the assumption that the variances of  $(e_{nt}, u_{nt})$  decrease, the ordinary least squares estimator  $\hat{\beta}_L$  obtained by minimizing

$$\sum_{t=1}^{b_n} [Y_{nt} - f(x_{nt}; \beta)]^2$$

is also consistent for  $\beta^0$ . Wolter and Fuller (1982b) showed that under certain assumptions including that  $\Sigma_n = O(a_n^{-1})$ ,

$$\hat{\beta}_L - \beta^0 = O_p(\max[n^{-1/2}, a_n^{-1}]) .$$

### 3. Limiting distribution

We introduce further assumptions.

Assumption 5.11.  $\beta^0$  is an interior point of  $\Omega$ .

Assumption 5.12.  $f_{\beta\beta}(\underline{z}; \underline{\theta})$  and  $f_{\beta x}(\underline{z}; \underline{\theta})$  exist and are continuous on  $\underline{\Gamma} \times \Omega$ .

Assumption 5.13. There exist constants  $K_2$  and  $K_3$  such that uniformly for all  $\underline{z}$  in  $\underline{\Gamma}$  and all  $\underline{\theta}$  in  $\Omega$  the absolute value of each element of  $f_{\beta\beta}(\underline{z}; \underline{\theta})$  is bounded by  $K_2$  and the absolute value of each element of  $f_{\beta x}(\underline{z}; \underline{\theta})$  is bounded by  $K_3$ .

Later, we will use different sets of assumptions to derive the limiting distribution, the order, and the bias expression of  $\hat{\beta}_n$ . A term playing an important role in the derivations is

$$b_n^{-1} \sum_{t=1}^{b_n} [Y_{nt} - f(\underline{x}_{nt}; \beta^0)] W_{nt}.$$

The order of this term varies under different assumptions. Hence, it is convenient to establish an expansion of  $\hat{\beta}_n$  in terms of this quantity. The following Lemma 5.1 provides such an expansion, and will be used repeatedly in later discussions.

Lemma 5.1. Let Assumptions 5.1 through 5.13 hold. Assume that

$$b_n^{-1} \sum_{t=1}^{b_n} [Y_{nt} - f(\underline{x}_{nt}; \beta^0)] W_{nt} = o_p(c_n^{-1}), \quad (5.30)$$

where  $c_n^{-1} = o(1)$ . Then,

$$\hat{\beta}_n - \beta^0 = \{h^0 \bar{m}_{WW}^{-1} h^{0'}\}^{-1} h^0 \bar{m}_{WW}^{-1} \{b_n^{-1} \sum_{t=1}^{b_n} W'_{nt} [Y_{nt} - f(X_{nt}; \beta^0)]\} + o_p(c_n^{-1}),$$

where

$$h^0 = h(\beta^0) \quad ,$$

$$\bar{m}_{WW} = \bar{m}_{ww} + \Sigma_{rr} \quad .$$

and  $h(\theta)$  is defined in Assumption 5.10.

Proof. Let  $f_{\beta i}(z; \theta)$  be the  $i$ -th element of  $f_{\beta}(z; \theta)$  and let  $f_{\beta \beta i}(z; \theta)$  be the  $i$ -th row of  $f_{\beta \beta}(z; \theta)$ . Also let

$$m_{WW} = b_n^{-1} \sum_{t=1}^{b_n} W'_{nt} W_{nt} \quad .$$

By Assumptions 5.5, 5.6, and 5.12, and by Lemma 3 of Jennrich (1969),  
for  $i = 1, 2, \dots, k$ ,

$$\{b_n^{-1} \sum_{t=1}^{b_n} [Y_{nt} - f(X_{nt}; \hat{\beta}_n)] W_{nt}\} m_{WW}^{-1} \{b_n^{-1} \sum_{s=1}^{b_n} W'_{ns} f_{\beta i}(X_{ns}; \hat{\beta}_n)\}$$

$$\begin{aligned}
&= \left\{ b_n^{-1} \sum_{t=1}^{b_n} [Y_{nt} - f(X_{nt}; \beta^0)] W_{nt} \right\} m_{WW}^{-1} \left\{ b_n^{-1} \sum_{t=1}^{b_n} W'_{nt} f_{\beta i}(X_{nt}; \beta^0) \right\} \\
&\quad + \left\{ \left[ b_n^{-1} \sum_{t=1}^{b_n} [Y_{nt} - f(X_{nt}; \beta_n^*)] W_{nt} \right] m_{WW}^{-1} \left\{ b_n^{-1} \sum_{s=1}^{b_n} W'_{ns} f_{\beta i}(X_{ns}; \beta_n^*) \right\} \right. \\
&\quad \left. - \left\{ b_n^{-1} \sum_{s=1}^{b_n} f_{\beta i}(X_{ns}; \beta_n^*) W_{ns} \right\} m_{WW}^{-1} \left\{ b_n^{-1} \sum_{t=1}^{b_n} W'_{nt} f'_{\beta}(X_{nt}; \beta_n^*) \right\} \right] (\hat{\beta}_n - \beta^0) \right\} \\
&\hspace{15em} (5.31)
\end{aligned}$$

where  $\beta_n^*$  is on the segment joining  $\beta^0$  and  $\hat{\beta}_n$ . The estimator  $\hat{\beta}_n$  minimizes  $Q_n(\beta)$  in (5.12) over  $\Omega$ , and  $Q_n(\beta)$  is differentiable on  $\Omega$  by Assumption 5.8. Thus, if  $\hat{\beta}_n$  is an interior point of  $\Omega$ , then

$$\begin{aligned}
\frac{\partial Q_n(\hat{\beta}_n)}{\partial \beta} &= -2 \left\{ b_n^{-1} \sum_{t=1}^{b_n} f_{\beta}(X_{nt}; \hat{\beta}_n) W_{nt} \right\} m_{WW}^{-1} \left\{ b_n^{-1} \sum_{s=1}^{b_n} [Y_{ns} - f(X_{ns}; \hat{\beta}_n)] W_{ns} \right\}' \\
&= 0.
\end{aligned} \hspace{15em} (5.32)$$

By Assumption 5.11, there is a  $\delta > 0$  such that if  $|\beta - \beta^0| < \delta$ , then  $\beta$  is an interior point of  $\Omega$ . Since  $\hat{\beta}_n \xrightarrow{P} \beta^0$  by Theorem 5.1, for every  $\varepsilon > 0$  given, there is an  $N_1$  such that if  $n > N_1$ , then

$$P\{|\hat{\beta}_n - \beta^0| < \delta\} > 1 - \varepsilon. \quad (5.33)$$

By (5.32) and (5.33), for every  $\varepsilon > 0$  and every  $\alpha > 0$ , if  $n > N_1$ , then

$$\begin{aligned} P\left\{n^\alpha \left| \frac{\partial Q_n(\hat{\beta}_n)}{\partial \beta} \right| < \varepsilon\right\} &> P\left\{ \frac{\partial Q_n(\hat{\beta}_n)}{\partial \beta} = 0 \right\} \\ &> P\{|\hat{\beta}_n - \beta^0| < \delta\} \\ &> 1 - \varepsilon. \end{aligned}$$

Hence, for every  $\alpha > 0$ ,

$$\left\{ b_n^{-1} \sum_{t=1}^{b_n} f_\beta(X_{nt}; \hat{\beta}_n) W_{nt} \right\}_{WW}^{-1} \left\{ b_n^{-1} \sum_{t=1}^{b_n} [Y_{nt} - f(X_{nt}; \hat{\beta}_n)] W_{nt} \right\}' = o_p(n^{-\alpha}). \quad (5.34)$$

It follows from (5.24) in the proof of Theorem 5.1 that

$$Q_n(\beta_n^*) - (\beta_n^* - \beta^0)' h(\beta^{**}) \bar{m}_{WW}^{-1} [h(\beta^{**})]' (\beta_n^* - \beta^0) = o_p(1), \quad (5.35)$$

where  $\beta_n^{**}$  is on the segment joining  $\beta^0$  and  $\beta_n^*$ . Since by Theorem 5.1  $\beta_n^* = \beta^0 + o_p(1)$ , we obtain from (5.26) and (5.35) that

$$Q_n(\beta_n^*) = o_p(1).$$

Hence, by (5.23) and the definition (5.12) of  $Q_n(\beta)$ ,

$$b_n^{-1} \sum_{t=1}^{b_n} [Y_{nt} - f(\underline{X}_{nt}; \beta_n^*)] W_{nt} = o_p(1). \quad (5.36)$$

Letting  $W_{nsm}$  be the  $m$ -th element of  $\underline{W}_{ns}$  and  $f_{\beta\beta ij}(\underline{X}_{ns}; \beta_n^*)$  be the  $(i,j)$ -th element of  $\underline{f}_{\beta\beta}(\underline{X}_{ns}; \beta_n^*)$ , we have by Assumptions 5.4 and 5.13,

$$\begin{aligned} \left| b_n^{-1} \sum_{s=1}^{b_n} W_{nsm} f_{\beta\beta ij}(\underline{X}_{ns}; \beta_n^*) \right| &< b_n^{-1} \sum_{s=1}^{b_n} |W_{nsm}| K_2 \\ &< K_2 (b_n^{-1} \sum_{s=1}^{b_n} W_{nsm}^2)^{1/2} \\ &= K_2. \end{aligned}$$

Thus,

$$b_n^{-1} \sum_{s=1}^{b_n} \underline{W}'_{ns} \underline{f}_{\beta\beta i}(\underline{X}_{ns}; \beta^0) = o_p(1). \quad (5.37)$$

By Assumption 5.12,

$$b_n^{-1} \sum_{t=1}^{b_n} f_{\beta i}(\underline{x}_{nt}; \underline{\beta}_n^*) \underline{w}_{nt} = b_n^{-1} \sum_{t=1}^{b_n} [f_{\beta i}(\underline{x}_t^0; \underline{\beta}_n^*) (\underline{w}_t^0 + \underline{r}_{nt}) + f_{\beta xi}(\underline{x}_{nt}^{**}; \underline{\beta}_n^*) \underline{u}'_{nt} \underline{w}_{nt}], \quad (5.38)$$

where  $\underline{x}_{nt}^{**}$  is on the segment joining  $\underline{x}_t^0$  and  $\underline{x}_{nt}$ , and  $f_{\beta xi}$  is the  $i$ -th row of  $f_{\beta x}$ . By Assumptions 5.4 and 5.13, and the argument used in (5.16),

$$b_n^{-1} \sum_{t=1}^{b_n} f_{\beta xi}(\underline{x}_{nt}^{**}; \underline{\beta}_n^*) \underline{u}'_{nt} \underline{w}_{nt} = o_p(a_n^{-1/2}). \quad (5.39)$$

By (5.20) in the proof of Theorem 5.1,

$$b_n^{-1} \sum_{t=1}^{b_n} f_{\beta i}(\underline{x}_t^0; \underline{\beta}_n^*) \underline{r}_{nt} = o_p(b_n^{-1/2}). \quad (5.40)$$

By Assumption 5.10, for every  $\varepsilon > 0$ , there exists an  $N_1(\varepsilon)$  such that if  $n > N_1(\varepsilon)$ , then for all  $\underline{\beta}$  in  $\Omega$ ,

$$|\text{vec}[b_n^{-1} \sum_{t=1}^{b_n} f_{\beta}(\underline{x}_t^0; \underline{\beta}) \underline{w}_t^0 - \underline{h}(\underline{\beta})]| < 2^{-1} \varepsilon,$$

and, thus,

$$\left| \text{vec} \left[ b_n^{-1} \sum_{t=1}^{b_n} \tilde{f}_\beta(x_t^0; \beta_n^*) \tilde{w}_t^0 - \tilde{h}(\beta_n^*) \right] \right| < 2^{-1} \varepsilon, \quad (5.41)$$

where we have used the convexity of  $\Omega$ . Since  $\tilde{h}(\theta)$  is the uniform limit of continuous functions,  $\tilde{h}(\theta)$  is continuous. Since

$\beta_n^* \xrightarrow{P} \beta^0$  by Theorem 5.1, there exists an  $N_2(\varepsilon)$  such that if  $n > N_2(\varepsilon)$ , then

$$P\{|\tilde{h}(\beta_n^*) - \tilde{h}(\beta^0)| > 2^{-1} \varepsilon\} < \varepsilon. \quad (5.42)$$

By (5.41) and (5.42), if  $n > \max[N_1(\varepsilon), N_2(\varepsilon)]$ , then

$$\begin{aligned} & P\left\{ \left| b_n^{-1} \sum_{t=1}^{b_n} \tilde{f}_\beta(x_t^0; \beta_n^*) \tilde{w}_t^0 - \tilde{h}(\beta^0) \right| > \varepsilon \right\} \\ & \leq P\left\{ \left| b_n^{-1} \sum_{t=1}^{b_n} \tilde{f}_\beta(x_t^0; \beta_n^*) \tilde{w}_t^0 - \tilde{h}(\beta_n^*) \right| > 2^{-1} \varepsilon \right\} + P\left\{ |\tilde{h}(\beta_n^*) - \tilde{h}(\beta^0)| > 2^{-1} \varepsilon \right\} \\ & < \varepsilon. \end{aligned}$$

Therefore,

$$b_n^{-1} \sum_{t=1}^{b_n} \tilde{f}_\beta(\tilde{x}_t^0; \beta^*) \tilde{w}_t^0 = \tilde{h}(\beta^0) + o_p(1) . \quad (5.43)$$

By (5.38), (5.39), (5.40), and (5.43),

$$b_n^{-1} \sum_{t=1}^{b_n} \tilde{f}_\beta(\tilde{x}_{nt}; \beta_n^*) \tilde{w}_{nt} = \tilde{h}^0 + o_p(1) . \quad (5.44)$$

Similarly,

$$b_n^{-1} \sum_{t=1}^{b_n} \tilde{f}_\beta(\tilde{x}_{nt}; \beta^0) \tilde{w}_{nt} = \tilde{h}^0 + o_p(1) . \quad (5.45)$$

Note that by Assumptions 5.4 and 5.10  $\tilde{h}^0 \bar{m}_{WW}^{-1} \tilde{h}^0$  is nonsingular.

Hence, by (5.23), (5.31), (5.34), (5.36), (5.37), (5.44), (5.45), and

Theorem 5.1, the result holds.  $\square$

Note that if  $\Sigma_{rrn} = O(a_n^{-1})$  in Lemma 5.1 then  $\Sigma_{rr} = 0$  and thus  $\bar{m}_{WW} = \bar{m}_{ww}$  .

In the linear instrumental variable estimation, it is assumed that the instrumental variables are either independent of, or uncorrelated with, errors in the variables contained in the relationship. In the

derivation of the limiting distribution of  $\hat{\beta}_n$ , the argument is somewhat simpler if  $\tilde{x}_{nt}$  are independent of  $(e_{nt}, u_{nt})$ . In practice, instrumental variables are often measured independently of other variables, and thus the assumption of independent errors holds. However, there are situations where we are not willing to assume the independence of  $\tilde{x}_{nt}$  and  $(e_{nt}, u_{nt})$ . Without the independence, some assumptions on the higher moments of  $\tilde{x}_{nt}$  and  $(e_{nt}, u_{nt})$  are required to obtain the limiting distribution of  $\hat{\beta}_n$ .

We introduce assumptions which will be used to derive the limiting distribution of  $\hat{\beta}_n$ .

Assumption 5.14.  $f_{xx}(z; \theta)$  exists and is continuous on  $\Gamma \times \Omega$ .

Assumption 5.15. There exists a constant  $K_4$  such that the absolute value of each element of  $f_{xx}(z; \theta)$  is bounded by  $K_4$  for all  $z$  in  $\Gamma$  and all  $\theta$  in  $\Omega$ .

Assumption 5.16. Let  $w_{ti}^0$  and  $f_{xi}(x_t^0; \beta^0)$  be the  $i$ -th elements of  $w_t^0$  and  $f_x(x_t^0; \beta^0)$ , respectively. Then, for all  $i_1, i_2 = 1, 2, \dots, p$ , and for all  $j_1, j_2 = 1, 2, \dots, q$ ,

$$\lim_{n \rightarrow \infty} b_n^{-1} \sum_{t=1}^{b_n} w_{ti_1}^0 w_{ti_2}^0 f_{xj_1}(x_t^0; \beta^0) f_{xj_2}(x_t^0; \beta^0) = \bar{m}_4(i_1, i_2, j_1, j_2).$$

Also,

$$\lim_{n \rightarrow \infty} a_n \sum_{n \sim n} = \Phi.$$

Assumption 5.17.  $E\{|a_n^{1/2}(e_{nt}, u_{nt})|^4\} = O(1)$ .

Assumption 5.18.  $E\{|\xi_{nt}|^4\} = O(1)$  . Also, for  $i = 1, 2, \dots, p$ ,

$$\lim_{n \rightarrow \infty} a_n V\{r_{nti}(e_{nt}, u_{nt})'\} = \psi_i .$$

The next theorem gives the limiting distribution of  $\hat{\beta}_n$  with the normalizing constant  $n^{1/2}$  .

Theorem 5.2. Let Assumptions 5.1 through 5.16 hold. Assume that

$$a_n^{-1} = o(n^{-1/2}) . \quad (5.46)$$

Also, assume either

- (a)  $(e_{nt}, u_{nt})$  and  $\xi_{nt}$  are independent, or
- (b)  $(e_{nt}, u_{nt})$  and  $\xi_{nt}$  are uncorrelated, and Assumptions 5.17 and 5.18 hold.

Then,

$$n^{1/2}(\hat{\beta}_n - \beta^0) \xrightarrow{L} N(0, V) ,$$

where

$$V = \{h^0 \bar{m}_{WW}^{-1} h^{0'}\}^{-1} h^0 \bar{m}_{WW}^{-1} v_{0\bar{m}_{WW}}^{-1} h^{0'} \{h^0 \bar{m}_{WW}^{-1} h^{0'}\}^{-1} ,$$

and for case (a),

$$\tilde{y}_0 = \text{plim}_{n \rightarrow \infty} b_n \sum_{t=1}^n a_n \sigma_{vnt}^2 \tilde{w}'_{nt} \tilde{w}_{nt} ,$$

$$\sigma_{vnt}^2 = [1, -\tilde{f}_{xt}^0] \tilde{\Sigma}_n [1, -\tilde{f}_{xt}^0]' ,$$

and for case (b),

$$\tilde{y}_0 = \lim_{n \rightarrow \infty} b_n^{-1} \sum_{t=1}^n V[\tilde{w}'_{nt} [1, -\tilde{f}_x(\tilde{x}_t^0; \beta^0)] a_n^{1/2} (e_{nt}, \tilde{u}_{nt})'] .$$

Proof. By Assumption 5.14,

$$Y_{nt} - f(\tilde{x}_{nt}; \beta^0) = e_{nt} - \tilde{f}_x(\tilde{x}_t^0; \beta^0) \tilde{u}_{nt} - \frac{1}{2} \tilde{u}_{nt} \tilde{f}_{xx}(\tilde{x}_{nt}^*; \beta^0) \tilde{u}_{nt}' , \quad (5.47)$$

where  $\tilde{x}_{nt}^*$  is on the segment joining  $\tilde{x}_t^0$  and  $\tilde{x}_{nt}$ . Let  $f_{xxij}$  be the  $(i,j)$ -th element of  $\tilde{f}_{xx}$ , let  $u_{nti}$  be the  $i$ -th element of  $\tilde{u}_{nt}$ , and let  $(w_{ntm}^0 + r_{ntm})$  be the  $m$ -th element of  $\tilde{w}_{nt} = \tilde{w}_t^0 + \tilde{r}_{nt}$ . Then, for either the case (a) or (b), and for  $i, j = 1, 2, \dots, q$ , and  $m = 1, 2, \dots, p$ ,

$$E\left\{ \left| b_n^{-1} \sum_{t=1}^n f_{xxij}(\tilde{x}_{nt}^*; \beta^0) u_{nti} u_{ntj} w_{ntm}^0 \right| \right\}$$

$$\begin{aligned}
& \leq K_4 b_n^{-1} \sum_{t=1}^{b_n} |w_{tm}^0| E\{|u_{nti} u_{ntj}|\} \\
& \leq K_4 \{b_n^{-1} \sum_{t=1}^{b_n} (w_{tm}^0)^2\}^{1/2} [E\{u_{nti}^2\}]^{1/2} [E\{u_{ntj}^2\}]^{1/2} \\
& = o_p(a_n^{-1}) .
\end{aligned}$$

Thus,

$$b_n^{-1} \sum_{t=1}^{b_n} u_{nt} f_{xx}(X_{nt}^*; \beta^0) u_{nt}' w_t^0 = o_p(a_n^{-1}) . \quad (5.48)$$

For case (a), by the independence of  $u_{nt}$  and  $r_{nt}$  and Assumption 5.15,

$$\begin{aligned}
E\{|b_n^{-1} \sum_{t=1}^{b_n} f_{xxij}(X_{nt}^*; \beta^0) u_{nti} u_{ntj} r_{ntm}|\} & \leq K_4 b_n^{-1} \sum_{t=1}^{b_n} E\{|u_{nti} u_{ntj}|\} E\{|r_{ntm}|\} \\
& \leq K_4 [E\{u_{nti}^2\} E\{u_{ntj}^2\} E\{r_{ntm}^2\}]^{1/2} \\
& = o_p(a_n^{-1}) .
\end{aligned} \quad (5.49)$$

For case (b), by Assumption 5.17,

$$\begin{aligned}
 E\left\{\left|b_n^{-1} \sum_{t=1}^{b_n} f_{xxij}(\tilde{x}_{nt}^*; \beta^0) u_{nti} u_{ntj} r_{ntm}\right|\right\} &\leq K_4 b_n^{-1} \sum_{t=1}^{b_n} E\left\{\left|u_{nti} u_{ntj} r_{ntm}\right|\right\} \\
 &\leq K_4 [E\{u_{nti}^2 u_{ntj}^2\} E\{r_{ntm}^2\}]^{1/2} . \\
 &= o(a_n^{-1}) . \tag{5.50}
 \end{aligned}$$

By (5.49) and (5.50), for both cases (a) and (b),

$$b_n^{-1} \sum_{t=1}^{b_n} u_{nt} f_{xx}(\tilde{x}_{nt}^*; \beta^0) u_{nt}' r_{nt} = o_p(a_n^{-1}) . \tag{5.51}$$

For case (a), by the independence of  $(e_{nt}, u_{nt})$  and  $r_{nt}$ , and by Assumption 5.8,

$$\begin{aligned}
 E\left\{\left|b_n^{-1} \sum_{t=1}^{b_n} [e_{nt} - f_x(\tilde{x}_t^0; \beta^0) u_{nt}'] r_{nt}\right|^2\right\} \\
 = b_n^{-2} \sum_{t=1}^{b_n} E\{[e_{nt} - f_x(\tilde{x}_t^0; \beta^0) u_{nt}']^2\} E\{r_{nt}^2\} = o(n^{-1}) . \tag{5.52}
 \end{aligned}$$

For case (b), by the zero correlation, and Assumptions 5.8, 5.17, and 5.18,

$$\begin{aligned}
 & E\left\{ \left[ b_n^{-1} \sum_{t=1}^{b_n} [e_{nt} - f_x(x_t^0; \beta^0) u'_{nt}] r_{ntm} \right]^2 \right\} \\
 &= b_n^{-2} \sum_{t=1}^{b_n} E\{ [e_{nt} - f_x(x_t^0; \beta^0) u'_{nt}]^2 r_{ntm}^2 \} \\
 &< b_n^{-2} \sum_{t=1}^{b_n} [E\{ [e_{nt} - f_x(x_t^0; \beta^0) u'_{nt}]^4 \} E\{ r_{ntm}^4 \}]^{1/2} \\
 &= o(n^{-1}) \quad . \quad (5.53)
 \end{aligned}$$

Since  $a_n^{-1} = o(n^{-1/2})$ , by (5.47), (5.48), (5.51), (5.52), (5.53), and Assumption 5.4, for both cases (a) and (b),

$$\begin{aligned}
 & b_n^{-1} \sum_{t=1}^{b_n} w'_{nt} [Y_{nt} - f(x_{nt}; \beta^0)] \\
 &= b_n^{-1} \sum_{t=1}^{b_n} (w_t^0 + r_{nt})' [1, -f_x(x_t^0; \beta^0)] (e_{nt}, u_{nt})' + o_p(n^{-1/2}) \\
 &= o_p(n^{-1/2}) \quad (5.54)
 \end{aligned}$$

We observe that each element of

$$(w_t^0 + \varepsilon_{nt})' [1, -f_x(x_t^0; \beta^0)](e_{nt}, u_{nt})'$$

is a linear combination of elements of a random vector

$\xi_{nt}^* = (e_{nt}, u_{nt}, e_{nt}\varepsilon_{nt}, u_{nt}\varepsilon_{nt})$ . If we let  $A_t$  be the coefficient matrix of the linear combinations, then by Assumptions 5.4 and 5.16, the limit

$$\lim_{n \rightarrow \infty} b_n^{-1} \sum_{t=1}^{b_n} \text{vec } A_t (\text{vec } A_t)' = H$$

exists. Also,  $a_n^{1/2} \xi_{nt}^*$  are independently and identically distributed.

For case (a), by the independence and Assumption 5.16, the mean of

$a_n^{1/2} \xi_{nt}^*$  is zero, and the covariance matrix of  $a_n^{1/2} \xi_{nt}^*$  converges to some matrix. For case (b), by the zero correlation and Assumption 5.18, the mean of  $a_n^{1/2} \xi_{nt}^*$  is zero, and the covariance matrix of  $a_n^{1/2} \xi_{nt}^*$  converges to some matrix. Hence, by Theorem 2.12, the limit

$$V_0 = \lim_{n \rightarrow \infty} b_n^{-1} V\{(w_t^0 + \varepsilon_{nt})' [1, -f_x(x_t^0; \beta^0)] a_n^{1/2} (e_{nt}, u_{nt})'\}$$

exists and

$$n^{1/2} b_n^{-1} \sum_{t=1}^{b_n} (\tilde{w}_t^0 + \tilde{\varepsilon}_{nt})' [1, -\tilde{f}_x(\tilde{x}_t^0; \tilde{\beta}^0)] (e_{nt}, \tilde{u}_{nt})' \xrightarrow{L} N(0, \tilde{V}_0) . \quad (5.55)$$

Note that for case (a), by the independence and Assumption 5.3,

$$\tilde{V}_0 = \lim_{n \rightarrow \infty} [b_n^{-1} \sum_{t=1}^{b_n} a_n \sigma_{vnt}^2 \tilde{w}_t^0 \tilde{w}_t^0 + b_n^{-1} \sum_{t=1}^{b_n} a_n \sigma_{vnt}^2 \tilde{\varepsilon}_{rrn}] .$$

Hence, the result follows from (5.54), (5.55), and Lemma 5.1.  $\square$

If  $\tilde{\varepsilon}_{rrn} = o(1)$ , then the result of Theorem 5.2 holds with  $\tilde{\varepsilon}_{rr} = 0$ . It can be shown that if  $\tilde{\varepsilon}_{rrn} = O(a_n^{-1})$  then the limiting result for  $n^{1/2}(\hat{\beta}_n - \beta^0)$  is valid under either the assumption of independence between  $\tilde{u}_{nt}$  and  $\tilde{\varepsilon}_{nt}$  or the assumption that

$$E\{|a_n^{1/2} \tilde{u}_{nt}|^4\} = O(1) .$$

That is, if the variance of the random part  $\tilde{\varepsilon}_{nt}$  is decreasing, then with the higher moment assumption,  $\tilde{w}_{nt}$  need not be uncorrelated with  $(e_{nt}, \tilde{u}_{nt})$  to produce an asymptotically normal estimator. In practice, we recommend that the zero correlation or the independence of  $\tilde{w}_{nt}$  and  $(e_{nt}, \tilde{u}_{nt})$  be examined before applying the result of Theorem 5.2.

The following theorem provides a consistent estimator of the covariance matrix of the limiting distribution of  $\hat{\beta}_n$ . The estimator

is identical for the case with independence of  $\mathbf{r}_{nt}$  and  $(e_{nt}, \mathbf{u}_{nt})$  and the case only with zero correlation between  $\mathbf{r}_{nt}$  and  $(e_{nt}, \mathbf{u}_{nt})$ . The consistency of the covariance matrix estimator requires stronger assumptions than those used to obtain the limiting distribution.

Assumption 5.19. There exists a constant  $K_5$  such that the absolute value of each element of  $\mathbf{f}_\beta(\mathbf{x}_t^0; \boldsymbol{\beta}^0)$  is bounded by  $K_5$  for all  $t$ .

Assumption 5.20. For  $i = 1, 2, \dots, p$ ,

$$b_n^{-1} \sum_{t=1}^{b_n} (w_{ti}^0)^4 = o(1) \quad .$$

The following theorem repeats the results in Theorem 5.2 under stronger assumptions, and presents a consistent estimator of the covariance matrix of the limiting distribution of  $\hat{\boldsymbol{\beta}}_n$ .

Theorem 5.3. Let Assumptions 5.1 through 5.20 hold. Assume that

$$a_n^{-1} = o(n^{-1/2}) \quad .$$

Also, assume either

- (a)  $(e_{nt}, \mathbf{u}_{nt})$  and  $\mathbf{r}_{nt}$  are independent, or
- (b)  $(e_{nt}, \mathbf{u}_{nt})$  and  $\mathbf{r}_{nt}$  are uncorrelated, and for  $i = 1, 2, \dots, p$ ,

$$E\left\{\left|a_n^{1/2}(e_{nt}, \mathbf{u}_{nt}) \mathbf{r}_{nti}\right|^4\right\} = o(1) \quad . \quad (5.56)$$

Then,

$$n^{1/2} (\hat{\beta}_n - \beta^0) \xrightarrow{L} N(0, \underline{V}) ,$$

and

$$n \hat{\underline{V}}(\hat{\beta}_n) \xrightarrow{P} \underline{V} ,$$

where

$$\hat{\underline{V}}(\hat{\beta}_n) = \{\hat{\underline{h}} \underline{m}_{WW}^{-1} \hat{\underline{h}}'\}^{-1} \hat{\underline{h}} \underline{m}_{WW}^{-1} \hat{\underline{V}}_O \underline{m}_{WW}^{-1} \hat{\underline{h}}' \{\hat{\underline{h}} \underline{m}_{WW} \hat{\underline{h}}'\}^{-1} ,$$

$$\hat{\underline{h}} = b_n^{-1} \sum_{t=1}^{b_n} \underline{f}_{\beta}(x_{nt}, \hat{\beta}_n) \underline{w}_{nt} ,$$

$$\underline{m}_{WW} = b_n^{-1} \sum_{t=1}^{b_n} \underline{w}_{nt}' \underline{w}_{nt} ,$$

$$\hat{\underline{V}}_O = b_n^{-2} \sum_{t=1}^{b_n} \underline{w}_{nt}' \underline{w}_{nt} [Y_{nt} - f(x_{nt}; \hat{\beta}_n)]^2 .$$

Proof. By the argument used to obtain (5.44),

$$\hat{\underline{h}} = \underline{h}(\beta^0) + o_p(1) . \quad (5.57)$$

By Assumption 5.4 and the law of large numbers,

$$\bar{m}_{WW} = \bar{m}_{WW} + o_p(1) . \quad (5.58)$$

We observe that

$$\begin{aligned} Y_{nt} - f(\tilde{x}_{nt}; \hat{\beta}_n) &= e_{nt} - \tilde{f}_x(\tilde{x}_t^0; \beta^0) \tilde{u}_{nt}' - \tilde{f}_\beta'(\tilde{x}_t^0; \beta^0) (\hat{\beta}_n - \beta^0) \\ &\quad - \frac{1}{2} \tilde{u}_{nt} \tilde{f}_{xx}(\tilde{x}_{nt}^*; \beta_n^*) \tilde{u}_{nt}' - \frac{1}{2} (\hat{\beta}_n - \beta^0)' \tilde{f}_{\beta\beta}(\tilde{x}_{nt}^*; \beta_n^*) (\hat{\beta}_n - \beta^0) \\ &\quad - (\hat{\beta}_n - \beta^0)' \tilde{f}_{\beta x}(\tilde{x}_{nt}^*; \beta_n^*) \tilde{u}_{nt}' , \end{aligned} \quad (5.59)$$

where  $\tilde{x}_{nt}^*$  is on the segment joining  $\tilde{x}_t^0$  and  $\tilde{x}_{nt}$ , and  $\beta_n^*$  is on the segment joining  $\beta^0$  and  $\hat{\beta}_n$ . By Theorem 5.2,

$$\hat{\beta}_n - \beta^0 = o_p(n^{-1/2}) . \quad (5.60)$$

Using (5.59) and (5.60), we obtain for both the cases (a) and (b),

$$\begin{aligned} b_n \hat{y}_o &= b_n^{-1} \sum_{t=1}^{b_n} \tilde{w}_{nt}' \tilde{w}_{nt} [Y_{nt} - f(\tilde{x}_{nt}; \hat{\beta}_n)]^2 \\ &= b_n^{-1} \sum_{t=1}^{b_n} \tilde{w}_{nt}' \tilde{w}_{nt} [e_{nt} - \tilde{f}_x(\tilde{x}_t^0; \beta^0) \tilde{u}_{nt}']^2 + o_p(n^{-1/2} a_n^{-1/2}) , \end{aligned} \quad (5.61)$$

where we have used, for example, if (a) holds, then

$$E\left\{b_n^{-1} \sum_{t=1}^{b_n} \left| r_{nt} r_{nti} u_{nt} f_{xx}(X_{nt}^*; \beta_n^*) u'_{nt} f'_{\beta}(x_t^0; \beta^0) (\hat{\beta}_n - \beta^0) \right| \right\}$$

$$= O(\max[n^{-1/2} a_n^{-1}, a_n^{-2}]) ,$$

and if (b) holds, then

$$E\left\{b_n^{-1} \sum_{t=1}^{b_n} \left| r_{nt} r_{nti} [u_{nt} f_{xx}(X_{nt}^*; \hat{\beta}_n^*) u'_{nt}]^2 \right| \right\}$$

$$\leq b_n^{-1} \sum_{t=1}^{b_n} K_4^2 [ |r_{nt} r_{nti} u_{nt} u'_{nt}|^2 ] E\{ |u_{nt} u'_{nt}|^2 \}^{1/2}$$

$$= O(a_n^{-3/2})$$

$$= o(n^{-1/2} a_n^{-1/2}) .$$

We note that by Theorem 5.2  $\underline{V}$  can always be written in the form of

$$\{ \underline{h}^0 \underline{m}_{WW}^{-1} \underline{h}^{0'} \}^{-1} \underline{h}^0 \underline{m}_{WW}^{-1} \underline{V}_O \underline{m}_{WW}^{-1} \underline{h}^{0'} \{ \underline{h}^0 \underline{m}_{WW}^{-1} \underline{h}^{0'} \}^{-1} ,$$

where

$$\underline{v}_0 = \lim_{n \rightarrow \infty} a_n b_n^{-1} \sum_{t=1}^{b_n} E\{\tilde{W}'_{nt} \tilde{W}_{nt} [e_{nt} - \tilde{f}_x(\tilde{x}_t^0; \beta^0) \tilde{u}'_{nt}]^2\} . \quad (5.62)$$

For case (a), by the independence and Assumption 5.17, 5.18, and 5.20,

$$V\{a_n b_n^{-1} \sum_{t=1}^{b_n} \tilde{W}'_{nt} \tilde{W}_{nt} [e_{nt} - \tilde{f}_x(\tilde{x}_t^0; \beta^0) \tilde{u}'_{nt}]^2\} = O(b_n^{-1}) . \quad (5.63)$$

For case (b), by (5.56) and Assumptions 5.17, 5.18, and 5.20, the expression (5.63) holds. By (5.62) and (5.63),

$$a_n b_n^{-1} \sum_{t=1}^{b_n} \tilde{W}'_{nt} \tilde{W}_{nt} [e_{nt} - \tilde{f}_x(\tilde{x}_t^0; \beta^0)]^2 = \underline{v}_0 + o_p(b_n^{-1/2}) . \quad (5.64)$$

By (5.61) and (5.64),

$$n \hat{\underline{v}}_0 = a_n b_n^{-1} \sum_{t=1}^{b_n} \tilde{W}'_{nt} \tilde{W}_{nt} [Y_{nt} - f(\tilde{x}_{nt}; \hat{\beta}_n)]^2$$

$$= \underline{y}_0 + o_p(b_n^{-1/2}) \quad . \quad (5.65)$$

The result follows from (5.57), (5.58), (5.65), and Theorem 5.2.  $\square$

We have shown that the estimator of the variance of the approximate distribution of  $\hat{\beta}_n$  is the same for both the cases (a) and (b) in Theorem 5.3. Thus, the inference procedure based on the asymptotic theory is the same for both the cases. As we have seen, the proofs for the case without independence of  $(e_{nt}, u_{nt})$  and  $\underline{r}_{nt}$  are somewhat more tedious and troublesome than the cases with the independence. However, all the results holding for the independent case still hold for the case without the independence under the assumptions of zero correlation and the existence of higher moments for the errors.

In later sections, we will consider estimation of the error covariance matrix  $\underline{\Sigma}$  and the true values  $\underline{x}_t^0$  based on  $\hat{\beta}_n$ , and will derive an improved estimator of  $\beta^0$  using  $\hat{\beta}_n$  as an initial estimator. Therefore, it is convenient to derive the order of  $(\hat{\beta}_n - \beta^0)$  without the assumptions which are needed to obtain the limiting distribution of  $\hat{\beta}_n$ .

Theorem 5.4. Let Assumptions 5.1 through 5.15 hold. Assume either

(a)  $(e_{nt}, u_{nt})$  and  $\underline{r}_{nt}$  are independent, or

(b)  $E\{|a_{nt} u_{nt}|^4\} = O(1)$  .

Then,

$$\hat{\xi}_n - \xi^0 = o_p(\max[n^{-1/2}, a_n^{-1}]) .$$

Proof. For case (a), the result follows from (5.47), (5.48), (5.51), (5.52), and Lemma 5.1. For case (b), we observe that by Assumption 5.8, for  $m = 1, 2, \dots, p$ ,

$$\begin{aligned} & E\left\{\left|b_n^{-1} \sum_{t=1}^{b_n} [e_{nt} - f_x(x_t^0; \xi^0)u'_{nt}]r_{ntm}\right|\right\} \\ & \leq b_n^{-1} \sum_{t=1}^{b_n} E\{|e_{nt}r_{ntm}|\} + b_n^{-1} \sum_{t=1}^{b_n} \sum_{i=1}^q |f_{xi}(x_t^0; \xi^0)| E\{|u_{nti}r_{ntm}|\} \\ & \leq [E\{e_{nt}^2\}E\{r_{ntm}^2\}]^{1/2} + K_1 \sum_{i=1}^q [E\{u_{nti}^2\}E\{r_{ntm}^2\}]^{1/2} \\ & = o(a_n^{-1}) , \end{aligned}$$

where  $f_{xi}$  is the  $i$ -th element of  $f_x$ . Thus,

$$b_n^{-1} \sum_{t=1}^{b_n} [e_{nt} - f_x(x_t^0; \xi^0)u'_{nt}]r_{nt} = o_p(a_n^{-1}) . \quad (5.66)$$

By (5.47), (5.48), (5.50), (5.66), and Lemma 5.1, the result also holds for the case (b).  $\square$

#### 4. Bias

In Chapter IV, we discussed the bias of the maximum likelihood type estimator of the nonlinear functional relationship model with known error covariance matrix. The bias was of  $O(a_n^{-1})$ , and was due to the nonlinearity of the functional form. The instrumental variable estimator  $\hat{\beta}_n$  in this chapter does not adjust for the bias due to the curvature. Thus, we expect the estimator  $\hat{\beta}_n$  to have a bias which has a larger order than  $n^{-1}$ . In order to simplify our derivation of the asymptotic bias of  $\hat{\beta}_n$ , we introduce the following assumptions.

Assumption 5.21. Let  $f_{xxi}(z; \theta)$  be the  $i$ -th row of  $f_{xx}(z; \theta)$ ,  $i = 1, 2, \dots, q$ . Then,

$$f_{xxix}(z; \theta) = \frac{\partial}{\partial z} f_{xxi}(z; \theta)$$

exists and is continuous on  $\Gamma \times \Omega$ .

Assumption 5.22. There exists a constant  $K_6$  such that for every  $i = 1, 2, \dots, q$ , all  $z$  in  $\Gamma$ , and all  $\theta$  in  $\Omega$ , the absolute value of each element of  $f_{xxix}(z; \theta)$  is bounded by  $K_6$ .

In the next theorem, we derive the asymptotic bias of the instrumental variable estimator  $\hat{\beta}$ .

Theorem 5.5. Let Assumptions 5.1 through 5.15, 5.21, and 5.22 hold.

Assume either

(a)  $(e_{nt}, u_{nt})$  and  $\varepsilon_{nt}$  are independent, and

$$E\{|a_n^{1/2} u_{nt}|^3\} = O(1) \quad , \quad (5.67)$$

or

(b)  $(e_{nt}, u_{nt})$  and  $\varepsilon_{nt}$  are uncorrelated, and

$$E\{|a_n^{1/2} (e_{nt}, u_{nt})|^6\} = O(1) \quad . \quad (5.68)$$

Then, by taking the expectation of the expansion of  $(\hat{\beta}_n - \beta^0)$  through the terms of  $O_p(a_n^{-1})$ , the bias of  $\hat{\beta}_n$  is, for case (a),

$$- \frac{1}{2} \{h^0 \bar{m}_{WW}^{-1} h^0\}^{-1} h^0 \bar{m}_{WW}^{-1} \{b_n^{-1} \sum_{t=1}^{b_n} w_t^0, \text{tr}[f_{xxt}^0 \Sigma_{uun}]\} \quad ,$$

and for case (b),

$$- \frac{1}{2} \{h^0 \bar{m}_{WW}^{-1} h^0\}^{-1} h^0 \bar{m}_{WW}^{-1} \{b_n^{-1} \sum_{t=1}^{b_n} w_t^0, \text{tr}[f_{xxt}^0 \Sigma_{uun}] + b_n^{-1} \sum_{t=1}^{b_n} R_{nt}\} \quad ,$$

where  $R_{nt}$  is a  $p \times 1$  vector with the  $i$ -th element being

$$\text{tr}[f_{\text{xx}t}^0 E\{r_{\text{nt}i} u'_{\text{nt}} u_{\text{nt}}\}] \quad .$$

Proof. We observe that (5.68) implies (5.67). By Assumption 5.21,

$$\begin{aligned} Y_{\text{nt}} - f(x_{\text{nt}}; \beta^0) &= e_{\text{nt}} - f_x(x_t^0; \beta^0) u'_{\text{nt}} - \frac{1}{2} u_{\text{nt}} f_{\text{xx}}(x_t^0; \beta^0) u'_{\text{nt}} \\ &\quad - \frac{1}{6} \sum_{i=1}^q u_{\text{nt}} f_{\text{xxix}}(x_{\text{nt}}^*; \beta^0) u'_{\text{nt}} u_{\text{nt}i} \quad , \end{aligned} \quad (5.69)$$

where  $x_{\text{nt}}^*$  is on the segment joining  $x_t^0$  and  $x_{\text{nt}}$ . We note that

$$b_n^{-1} \sum_{t=1}^{b_n} [e_{\text{nt}} - f_x(x_t^0; \beta^0) u'_{\text{nt}}] w_t^0 = o_p(n^{-1/2}) \quad , \quad (5.70)$$

and that

$$b_n^{-1} \sum_{t=1}^{b_n} u_{\text{nt}} f_{\text{xx}}(x_t^0; \beta^0) u'_{\text{nt}} w_t^0 = o_p(a_n^{-1}) \quad . \quad (5.71)$$

By (5.67) and Assumption 5.22, for  $i = 1, 2, \dots, q$ ,

$$b_n^{-1} \sum_{t=1}^{b_n} u_{nt} f_{xxix}(\bar{x}_{nt}^*; \beta^0) u'_{nt} u_{nti} w_t^0 = o_p(a_n^{-3/2}) . \quad (5.72)$$

For case (a), by (5.52),

$$b_n^{-1} \sum_{t=1}^{b_n} [e_{nt} - f_x(\bar{x}_t^0; \beta^0) u'_{nt}] r_{nt} = o_p(n^{-1/2}) . \quad (5.73)$$

For case (b), by (5.68), the zero correlation, and the argument used to obtain (5.53), the expression (5.73) holds. For case (a), by (5.67), the independence, and Assumption 5.22,

$$E\{b_n^{-1} \sum_{t=1}^{b_n} |u_{nt} f_{xxt}^0 u'_{nt} r_{nt}|\} = o(a_n^{-1}) , \quad (5.74)$$

and for  $i = 1, 2, \dots, q$ ,

$$E\{b_n^{-1} \sum_{t=1}^{b_n} |u_{nt} f_{xxix}(\bar{x}_{nt}^*; \beta^0) u'_{nt} u_{nti} r_{nt}|\} = o(a_n^{-3/2}) . \quad (5.75)$$

For case (b), by (5.68), Assumption 5.22, and the Cauchy-Schwartz inequality, the expressions (5.74) and (5.75) hold. Therefore, by (5.69), (5.70), (5.71), (5.72), (5.73), (5.74), (5.75), and Lemma 5.1,

$$\hat{\beta}_n - \beta^0 = \{h^0 \bar{m}_{WW} h^{0'}\}^{-1} h^0 \bar{m}_{WW}^{-1} \{\Delta_1 \hat{\beta} + \Delta_2 \hat{\beta}\} + o_p(a_n^{-1}) , \quad (5.76)$$

where

$$\Delta_1 \hat{\beta} = b_n^{-1} \sum_{t=1}^{b_n} (\tilde{w}_t^0 + \tilde{r}_{nt})' [e_{nt} - \tilde{f}_x(\tilde{x}_t^0; \beta^0) \tilde{u}_{nt}'] , \quad \text{if } a_n^{-1} = o(n^{-1/2}) ,$$

$$= 0 , \quad \text{if } n^{-1/2} = o(a_n^{-1}) ,$$

$$\Delta_2 \hat{\beta} = -\frac{1}{2} b_n^{-1} \sum_{t=1}^{b_n} (\tilde{w}_t^0 + \tilde{r}_{nt})' \tilde{u}_{nt} \tilde{f}_{xx}^0 \tilde{u}_{nt}' .$$

Since under either (a) or (b),

$$E\{\Delta_1 \hat{\beta}\} = 0 ,$$

the results follow by taking the expectation of the leading term of (5.76).  $\square$

We point out that if  $\Sigma_{rrn} = o(a_n^{-1})$  then under either (a) or (b) the bias of  $\hat{\beta}_n$  is given by

$$-\frac{1}{2} \{h^0 \tilde{m}_{ww}^{-1} h^{0'}\}^{-1} h^0 \tilde{m}_{ww}^{-1} \{b_n^{-1} \sum_{t=1}^{b_n} \tilde{w}_t^0 \text{tr}[\tilde{f}_{xx}^0 \Sigma_{uun}]\} .$$

The asymptotic bias derived in Theorem 5.5 is a function of the second derivative  $\tilde{f}_{xx}(\tilde{x}_t^0; \beta^0)$ . Thus, as in Chapter IV, the bias of  $O(a_n^{-1})$  in  $\hat{\beta}_n$  can be considered as the bias due to the curvature of the function  $f(\tilde{x}; \beta^0)$ . For the model linear in all elements of

$x_t^0$ , the bias of  $\hat{\beta}_n$  is  $o(a_n^{-1})$ . We note that in the bias of  $\hat{\beta}_n$  the covariance matrix multiplying  $f_{xxt}^0$  is  $\Sigma_{uun}$  while the multiplier is  $\Lambda_{nt}^{-1}$  in the bias of the maximum likelihood type estimator considered in Chapter IV.

### C. An Estimator of the Error Covariance Matrix

In the previous sections, we discussed the instrumental variable estimator  $\hat{\beta}_n$  of  $\beta^0$  in the model (5.8). We assumed that the covariance matrix  $\Sigma_n$  of  $(e_{nt}, u_{nt})$  is unknown. In this section, we introduce an estimator of  $\Sigma_n$ , and discuss properties of the estimator. The estimator will be used in a later section to obtain an improved estimator of  $\beta^0$ .

In the linear errors-in-variables model with instrumental variables, a general unknown error covariance matrix cannot be estimated without an additional assumption such as zero correlation between errors in  $Y_{nt}$  and  $X_{nt}$ . For the estimation of the error covariance matrix, the residual

$$\hat{v}_{nt} = Y_{nt} - X_{nt} \hat{\beta}_{IV},$$

where  $\hat{\beta}_{IV}$  is the instrumental variable estimator, plays an important role in the linear model. The sample covariances of  $\hat{v}_{nt}$  with  $(Y_{nt}, X_{nt}, \hat{v}_{nt})$  provide estimates of linear combinations of the elements in the error covariance matrix. In the nonlinear model, the residual

$$\hat{v}_{nt} = Y_{nt} - f(\tilde{x}_{nt}; \hat{\beta}_n)$$

contains the information on the measurement errors. As we observed earlier,

$$v_{nt}^0 = Y_{nt} - f(\tilde{x}_{nt}; \beta^0) \quad , \quad t = 1, 2, \dots, b_n,$$

are not identically distributed for the models nonlinear in  $\tilde{x}_t^0$ . This fact enables us to estimate the error covariance matrix in the nonlinear model without additional knowledge on  $\Sigma_n$ .

We observe that under the assumptions of Theorem 5.4,

$$\hat{\beta}_n - \beta^0 = O_p(\max[n^{-1/2}, a_n^{-1}]) \quad ,$$

$$\hat{v}_{nt} = Y_{nt} - f(\tilde{x}_{nt}; \hat{\beta}_n)$$

$$= e_{nt} - f_{xt}^0 \tilde{u}_{nt}' + O_p(\max[n^{-1/2}, a_n^{-1}]),$$

and  $f_{xt}^0$  varies over  $t = 1, 2, \dots, b_n$ , for the models nonlinear in  $\tilde{x}_t^0$ . Hence, the expected values of  $\hat{v}_{nt}^2$  are approximately given by

$$\sigma_{vnt}^2 = [1, -f_{xt}^0] \Sigma_n [1, -f_{xt}^0]'$$

$$= \{[1, -f_{xt}^0] \otimes [1, -f_{xt}^0]\} \text{vec } \Sigma_n$$

$$= \eta_t \text{vech } \Sigma_n, \quad (5.77)$$

where

$$\eta_t = (\tilde{F}_t^0 \otimes \tilde{F}_t^0) \tilde{\phi}_{q+1},$$

$$\tilde{F}_t^0 = [1, -\tilde{f}_{xt}^0],$$

and  $\tilde{\phi}_{q+1}$  was defined in Definition 2.3. The vector  $\eta_t$  can be estimated by

$$\hat{\eta}_{nt} = (\hat{\tilde{F}}_{nt} \otimes \hat{\tilde{F}}_{nt}) \hat{\phi}_{q+1}, \quad (5.78)$$

where

$$\hat{\tilde{F}}_{nt} = [1, -\tilde{f}_x(X_{nt}; \hat{\beta}_n)] .$$

The expressions (5.77) and (5.78) suggest that the estimated coefficients in the regression of  $\hat{v}_{nt}^2$  on  $\hat{\eta}_{nt}$  estimate the elements of  $\Sigma_n$ . In order to investigate the properties of such an estimator of  $\Sigma_n$ , we introduce an assumption.

Assumption 5.23. The limit

$$H_0 = \lim_{n \rightarrow \infty} b_n^{-1} \sum_{t=1}^{b_n} \{(\tilde{F}_t^0 \tilde{F}_t^0) \otimes (\tilde{F}_t^0 \tilde{F}_t^0)\},$$

exists, and the matrix

$$\underline{H} = \underline{\Phi}'_{q+1} \underline{H}_0 \underline{\Phi}_{q+1}$$

is nonsingular.

Note that for a  $1 \times (q+1)$  vector  $\underline{A}$  each element of  $(\underline{A} \otimes \underline{A}) \underline{\Phi}_{q+1}$  has the form of either  $a_i^2$  or  $2a_i a_j$ ,  $i \neq j$ , where  $a_i$  is the  $i$ -th element of  $\underline{A}$ . Thus, the multiplication of  $\underline{\Phi}_{q+1}$  combines repeated terms in  $(\underline{A} \otimes \underline{A})$ .

If the model is linear in one of the elements of  $\underline{x}_t^0$ , then the corresponding element of  $\underline{f}_{xt}^0$  is a constant. Hence, the model linear in at least one element of  $\underline{x}_t^0$  does not satisfy the nonsingularity of  $\underline{H}$  in Assumption 5.23. An intuitive interpretation of Assumption 5.23 is that the sequence  $\underline{x}_t^0$  is chosen in a way such that for large  $n$  the vectors  $(\underline{F}_t^0 \otimes \underline{F}_t^0) \underline{\Phi}_{q+1}$ ,  $t = 1, 2, \dots, b_n$ , do not concentrate on any subspaces of less than full dimension in the  $[1/2(q+1)(q+2)]$ -dimensional Euclidean space. Hence, intuitively speaking, Assumption 5.23 is satisfied if the elements of  $\underline{f}_{xt}^0$  vary sufficiently, but without increasing in absolute magnitude.

We define the estimator  $\text{vech } \hat{\underline{\Sigma}}_n$  of  $\text{vech } \underline{\Sigma}_n$  to be the estimated regression vector given by the regression of  $\hat{v}_{nt}^2$  on  $\hat{\eta}_{nt}$ ,  $t = 1, 2, \dots, b_n$ . That is,

$$\text{vech } \hat{\underline{\Sigma}}_n = \left\{ \sum_{t=1}^{b_n} \hat{\eta}_{nt}' \hat{\eta}_{nt} \right\}^{-1} \sum_{t=1}^{b_n} \hat{\eta}_{nt}' \hat{v}_{nt}^2.$$

We recall that  $\Sigma_n$  is of  $O(a_n^{-1})$ . Thus, it is desirable that the estimator  $\hat{\Sigma}_n$  satisfies

$$a_n(\hat{\Sigma}_n - \Sigma_n) \xrightarrow{P} 0. \quad (5.79)$$

The following theorem shows that our estimator  $\hat{\Sigma}_n$  satisfies the consistency condition (5.79).

Theorem 5.6. Let Assumptions 5.1 through 5.15, 5.17, and 5.23 hold.

Then,

$$\hat{\Sigma}_n - \Sigma_n = o_p(\max[a_n^{-3/2}, n^{-1/2} a_n^{-1/2}]) .$$

Proof. By Assumption 5.17 and Theorem 5.4,

$$\hat{\beta}_n - \beta^0 = o_p(\max[n^{-1/2}, a_n^{-1}]) . \quad (5.80)$$

We observe that

$$\begin{aligned} \hat{v}_{nt} &= e_{nt} - f_{xt}^0 u'_{nt} - f_{\beta t}^0 (\hat{\beta}_n - \beta^0) - \frac{1}{2} u_{nt} f_{xx}(X_{nt}^*; \beta_n^*) u'_{nt} \\ &\quad - \frac{1}{2} (\hat{\beta}_n - \beta^0)' f_{\beta\beta}(X_{nt}^*; \beta_n^*) (\hat{\beta}_n - \beta^0) - (\hat{\beta}_n - \beta^0)' f_{\beta x}(X_{nt}^*; \beta_n^*) u'_{nt} , \end{aligned} \quad (5.81)$$

where  $X_{nt}^*$  is on the segment joining  $x_t^0$  and  $X_{nt}$ , and  $\beta_n^*$  is on the segment joining  $\beta^0$  and  $\hat{\beta}_n$ . Hence, it follows from (5.80),

(5.81), and Assumptions 5.8, 5.9, 5.13, 5.15, and 5.17 that

$$\begin{aligned} b_n^{-1} \sum_{t=1}^{b_n} \hat{\eta}_{nt}' \hat{v}_{nt}^2 &= b_n^{-1} \sum_{t=1}^{b_n} \eta_t' \eta_t \text{vech } \Sigma_n + o_p(\max[a_n^{-3/2}, n^{-1/2} a_n^{-1/2}]) \\ &= o_p(a_n^{-1}) \quad , \end{aligned} \quad (5.82)$$

where we have used, for example,

$$b_n^{-1} \sum_{t=1}^{b_n} \hat{\eta}_{nt}' [(\hat{\beta}_n - \beta^0)' \tilde{f}_{\beta t}^0]^2 = o_p(\max[n^{-1}, a_n^{-2}]) \quad ,$$

$$b_n^{-1} \sum_{t=1}^{b_n} \tilde{f}_{xt}^0 u_{nt}' u_{nt} \tilde{f}_{xx}(\tilde{x}_{nt}^*; \beta_n^*) u_{nt}' = o_p(a_n^{-3/2}) \quad ,$$

$$b_n^{-1} \sum_{t=1}^{b_n} \tilde{f}_{xt}^0 u_{nt}' \tilde{f}_{\beta t}^0 (\hat{\beta}_n - \beta^0) = o_p(\max[n^{-1/2} a_n^{-1/2}, a_n^{-3/2}]) \quad ,$$

$$E\{b_n^{-1} \sum_{t=1}^{b_n} \eta_t' (e_{nt} - \tilde{f}_{xt}^0 u_{nt}')^2\} = b_n^{-1} \sum_{t=1}^{b_n} \eta_t' \sigma_{vnt}^2$$

$$= b_n^{-1} \sum_{t=1}^{b_n} \eta_t' \eta_t \text{vech } \Sigma_n \quad ,$$

$$V\{b_n^{-1} \sum_{t=1}^{b_n} \eta_t' (e_{nt} - f_{nt}^0 u_{nt})^2\} = o(n^{-1} a_n^{-1}) ,$$

and in  $\hat{\eta}_{nt}$  for some  $\tilde{x}_{nt}^{**}$  and  $\tilde{\beta}_n^{**}$  ,

$$\tilde{f}_x(\tilde{x}_{nt}; \hat{\beta}_n) = \tilde{f}_{xt}^0 + \tilde{f}_{xx}(\tilde{x}_{nt}^{**}; \tilde{\beta}_n^{**}) u_{nt}' + (\tilde{\beta}_n - \beta^0)' \tilde{f}_{\beta x}(\tilde{x}_{nt}^{**}; \tilde{\beta}_n^{**}) . \quad (5.83)$$

Also, by (5.83) and Assumptions 5.17 and 5.23,

$$\begin{aligned} b_n^{-1} \sum_{t=1}^{b_n} \hat{\eta}_{nt}' \hat{\eta}_{nt} &= b_n^{-1} \sum_{t=1}^{b_n} \eta_t' \eta_t + o_p(\max[n^{-1/2}, a_n^{-1}]) \\ &= H + o_p(1) , \end{aligned} \quad (5.84)$$

where  $H$  is nonsingular. Therefore, the result follows from (5.82) and (5.84).  $\square$

We recall that the assumption  $a_n^{-1} = o(n^{-1/2})$  was required to obtain the limiting distribution of  $n^{1/2}(\hat{\beta}_n - \beta^0)$ . The limiting distribution of  $\hat{\Sigma}_n$  with normalizing constant  $n^{1/2}$  can be derived under a weaker assumption on the rate of increase of  $a_n$ .

Theorem 5.7. Let Assumptions 5.1 through 5.15, 5.17, and 5.23 hold.

Also, let

$$\tilde{G}_n = a_n^2 V\{\text{vech}[(e_{nt}, \tilde{u}_{nt})'(e_{nt}, \tilde{u}_{nt}) - \tilde{\Sigma}_n]\} .$$

Assume

$$a_n^{-1} = o(n^{-1/3}) , \quad (5.85)$$

$$\lim_{n \rightarrow \infty} b_n^{-1} \sum_{t=1}^{b_n} (\eta_t' \eta_t) \otimes (\eta_t' \eta_t) = A_1 , \quad (5.86)$$

$$\lim_{n \rightarrow \infty} \tilde{G}_n = G .$$

Then,

$$n^{1/2} \text{vech}(\hat{\tilde{\Sigma}}_n - \tilde{\Sigma}_n) \xrightarrow{L} N(0, \tilde{V}_\sigma) ,$$

where

$$\tilde{V}_\sigma = H^{-1} A_2 H^{-1} ,$$

$$A_2 = \lim_{n \rightarrow \infty} b_n^{-1} \sum_{t=1}^{b_n} \eta_t' \eta_t G_{nt} \eta_t' \eta_t .$$

Proof. By (5.82), (5.83), and (5.85),

$$b_n^{-1} \sum_{t=1}^{b_n} \hat{\eta}_{nt}' (\hat{v}_{nt}^2 - \hat{\eta}_{nt}) \text{vech } \hat{\Sigma}_n = n^{-1} \sum_{t=1}^{b_n} \hat{\eta}_{nt}' \hat{\eta}_{nt} a_n \text{vech}[(e_{nt}, u_{nt})'(e_{nt}, u_{nt}) - \hat{\Sigma}_n] \\ + o_p(n^{-1/2}) . \quad (5.87)$$

It follows from (5.86), (5.87), and Theorem 2.12,

$$n^{1/2} b_n^{-1} \sum_{t=1}^{b_n} \hat{\eta}_{nt}' (\hat{v}_{nt}^2 - \hat{\eta}_{nt}) \text{vech } \hat{\Sigma}_n \xrightarrow{L} N(0, A_2) . \quad (5.88)$$

The result follows from (5.84) and (5.88).  $\square$

Without knowledge of a particular form of  $G_n$ , we do not have a consistent estimator of the covariance matrix  $V_G$  of the limiting distribution of  $\text{vech } \hat{\Sigma}_n$ . Thus, to make inferences on  $\hat{\Sigma}_n$ , we need more information on distributional properties of  $(e_{nt}, u_{nt})$ .

It should be pointed out that with a finite number of observations the estimator  $\hat{\Sigma}_n$  may not be a nonnegative definite matrix. To use  $\hat{\Sigma}_n$  in the derivation of estimators of  $x_t^0$  and  $\beta^0$ , we replace  $\hat{\Sigma}_n$  which is not nonnegative definite by its closest positive semidefinite matrix, i.e., the matrix with the same eigenvectors as  $\hat{\Sigma}_n$  and zeros

replacing the negative eigenvalues of  $\hat{\Sigma}_n$ . In order to simplify our theoretical development, we assume, from now on, that  $\hat{\Sigma}_n$  is nonnegative definite. It is understood that the inverse of a matrix is replaced by the Moore-Penrose generalized inverse whenever the matrix is singular in the ensuing sections.

#### D. An Estimator of the True Value

Having obtained an estimator  $\hat{\Sigma}_n$  of the error covariance matrix in the previous section, we can now obtain an estimator of the unknown true value  $\tilde{x}_t^0$  which utilizes both  $Y_{nt}$  and  $X_{nt}$ . Let  $\hat{x}_t$  be the value of  $x_t$  contained in  $\Gamma$  that minimizes

$$D_n(\tilde{x}_t) = [Y_{nt} - f(\tilde{x}_t; \hat{\beta}_n), X_{nt} - \tilde{x}_t] \hat{\Sigma}_n^{-1} [Y_{nt} - f(\tilde{x}_t; \hat{\beta}_n), X_{nt} - \tilde{x}_t]' .$$

We introduce additional assumptions.

Assumption 5.24. For all  $n$ ,  $\Sigma_n$  is nonsingular.

Assumption 5.25. For all  $n$ , the  $\tilde{x}_t^0$ ,  $t = 1, 2, \dots, b_n$ , are interior points of  $\Gamma$ .

The property of the estimator  $\hat{x}_t$  is given in the following theorem.

Theorem 5.8. Let Assumptions 5.1 through 5.15, 5.17, 5.19, 5.23, 5.24, and 5.25 hold. Then,

$$\hat{x}_t - \tilde{x}_t^0 = \hat{\xi}_{nt} + o_p(\max[n^{-1/2}, a_n^{-1}]) ,$$

where

$$\begin{aligned}\delta_{nt} &= (e_{nt}, u_{nt}) \Sigma_n^{-1} [f_{xt}^0, I] \Lambda_{nt}^{-1} \\ &= u_{nt} - v_{nt}^* \sigma_{vnt}^{-2} \Sigma_{vunt} \quad ,\end{aligned}$$

$$\Lambda_{nt} = [f_{xt}^0, I] \Sigma_n^{-1} [f_{xt}^0, I]' \quad ,$$

$$v_{nt}^* = e_{nt} - f_{xt}^0 u_{nt}' \quad ,$$

$$\Sigma_{vunt} = [1, -f_{xt}^0] \Sigma_n [0, I]' \quad ,$$

and  $\sigma_{vnt}^2$  is given in (5.77).

Proof. By Assumption 5.24 and Theorem 5.6,  $\hat{\Sigma}_n$  is nonsingular for large  $n$ . Since  $\hat{x}_t$  minimizes  $D_n(x_t)$  over  $\Gamma$ ,

$$D_n(\hat{x}_t) \leq D_n(x_t^0) \quad . \quad (5.89)$$

Observe that

$$y_{nt} - f(x_t^0; \hat{\beta}_n) = e_{nt} - f_{\beta t}^0 (\hat{\beta}_n - \beta^0) - \frac{1}{2} (\hat{\beta}_n - \beta^0)' f_{\beta\beta}(x_t^0; \beta_n^*) (\hat{\beta}_n - \beta^0) \quad , \quad (5.90)$$

where  $\beta_n^*$  is on the segment joining  $\beta^0$  and  $\hat{\beta}_n$ . Since  $\hat{\beta}_n$  is in  $\Omega$ , and since  $\Omega$  is convex,  $\beta_n^*$  is also in  $\Omega$ . Thus, by Assumptions

5.13 and 5.19, (5.80), and (5.90),

$$a_n^{1/2} [Y_{nt} - f(\tilde{x}_t^0; \hat{\beta}_n), \tilde{x}_{nt} - \tilde{x}_t^0] = o_p(1) . \quad (5.91)$$

By Theorem 5.6,

$$a_n \hat{\Sigma}_{\tilde{x}} = o_p(1) . \quad (5.92)$$

It follows from (5.89), (5.91), and (5.92) that

$$D_n(\hat{\tilde{x}}_t) \leq D_n(\tilde{x}_t^0) = o_p(1) . \quad (5.93)$$

We observe that

$$D_n(\hat{\tilde{x}}_t) = (\tilde{x}_{nt} - \hat{\tilde{x}}_t) \hat{\Sigma}_{\tilde{x}}^{-1} (\tilde{x}_{nt} - \hat{\tilde{x}}_t)' + \hat{\sigma}_{ee.un}^{-1} [Y_{nt} - f(\hat{\tilde{x}}_t; \hat{\beta}_n) - \hat{\gamma}_n(\tilde{x}_{nt} - \hat{\tilde{x}}_t)]^2 , \quad (5.94)$$

where

$$\hat{\Sigma}_{\tilde{x}} = \begin{bmatrix} \hat{\sigma}_{een} & \hat{\Sigma}_{\tilde{x}eun} \\ \hat{\Sigma}_{\tilde{x}uen} & \hat{\Sigma}_{\tilde{x}uun} \end{bmatrix} ,$$

$$\hat{\sigma}_{ee.un} = \hat{\sigma}_{een} - \hat{\Sigma}_{\tilde{x}eun} \hat{\Sigma}_{\tilde{x}uun}^{-1} \hat{\Sigma}_{\tilde{x}uen} ,$$

$$\hat{\gamma}_n = \hat{\Sigma}_{eun} \hat{\Sigma}_{uun}^{-1} .$$

By (5.93) and (5.94),

$$(\hat{x}_{nt} - \hat{x}_t) \hat{\Sigma}_{uun}^{-1} (\hat{x}_{nt} - \hat{x}_t)' = o_p(1) . \quad (5.95)$$

Since by Theorem 5.6,

$$a_n \hat{\Sigma}_{uun} = o_p(1) ,$$

it follows from Assumption 5.24 and (5.95) that

$$\hat{x}_t - \hat{x}_{nt} = o_p(a_n^{-1/2}) .$$

Since

$$\hat{x}_{nt} - x_t^0 = o_p(a_n^{-1/2}) ,$$

we have

$$\hat{x}_t - x_t^0 = o_p(a_n^{-1/2}) . \quad (5.96)$$

Let  $f_{xi}$  be the  $i$ -th element of  $\tilde{f}_x$ , let  $\tilde{f}_{xxi}$  be the  $i$ -th row of  $\tilde{f}_{xx}$ , and let  $I_i$  be the  $i$ -th row of the  $q \times q$  identity matrix. Then, for  $i = 1, 2, \dots, q$ ,

$$\begin{aligned}
 & [Y_{nt} - f(\hat{x}_t; \hat{\beta}_n), X_{nt} - \tilde{x}_t] \hat{\Sigma}_n^{-1} [f_{xi}(\hat{x}_t; \hat{\beta}_n), I_i]' \\
 &= [Y_{nt} - f(\tilde{x}_t^0; \hat{\beta}_n), u_{nt}] \hat{\Sigma}_n^{-1} [f_{xi}(\tilde{x}_t^0; \hat{\beta}_n), I_i]' \\
 &+ (\hat{x}_t - \tilde{x}_t^0) \{ [\tilde{f}'_{xxi}(\tilde{x}_t^*; \hat{\beta}_n), 0] \hat{\Sigma}_n^{-1} [Y_{nt} - f(\tilde{x}_t^*; \hat{\beta}_n), X_{nt} - \tilde{x}_t^*]' \\
 &- [\tilde{f}'_x(\tilde{x}_t^*; \hat{\beta}_n), I] \hat{\Sigma}_n^{-1} [f_{xi}(\tilde{x}_t^*; \hat{\beta}_n), I_i] \} , \tag{5.97}
 \end{aligned}$$

where  $\tilde{x}_t^*$  is on the segment joining  $\tilde{x}_t^0$  and  $\hat{x}_t$ . The left hand side of (5.97) is a multiple of the derivative of  $D_n(\tilde{x}_t)$  evaluated at  $\hat{x}_t$ . Therefore, by Assumption 5.25 and the argument used in (5.34) for  $i = 1, 2, \dots, q$ , the left hand

$$[Y_{nt} - f(\hat{x}_t; \hat{\beta}_n), X_{nt} - \hat{x}_t] \hat{\Sigma}_n^{-1} [f_{xi}(\hat{x}_t; \hat{\beta}_n), I_i]' = o_p(a_n^{-2}) . \tag{5.98}$$

It follows from (5.80), (5.90), (5.96), (5.98), and Theorem 5.6 that

$$\begin{aligned}
 \hat{x}_t - \tilde{x}_t^0 &= (e_{nt}, u_{nt}) \hat{\Sigma}_n^{-1} [\tilde{f}_{xt}^0, I]' \Lambda_{nt}^{-1} + o_p(\max[n^{-1/2}, a_n^{-1}]) \\
 &= u_{nt} + [v_{nt}^*, 0] \hat{\Sigma}_n^{-1} [\tilde{f}_{xt}^0, I]' \Lambda_{nt}^{-1} + o_p(\max[n^{-1/2}, a_n^{-1}]) ,
 \end{aligned}$$

where

$$\Lambda_{nt} = [f_{xt}^0, I] \Sigma_n^{-1} [f_{xt}^0, I]' .$$

Using Theorem 2.1, we obtain the result.  $\square$

If we ignore the terms of  $O_p(\max[n^{-1/2}, a_n^{-1}])$ , then the error in  $\hat{x}_t$  as an estimator of  $x_t^0$  is  $\delta_{nt}$  with mean zero and covariance matrix  $\Lambda_{nt}^{-1}$ . Since by Theorem 5.6

$$a_{n\sim n} \hat{\Sigma} - a_{n\sim n} \Sigma = O_p(\max[a_n^{-1/2}, b_n^{-1/2}]) ,$$

the errors of  $O_p(a_n^{-1})$  in  $(\hat{x}_t - x_t^0)$  include the terms due to the error in  $\hat{\Sigma}_n$  along with the terms due to the curvature of  $f$ .

The unknown true value  $y_t^0$  can be estimated by

$$\hat{y}_t = f(\hat{x}_t; \hat{\beta}_n) .$$

Corollary 5.8.1. Let Assumptions 5.1 through 5.15, 5.17, 5.19, 5.23, 5.24, and 5.25 hold. Then,

$$\hat{y}_t - y_t^0 = \delta_{nt}^* + O_p(\max[n^{-1/2}, a_n^{-1}]) ,$$

where

$$\delta_{nt}^* = e_{nt} - v_{nt}^* \sigma_{vnt}^{-2} \sigma_{vent} ,$$

$$\sigma_{\text{vent}} = [1, -f_{\text{xt}}^{0'}] \Sigma_n (1, 0)'$$

Proof. We observe that by Theorem 5.8,

$$\begin{aligned} \hat{y}_t - y_t^0 &= f_{\text{xt} \sim \text{nt}}^0 \delta' + O_p(\max[n^{-1/2}, a_n^{-1}]) \\ &= e_{\text{nt}} + \{f_{\text{xt} \sim \text{nt}}^0 \Lambda^{-1} [f_{\text{xt}}^{0'}, I] \Sigma_n^{-1} (I, 0) - I\} v_{\text{nt}}^* + O_p(\max[n^{-1/2}, a_n^{-1}]) . \end{aligned}$$

Thus, the result follows from Theorem 2.1.  $\square$

In practice, the minimization of  $D_n(\hat{x}_t)$  for each  $t$  may not be feasible. For computational purposes, we present the following local approximation to  $\hat{x}_t$ . Let

$$\bar{x}_t = \hat{x}_{\text{nt}} - \hat{u}_{\text{nt}} ,$$

where

$$\hat{u}_{\text{nt}} = \hat{v}_{\text{nt}} \hat{\sigma}_{\text{vnt}}^{-2} \hat{\Sigma}_{\text{vunt}} ,$$

$$\hat{v}_{\text{nt}} = Y_{\text{nt}} - f(\hat{x}_{\text{nt}}; \hat{\beta}_n) ,$$

$$\hat{\sigma}_{\text{vnt}}^2 = [1, -f_{\text{x}}(\hat{x}_{\text{nt}}; \hat{\beta}_n)] \hat{\Sigma}_n [I, -f_{\text{x}}(\hat{x}_{\text{nt}}; \hat{\beta}_n)]' ,$$

$$\hat{\Sigma}_{\text{vunt}} = [1, -f_{\text{x}}(\hat{x}_{\text{nt}}; \hat{\beta}_n)] \hat{\Sigma}_n [0, I]' .$$

The next theorem shows that the estimator  $\bar{x}_t$  is equivalent to  $\hat{x}_t$  up to terms of  $O_p(a_n^{-1/2})$ .

Theorem 5.9. Let Assumption 5.1 through 5.15, 5.17, 5.19, 5.23, 5.24, and 5.25 hold. Then,

$$\bar{x}_t - x_t^0 = \delta_{nt} + O_p(\max[n^{-1/2}, a_n^{-1}]) ,$$

where  $\delta_{nt}$  is defined in Theorem 5.8.

Proof. By (5.80) and (5.81),

$$\hat{v}_{nt} = v_{nt}^* + O_p(\max[n^{-1/2}, a_n^{-1}]) ,$$

$$\hat{f}_x(\hat{x}_{nt}; \hat{\beta}_n) = f_{xt}^0 + O_p(a_n^{-1/2}) .$$

Hence, the result follows from Theorem 5.6. □

Corollary 5.9.1. Let Assumptions 5.1 through 5.15, 5.17, 5.19, 5.23, 5.24, and 5.25 hold. Also, let

$$\bar{y}_t = f(\bar{x}_t; \hat{\beta}_n) .$$

Then,

$$\bar{y}_t - y_t^0 = \delta_{nt}^* + O_p(\max[n^{-1/2}, a_n^{-1}]) ,$$

where  $\delta_{nt}^*$  is defined in Corollary 5.8.1.

Proof. The result follows from Theorem 5.9 and the argument used in the proof of Corollary 5.8.1.  $\square$

## E. The Modified Instrumental Variable Estimator

### 1. Introduction

In the previous sections, we discussed the properties of the instrumental variable estimator  $\hat{\beta}_n$  and the estimators of the error covariance matrix and of the true values based on  $\hat{\beta}_n$ . The estimator  $\hat{\beta}_n$  was introduced as an analogue of the two-stage least squares estimator in the simultaneous equation system. As we discussed earlier, the quantities  $[Y_{nt} - f(X_{nt}; \beta^0)]$ ,  $t = 1, 2, \dots, b_n$ , are not identically distributed. The estimator  $\hat{\beta}_n$  was obtained by minimizing  $Q_n(\beta)$  which does not take into account the differences among the variances of  $[Y_{nt} - f(X_{nt}; \beta^0)]$ . Having obtained an estimator of  $\Sigma_n$ , we now seek to improve the estimator of  $\beta^0$  by modifying  $Q_n(\beta)$  with estimated variances of  $[Y_{nt} - f(X_{nt}; \beta^0)]$ . That is, we expect to obtain an improvement analogous to that of the estimated generalized least squares estimator over the ordinary least squares estimator.

We have discovered another unpleasant property of the estimator  $\hat{\beta}_n$ . The derivation of the limiting distribution of  $n^{1/2}(\hat{\beta}_n - \beta^0)$  required a rather strong assumption for the rate of increase in  $a_n$ . The error variances were required to decrease at a faster rate than  $n^{-1/2}$ . This assumption about  $a_n$  may be weakened by considering the second derivative adjustment used in Chapter IV. We also expect that such an adjustment will reduce the bias in  $\hat{\beta}_n$  due to the

nonlinearity of the relationship.

In the derivation of the properties of  $\hat{\beta}_n$ , we considered two cases depending on whether  $(e_{nt}, u_{nt})$  and  $\tilde{r}_{nt}$  are independent or uncorrelated. Such separate considerations made our discussion rather tedious. We had to introduce different assumptions and proofs for different cases. But, every result we derived was equally valid for the two cases with appropriate adjustment in the assumptions. As shown in Theorem 5.3, the inferences on  $\beta^0$  based on the asymptotic results for  $\hat{\beta}_n$  are identical for the independent case and the uncorrelated case. Thus, in order to simplify our discussion on the modified estimator of  $\beta^0$ , we consider only the case where  $\tilde{w}_{nt}$  are fixed constants.

The assumption of fixed  $\tilde{w}_{nt}$  does not necessarily mean that we observe fixed constants  $\tilde{w}_{nt}$  without measurement errors. The  $\tilde{w}_{nt}$  may have random components which are independent of  $(e_{nt}, u_{nt})$ . In such a case, we can make inferences on the equation of interest, conditional on the observed  $\tilde{w}_{nt}$ , without altering the nature of the equation. Also, as we have seen, the estimated covariance matrix of the estimator for unconditional inference has the same form as that for conditional inference. Thus, our asymptotic results under the assumption of fixed  $\tilde{w}_{nt}$  can be applied to the case with the assumption that  $\tilde{w}_{nt}$  are independent of  $(e_{nt}, u_{nt})$ . The results in this section would hold for the cases with only the zero correlation of  $\tilde{w}_{nt}$  and  $(e_{nt}, u_{nt})$  and with higher moment assumptions. Since our purpose in this section is the presentation of a modified estimator, we concentrate on the case with fixed  $\tilde{w}_{nt}$ , and avoid the complication due to the differences in

the assumptions on  $\tilde{W}_{nt}$  .

We also transform  $\tilde{W}_{nt}$  so that for all  $n$

$$b_n^{-1} \sum_{t=1}^{b_n} \tilde{W}_{nt}' \tilde{W}_{nt} = I \quad .$$

Since  $\tilde{W}_{nt}$  are fixed and, thus, independent of  $(Y_{nt}, X_{nt})$  , such a transformation does not change our model specification.

The assumption of fixed  $\tilde{W}_{nt}$  may be considered as a special case of the general model discussed in previous sections, where  $\tilde{r}_{nt} = 0$  for all  $t$  and  $n$  . But, we choose to introduce some assumptions to clarify the model considered in this section.

Assumption 5.1a. The observations satisfy

$$Y_{nt} = f(\tilde{x}_t^0; \beta^0) + e_{nt} \quad ,$$

$$X_{nt} = \tilde{x}_t^0 + \tilde{u}_{nt} \quad , \quad t = 1, 2, \dots, b_n \quad ,$$

$$(e_{nt}, \tilde{u}_{nt})' \sim II(0, \Sigma_n) \quad ,$$

$$\Sigma_n = o(a_n^{-1}) \quad .$$

Assumption 5.4a. The  $\tilde{W}_{nt}$  form a triangular array of fixed constants satisfying, for all  $n$  ,

$$b_n^{-1} \sum_{t=1}^{b_n} \tilde{W}_{nt}' \tilde{W}_{nt} = \tilde{I} \quad .$$

Assumption 5.10a. Uniformly for all  $\theta$  in  $\Omega$  ,

$$\lim_{n \rightarrow \infty} b_n^{-1} \sum_{t=1}^{b_n} \tilde{f}_{\beta}^0(\tilde{x}_t^0; \theta) \tilde{W}_{nt} = h(\theta) \quad ,$$

where the rank of  $h(\theta)$  is  $k$  for all  $\theta$  in  $\Omega$  .

Assumption 5.16a. Let  $W_{nti}$  and  $f_{xti}^0$  be the  $i$ -th elements of  $\tilde{W}_{nt}$  and  $\tilde{f}_{xt}^0$  , respectively. Then, for all  $i_1, i_2 = 1, 2, \dots, p$ , and for all  $j_1, j_2 = 1, 2, \dots, q$ ,

$$\lim_{n \rightarrow \infty} b_n^{-1} \sum_{t=1}^{b_n} W_{nti_1} W_{nti_2} f_{xtj_1}^0 f_{xtj_2}^0 = \bar{m}_4(i_1, i_2, j_1, j_2) \quad .$$

Also,

$$\lim_{n \rightarrow \infty} a_n \sum_{n \sim n} = \Phi \quad .$$

Assumption 5.20a. For  $i = 1, 2, \dots, p$ ,

$$b_n^{-1} \sum_{t=1}^{b_n} (W_{nti})^4 = o(1) .$$

Under Assumption 5.4a, the instrumental variable estimator  $\hat{\beta}_n$  minimizes

$$Q_n(\beta) = \left\{ \sum_{t=1}^{b_n} [Y_{nt} - f(X_{nt}; \beta)] W_{nt} \right\} \left\{ \sum_{t=1}^{b_n} W'_{nt} [Y_{nt} - f(X_{nt}; \beta)] \right\} .$$

We summarize the properties of  $\hat{\beta}_n$  for the case with fixed  $W_{nt}$ . By Theorem 5.1, under Assumptions 5.1a, 5.2, 5.4a, 5.5 through 5.9, and 5.10a,  $\hat{\beta}_n$  is consistent for  $\beta^0$ . By Theorem 5.2, if Assumptions 5.1a, 5.2, 5.4a, 5.5 through 5.9, 5.10a, 5.11 through 5.15, and 5.16a hold, and if  $a_n^{-1} = o(n^{-1/2})$ , then

$$n^{1/2} (\hat{\beta}_n - \beta^0) \xrightarrow{L} N(0, V) , \quad (5.99)$$

where

$$V = \{h^0 h^{0'}\}^{-1} h^0 \bar{m} h^{0'} \{h^0 h^{0'}\}^{-1} ,$$

$$\bar{m} = \lim_{n \rightarrow \infty} b_n^{-1} \sum_{t=1}^{b_n} a_n \sigma_{vnt}^2 \tilde{w}'_{nt} \tilde{w}_{nt} ,$$

$$\tilde{h}^0 = \tilde{h}(\beta^0) ,$$

$$\sigma_{vnt}^2 = [1, -\tilde{f}_{xt}^0] \Sigma_n [1, -\tilde{f}_{xt}^0]' .$$

By Theorem 5.4, under Assumptions 5.1a, 5.2, 5.4a, 5.5 through 5.9, 5.10a, and 5.11 through 5.15,

$$\hat{\tilde{\beta}}_n - \beta^0 = O_p(\max[n^{-1/2}, a_n^{-1}]) . \quad (5.100)$$

Also, by Theorem 5.5, if Assumptions 5.1a, 5.2, 5.4a, 5.5 through 5.9, 5.10a, 5.11 through 5.15, 5.21, and 5.22 hold, and if

$$E\{|a_n^{1/2} \tilde{u}_{nt}|^3\} = O(1) ,$$

then the bias of  $\hat{\tilde{\beta}}_n$  through terms of  $O_p(a_n^{-1})$  is

$$-\frac{1}{2} \{\tilde{h}^0 \tilde{h}^{0'}\}^{-1} \tilde{h}^0 \{b_n^{-1} \sum_{t=1}^{b_n} \tilde{w}'_{nt} \text{tr}[\tilde{f}_{xxt}^0 \Sigma_{uun}]\} . \quad (5.101)$$

By Theorem 5.7, under Assumptions 5.1a, 5.2, 5.4a, 5.5 through 5.9, 5.10a, 5.11 through 5.15, 5.17, and 5.23,

$$\hat{\Sigma}_n - \Sigma_n = O_p(\max[a_n^{-3/2}, n^{-1/2} a_n^{-1/2}]) \quad (5.102)$$

Finally, by Theorem 5.8, under Assumptions 5.1a, 5.2, 5.4a, 5.5 through 5.9, 5.10a, 5.11 through 5.15, 5.17, 5.19, 5.23, 5.24, and 5.25,

$$\hat{x}_t - x_t^0 = \delta_{nt} + O_p(\max[n^{-1/2}, a_n^{-1}]) \quad (5.103)$$

## 2. The estimator

The covariance matrix of the limiting distribution of  $\hat{\beta}_n$  given in (5.99) has the form of the covariance matrix of the ordinary least squares estimator with heteroscedastic residuals. The variance of  $[Y_{nt} - f(x_{nt}; \beta^0)]$  is approximately  $\sigma_{vnt}^2$ , which can be estimated using  $\hat{\Sigma}_n$  and  $\hat{x}_t$ . Also, we notice from (5.101) that the source of the bias of  $\hat{\beta}_n$  is the same as that for the maximum likelihood type estimator discussed in Chapter IV with  $\Sigma_{uun}$  replacing  $\Lambda_{nt}^{-1}$ . Thus, the bias of  $O(a_n^{-1})$  may be removed by an adjustment similar to that considered in Chapter IV. Even though there are more than one possible type of adjustment, we consider only one of such adjustments in this chapter. We choose the adjustment based on the local quadratic approximation because of its relative simplicity.

Let

$$Y_{nt}^* = Y_{nt} + d\hat{Y}_{nt} ,$$

where

$$d\hat{Y}_{nt} = \frac{1}{2} \text{tr}[\hat{f}_{xx}(\hat{x}_t; \hat{\beta}_n) \hat{\Sigma}_{uun}] ,$$

$$\hat{\Sigma}_{\sim n} = \begin{bmatrix} \hat{\sigma}_{een} & \hat{\Sigma}_{eun} \\ \hat{\Sigma}_{uen} & \hat{\Sigma}_{uun} \end{bmatrix} .$$

The modified instrumental variable estimator  $\tilde{\beta}_n$  is defined to be the value of  $\beta$  contained in  $\Omega$  that minimizes

$$Q_n^*(\beta) = \{b_n^{-1} \sum_{t=1}^{b_n} [Y_{nt}^* - f(\tilde{x}_{nt}; \beta)] \tilde{w}_{nt}\} \{b_n^{-1} \sum_{t=1}^{b_n} \hat{\sigma}_{vnt}^2 \tilde{w}_{nt}' \tilde{w}_{nt}\}^{-1}$$

$$\{b_n^{-1} \sum_{t=1}^{b_n} \tilde{w}_{nt}' [Y_{nt}^* - f(\tilde{x}_{nt}; \beta)]\} , \quad (5.104)$$

where

$$\hat{\sigma}_{vnt}^2 = [1, -\hat{f}_x(\hat{x}_t; \hat{\beta}_n)] \hat{\Sigma}_n [1, -\hat{f}_x(\hat{x}_t; \hat{\beta}_n)]' .$$

The weight matrix

$$b_n^{-1} \sum_{t=1}^{b_n} \hat{\sigma}_{vnt}^2 \tilde{w}_{nt}' \tilde{w}_{nt}$$

estimates the variance of

$$b_n^{-1} \sum_{t=1}^{b_n} [Y_{nt} - f(\tilde{x}_{nt}; \beta^0)] \tilde{w}_{nt} \quad .$$

### 3. Preliminary lemmas

We present three lemmas which will be used to study the asymptotic properties of  $\tilde{\beta}_n$  .

Lemma 5.2. Let Assumptions 5.1a, 5.2, 5.4a, 5.5 through 5.9, 5.10a, 5.11 through 5.15, 5.17, 5.19, 5.23, 5.24 and 5.25 hold. Then,

$$b_n^{-1} \sum_{t=1}^{b_n} (\hat{x}_t - \tilde{x}_t^0)' (\hat{x}_t - \tilde{x}_t^0) = o_p(a_n^{-1}) \quad .$$

Proof. By (5.93) and (5.94),

$$\text{tr}\{\hat{\Sigma}_{uun}^{-1} b_n^{-1} \sum_{t=1}^{b_n} (\tilde{x}_{nt} - \hat{x}_t)' (\tilde{x}_{nt} - \hat{x}_t)\} < b_n^{-1} \sum_{t=1}^{b_n} D_n(\tilde{x}_t^0) \quad . \quad (5.105)$$

By (5.90), (5.100), (5.102), and Assumption 5.13,

$$\begin{aligned}
 b_n^{-1} \sum_{t=1}^{b_n} D_n(\tilde{x}_t^0) &= \text{tr} \left\{ \hat{\Sigma}_n^{-1} b_n^{-1} \sum_{t=1}^{b_n} [Y_{nt} - f(\tilde{x}_t^0; \hat{\beta}_n), \tilde{u}_{nt}]' [Y_{nt} - f(\tilde{x}_t^0; \hat{\beta}_n), \tilde{u}_{nt}] \right\} \\
 &= \text{tr} \left\{ [a_{n\tilde{n}} \hat{\Sigma}_n]^{-1} b_n^{-1} \sum_{t=1}^{b_n} a_n (e_{nt}, \tilde{u}_{nt})' (e_{nt}, \tilde{u}_{nt}) \right\} \\
 &\quad + O_p(\max[n^{-1/2} a_n^{1/2}, a_n^{-1/2}]) \\
 &= O_p(1) .
 \end{aligned} \tag{5.106}$$

Since by (5.102) each element of  $\hat{\Sigma}_{uun}$  is  $O_p(a_n^{-1})$ , there is a matrix  $\hat{T}_n$  such that

$$\begin{aligned}
 (a_{n\tilde{n}} \hat{\Sigma}_n)^{-1} &= \hat{T}_n \hat{T}_n' , \\
 \hat{T}_n &= O_p(1) .
 \end{aligned} \tag{5.107}$$

Hence, by (5.105) and (5.106),

$$\text{tr} \left\{ b_n^{-1} \sum_{t=1}^{b_n} \hat{T}_n' (\tilde{x}_{nt} - \hat{x}_t) (\tilde{x}_{nt} - \hat{x}_t)' \hat{T}_n \right\} = O_p(a_n^{-1}) .$$

Thus, the average of  $b_n$  values of squares of the elements of  $(\underline{x}_{nt} - \hat{x}_t) \hat{T}_n$  is  $O_p(a_n^{-1})$ . Therefore, by the Cauchy-Schwarz inequality,

$$b_n^{-1} \sum_{t=1}^{b_n} \hat{T}_n' (\underline{x}_{nt} - \hat{x}_t)' (\underline{x}_{nt} - \hat{x}_t) \hat{T}_n = O_p(a_n^{-1}) . \quad (5.108)$$

It follows from (5.107) and (5.108) that

$$b_n^{-1} \sum_{t=1}^{b_n} (\underline{x}_{nt} - \hat{x}_t)' (\underline{x}_{nt} - \hat{x}_t) = O_p(a_n^{-1}) . \quad (5.109)$$

We observe that

$$\begin{aligned} b_n^{-1} \sum_{t=1}^{b_n} (\hat{x}_t - \underline{x}_t^0)' (\hat{x}_t - \underline{x}_t^0) &= b_n^{-1} \sum_{t=1}^{b_n} (\underline{x}_{nt} - \hat{x}_t)' (\underline{x}_{nt} - \hat{x}_t) - b_n^{-1} \sum_{t=1}^{b_n} (\underline{x}_{nt} - \hat{x}_t)' \underline{u}_{nt} \\ &\quad - b_n^{-1} \sum_{t=1}^{b_n} \underline{u}_{nt}' (\underline{x}_{nt} - \hat{x}_t) + b_n^{-1} \sum_{t=1}^{b_n} \underline{u}_{nt}' \underline{u}_{nt} . \end{aligned} \quad (5.110)$$

By Assumption 5.1a,

$$b_n^{-1} \sum_{t=1}^{b_n} \tilde{u}_{nt}' \tilde{u}_{nt} = o_p(a_n^{-1}) \quad (5.111)$$

By (5.109), (5.111) and the Cauchy-Schwarz inequality,

$$b_n^{-1} \sum_{t=1}^{b_n} (\tilde{x}_{nt} - \hat{\tilde{x}}_t)' \tilde{u}_{nt} = o_p(a_n^{-1}) \quad (5.112)$$

The result follows from (5.109) through (5.112).  $\square$

Lemma 5.3. Let Assumptions 5.1a, 5.2, 5.4a, 5.5 through 5.9, 5.10a, 5.11 through 5.15, 5.17, 5.19, 5.23, 5.24, and 5.25 hold. Then,

$$b_n^{-1} \sum_{t=1}^{b_n} [f_{\tilde{x}}(\hat{\tilde{x}}_t; \hat{\beta}_n) - f_{\tilde{x}\tilde{x}}^0]' [f_{\tilde{x}}(\hat{\tilde{x}}_t; \hat{\beta}_n) - f_{\tilde{x}\tilde{x}}^0] = o_p(a_n^{-1}) \quad .$$

Proof. We observe that

$$f_{\tilde{x}}(\hat{\tilde{x}}_t; \hat{\beta}_n) - f_{\tilde{x}\tilde{x}}^0 = f_{\tilde{x}\tilde{x}}(\tilde{x}_t^*; \beta_n^*)(\hat{\tilde{x}}_t - \tilde{x}_t^0)' + f_{\tilde{x}\beta}(\tilde{x}_t^*; \beta_n^*)(\hat{\beta}_n - \beta^0) \quad , \quad (5.113)$$

where  $\tilde{x}_t^*$  is on the segment joining  $\tilde{x}_t^0$  and  $\hat{\tilde{x}}_t$ , and  $\beta_n^*$  is on the

segment joining  $\beta^0$  and  $\hat{\beta}_n$ . Hence, the result follows from (5.100), (5.113), Assumptions 5.13 and 5.15, the Cauchy-Schwarz inequality, and Lemma 5.2.  $\square$

Lemma 5.4. Let Assumptions 5.1a, 5.2, 5.4a, 5.5 through 5.9, 5.10a, 5.11 through 5.17, 5.19, 5.23, 5.24, and 5.25 hold. Then,

$$b_n^{-1} \sum_{t=1}^{b_n} [a_n(\hat{\sigma}_{vnt}^2 - \sigma_{vnt}^2)]^2 = O_p(\max[b_n^{-1}, a_n^{-1}]) .$$

Proof. We observe that

$$\begin{aligned} \hat{\sigma}_{vnt}^2 - \sigma_{vnt}^2 &= [1, -\tilde{f}_x(\tilde{x}_t; \hat{\beta}_n)](\hat{\Sigma}_n - \Sigma_n)[1, -\tilde{f}_x(\tilde{x}_t; \hat{\beta}_n)]' \\ &\quad - [1, -\tilde{f}_x(\tilde{x}_t; \hat{\beta}_n)]\Sigma_n[0, \tilde{f}_x(\tilde{x}_t; \hat{\beta}_n) - \tilde{f}_{xt}^0]' \\ &\quad - [0, \tilde{f}_x(\tilde{x}_t; \hat{\beta}_n) - \tilde{f}_{xt}^0]\Sigma_n[1, -\tilde{f}_{xt}^0]' . \end{aligned}$$

Hence, by Assumptions 5.1a and 5.8, for some constants  $K_1^*$  and  $K_2^*$ ,

$$|a_n(\hat{\sigma}_{vnt}^2 - \sigma_{vnt}^2)| \leq K_1^* |\text{vech}[a_n(\hat{\Sigma}_n - \Sigma_n)]| + K_2^* |\tilde{f}_x(\tilde{x}_t; \hat{\beta}_n) - \tilde{f}_{xt}^0| .$$

Therefore, the result follows from (5.102), Lemma 5.3, and the Cauchy-Schwarz inequality.  $\square$

#### 4. Consistency

To show the consistency of the modified instrumental variable estimator  $\tilde{\beta}_n$ , we introduce an additional assumption.

Assumption 5.26. The limit

$$\lim_{n \rightarrow \infty} \tilde{m}_n = \bar{m}$$

is positive definite, where

$$\tilde{m}_n = b_n^{-1} \sum_{t=1}^{b_n} a_n \sigma_{vnt}^2 \frac{W' W}{\tilde{m}_{nt} \tilde{m}_{nt}} .$$

Note that the existence of the limit in Assumption 5.26 follows from Assumptions 5.1a and 5.16a. But, the nonsingularity of the limit  $\bar{m}$  will be needed for the consistency of  $\tilde{\beta}_n$  defined in (5.104). In the following theorem, we show the consistency of  $\tilde{\beta}_n$ . Note that Assumptions 5.20a and 5.25 are used, but Assumption 5.16a is not used to obtain the consistency.

Theorem 5.10. Let Assumptions 5.1a, 5.2, 5.4a, 5.5 through 5.9, 5.10a, 5.11 through 5.15, 5.17, 5.19, 5.20a, and 5.23 through 5.26 hold. Then,

$$\tilde{\beta}_n \xrightarrow{P} \beta^0 .$$

Proof. The estimator  $\tilde{\beta}_n$  minimizes over  $\Omega$

$$Q_n^{**}(\beta) = a_n^{-1} Q_n^*(\beta)$$

$$= \{b_n^{-1} \sum_{t=1}^{b_n} [Y_{nt} - f(\underline{x}_{nt}; \beta) + d\hat{Y}_{nt} W_{nt}] \hat{m}_n^{-1}$$

$$\{b_n^{-1} \sum_{t=1}^{b_n} [Y_{nt} - f(\underline{x}_{nt}; \beta) + d\hat{Y}_{nt} W_{nt}] \}' , \quad (5.114)$$

where

$$\hat{m}_n = b_n^{-1} \sum_{t=1}^{b_n} a_n \hat{\sigma}_{vnt}^2 W_{nt}' W_{nt} .$$

Let  $W_{nti}$  be the  $i$ -th element of  $W_{nt}$ , and let  $\hat{m}_{nij}$  and  $m_{nij}$  be the  $(i,j)$ -th elements of  $\hat{m}_n$  and  $m_n$ , respectively. Then, by Assumption 5.20a, Lemma 5.4, and the Cauchy-Schwarz inequality, for  $i, j = 1, 2, \dots, p$ ,

$$|\hat{m}_{nij} - m_{nij}| = |b_n^{-1} \sum_{t=1}^{b_n} a_n (\hat{\sigma}_{vnt}^2 - \sigma_{vnt}^2) W_{nti} W_{ntj}|$$

$$\leq \{b_n^{-1} \sum_{t=1}^{b_n} [a_n (\hat{\sigma}_{vnt}^2 - \sigma_{vnt}^2)]^2\}^{1/2} \{b_n^{-1} \sum_{t=1}^{b_n} (W_{nti})^4\}^{1/4} \{b_n^{-1} \sum_{t=1}^{b_n} (W_{ntj})^4\}^{1/4}$$

$$= o_p(\max[b_n^{-1/2}, a_n^{-1/2}]) .$$

Thus,

$$\hat{\tilde{m}}_n - \tilde{m}_n = o_p(\max[b_n^{-1/2}, a_n^{-1/2}]) . \quad (5.115)$$

By Assumption 5.26,

$$\lim_{n \rightarrow \infty} \tilde{m}_n = \bar{m} . \quad (5.116)$$

It follows from (5.115) and (5.116) that

$$\hat{\tilde{m}}_n = \bar{m} + o_p(1) . \quad (5.117)$$

By (5.102) and Assumptions 5.4a and 5.15,

$$\begin{aligned} b_n^{-1} \sum_{t=1}^{b_n} |(\hat{dY}_{nt}) \tilde{w}_{nt}| &\leq \frac{1}{2} K_4 \text{tr}[\hat{\Sigma}_{uun}] \{ b_n^{-1} \sum_{t=1}^{b_n} |\tilde{w}_{nt}| \} \\ &= o_p(a_n^{-1}) . \end{aligned} \quad (5.118)$$

By (5.114), (5.117), (5.118), and the argument used in the proof of Theorem 5.1, for every  $\epsilon > 0$  there is an  $N(\epsilon)$  such that if  $n > N(\epsilon)$  then for all  $\beta$  in  $\Omega$

$$P\{|Q_n^{**}(\beta) - (\beta - \beta^0)' h(\beta^*) \bar{m}^{-1} [h(\beta^*)]' (\beta - \beta^0)| > \varepsilon\} < \varepsilon, \quad (5.119)$$

where  $\beta^*$  is on the segment joining  $\beta^0$  and  $\beta$ . Note that by Assumption 5.26  $\bar{m}$  is positive definite. Hence, by the argument used in (5.26), the smallest eigenvalue of  $h(\theta) \bar{m}^{-1} [h(\theta)]'$  is bounded below for all  $\theta$  in  $\Omega$  by a positive constant. Thus, the consistency of  $\tilde{\beta}_n$  follows from the argument used in the proof of Theorem 5.1.  $\square$

### 5. Limiting distribution

Since the estimator  $\tilde{\beta}_n$  involves the term  $d\hat{Y}_{nt}$  which is a function of  $f_{xx}(\hat{x}_t; \hat{\beta}_n)$ , we need assumptions on the third derivative of  $f$  to derive the limiting distribution of  $\tilde{\beta}_n$ .

Assumption 5.27. Let  $f_{xxi}(z; \theta)$  be the  $i$ -th row of  $f_{xx}(z; \theta)$ ,  $i = 1, 2, \dots, q$ . Then,

$$f_{xxi\beta}(z; \theta) = \frac{\partial}{\partial \theta'} f_{xxi}(z; \theta)$$

exists and is continuous on  $\Gamma \times \Omega$ , for  $i = 1, 2, \dots, q$ .

Assumption 5.28. There exists a constant  $K_7$  such that for every  $i = 1, 2, \dots, q$ , all  $z$  in  $\Gamma$ , and all  $\theta$  in  $\Omega$ , the absolute value of each element of  $f_{xxi\beta}(z; \theta)$  is bounded by  $K_7$ .

We recall that the condition  $a_n^{-1} = o(n^{-1/2})$  was required to obtain the limiting distribution of  $n^{1/2}(\hat{\beta}_n - \beta^0)$ . The next theorem shows that the condition  $a_n^{-1} = o(n^{-1/3})$  is enough to derive the limiting distribution of  $n^{1/2}(\tilde{\beta}_n - \beta^0)$ .

Theorem 5.11. Let Assumptions 5.1a, 5.2, 5.4a, 5.5 through 5.9, 5.10a,

5.11 through 5.15, 5.16a, 5.17, 5.19, 5.20a, and 5.21 through 5.28 hold.

Assume that

$$a_n^{-1} = o(n^{-1/3}) \quad (5.120)$$

Then,

$$n^{1/2} (\tilde{\beta}_n - \beta^0) \xrightarrow{L} N(Q, Y^*),$$

where

$$Y^* = \{h^0 \bar{m}^{-1} h^{0'}\}^{-1}.$$

$$h^0 = h(\beta^0) = \lim_{n \rightarrow \infty} b_n^{-1} \sum_{t=1}^{b_n} f_{\beta^0 t}^0 W_{nt},$$

$$\bar{m} = \lim_{n \rightarrow \infty} b_n^{-1} \sum_{t=1}^{b_n} a_n \sigma_{vnt}^2 W_{nt}' W_{nt}.$$

Proof. We observe that for  $i = 1, 2, \dots, k$ ,

$$\begin{aligned} & \{b_n^{-1} \sum_{t=1}^{b_n} [Y_{nt}^* - f(X_{nt}; \tilde{\beta}_n)] W_{nt}\} \hat{m}_n^{-1} \{b_n^{-1} \sum_{s=1}^{b_n} W_{ns}' f_{\beta^0 i}(X_{ns}; \tilde{\beta}_n)\} \\ &= \{b_n^{-1} \sum_{t=1}^{b_n} [Y_{nt}^* - f(X_{nt}; \beta^0)] W_{nt}\} \hat{m}_n^{-1} \{b_n^{-1} \sum_{s=1}^{b_n} W_{ns}' f_{\beta^0 i}(X_{ns}; \beta^0)\} \end{aligned}$$

$$\begin{aligned}
& + \left\{ b_n^{-1} \sum_{t=1}^{b_n} [Y_{nt}^* - f(\tilde{X}_{nt}; \beta_n^*)] W_{nt} \right\} \hat{m}_n^{-1} \left\{ b_n^{-1} \sum_{s=1}^{b_n} W'_{ns} f_{\beta\beta i}(\tilde{X}_{ns}; \beta_n^*) \right\} \\
& - \left\{ b_n^{-1} \sum_{s=1}^{b_n} f_{\beta i}(\tilde{X}_{ns}; \beta_n^*) W_{ns} \right\} \hat{m}_n^{-1} \left\{ b_n^{-1} \sum_{t=1}^{b_n} W'_{nt} f'_{\beta}(\tilde{X}_{nt}; \beta_n^*) \right\} (\tilde{\beta}_n - \beta^0),
\end{aligned} \tag{5.121}$$

where  $\beta_n^*$  is on the segment joining  $\beta^0$  and  $\tilde{\beta}_n$ . By the argument used to obtain (5.34), for every  $\alpha > 0$  and  $i = 1, 2, \dots, k$ ,

$$\left\{ b_n^{-1} \sum_{t=1}^{b_n} [Y_{nt}^* - f(\tilde{X}_{nt}; \tilde{\beta}_n)] W_{nt} \right\} \hat{m}_n^{-1} \left\{ b_n^{-1} \sum_{s=1}^{b_n} W'_{ns} f_{\beta i}(\tilde{X}_{ns}; \tilde{\beta}_{ns}) \right\} = o_p(n^{-\alpha}). \tag{5.122}$$

Also, by Assumptions 5.21 and 5.27,

$$\begin{aligned}
Y_{nt}^* - f(\tilde{X}_{nt}; \beta^0) &= v_{nt}^* - \frac{1}{2} u_{nt} f_{xx}^0 u'_{nt} - \frac{1}{6} \sum_{i=1}^q u_{nt} f_{xxix}(\tilde{X}_{nt}; \beta^0) u_{nt} u_{nti} \\
&+ \frac{1}{2} \text{tr} \{ [f_{xx}^0 + \sum_{i=1}^q f_{xxix}(\tilde{x}_t; \beta_n^*) (\hat{x}_{ti} - x_{ti}^0)] \hat{\Sigma}_{uun} \} \\
&+ \sum_{j=1}^k f_{xxj\beta}(\tilde{x}_t; \beta_n^*) (\hat{\beta}_{nj} - \beta^0) \hat{\Sigma}_{uun} \}
\end{aligned}$$

$$\begin{aligned}
&= v_{nt}^* - \frac{1}{2} \operatorname{tr} \{ f_{\tilde{x}xt}^0 (\tilde{u}_{nt}' \tilde{u}_{nt} - \tilde{\Sigma}_{uun}) \} \\
&\quad + \frac{1}{2} \operatorname{tr} \{ f_{\tilde{x}xt}^0 (\hat{\tilde{\Sigma}}_{uun} - \tilde{\Sigma}_{uun}) \} \\
&\quad - \frac{1}{6} \sum_{i=1}^q \tilde{u}_{nt} f_{\tilde{x}xi}(\tilde{x}_{nt}^*; \beta^0) \tilde{u}_{nt} u_{nti} \\
&\quad + \frac{1}{2} \operatorname{tr} \left\{ \left[ \sum_{i=1}^q f_{\tilde{x}xi}(\tilde{x}_t^*; \beta_n^*) (\hat{x}_{ti} - x_{ti}^0) \right. \right. \\
&\quad \left. \left. + \sum_{j=1}^q f_{\tilde{x}xj\beta}(\tilde{x}_t^*; \beta_n^*) (\hat{\beta}_{nj} - \beta_j^0) \right] \hat{\tilde{\Sigma}}_{uun} \right\}, \\
\end{aligned} \tag{5.123}$$

where  $\tilde{x}_{nt}^*$  is on the segment joining  $\tilde{x}_t^0$  and  $\tilde{x}_{nt}$ ,  $\tilde{x}_t^*$  is on the segment joining  $\tilde{x}_t^0$  and  $\hat{\tilde{x}}_t$ ,  $\beta_n^*$  is on the segment joining  $\beta^0$  and  $\hat{\beta}_n$ ,  $\hat{x}_{ti}$  and  $x_{ti}^0$  are the  $i$ -th elements of  $\hat{\tilde{x}}_t$  and  $\tilde{x}_t^0$ , and  $\hat{\beta}_{nj}$  and  $\beta_j^0$  are the  $j$ -th elements of  $\hat{\beta}_n$  and  $\beta^0$ , and

$$v_{nt}^* = e_{nt} - f_{xt}^0 \tilde{u}_{nt}.$$

By Assumptions 5.4a, 5.15, and 5.17,

$$E\left\{\left|b_n^{-1} \sum_{t=1}^{b_n} \text{tr}[\tilde{f}_{\text{xxt}}^0(\tilde{u}'_{nt} \tilde{u}_{nt} - \tilde{\Sigma}_{uun})] \tilde{W}_{nt}\right|^2\right\} = O(b_n^{-1} a_n^{-2}) .$$

Thus,

$$b_n^{-1} \sum_{t=1}^{b_n} \text{tr}[\tilde{f}_{\text{xxt}}^0(\tilde{u}'_{nt} \tilde{u}_{nt} - \tilde{\Sigma}_{uun})] \tilde{W}_{nt} = o_p(n^{-1/2} a_n^{-1/2}) . \quad (5.124)$$

It follows from (5.102) and Assumptions 5.4a and 5.15 that

$$b_n^{-1} \sum_{t=1}^{b_n} \text{tr}[\tilde{f}_{\text{xxt}}^0(\hat{\Sigma}_{uun} - \tilde{\Sigma}_{uun})] \tilde{W}_{nt} = o_p(\max\{n^{-1/2} a_n^{-1/2}, a_n^{-3/2}\}) . \quad (5.125)$$

By Assumptions 5.1a and 5.17, for  $i, j = 1, 2, \dots, q$ ,

$$\begin{aligned} E\{|\tilde{u}_{nti} \tilde{u}_{ntj} \tilde{u}_{nt}|}\} &< [E\{\tilde{u}_{nti}^2 \tilde{u}_{ntj}^2\} E\{|\tilde{u}_{nt}|^2\}]^{1/2} \\ &= O(a_n^{-3/2}) . \end{aligned}$$

Therefore, by Assumptions 5.4a and 5.22, for  $i = 1, 2, \dots, q$ ,

$$b_n^{-1} \sum_{t=1}^{b_n} \tilde{u}_{nt} f_{xxix}(\tilde{x}_{nt}^*; \beta^0) \tilde{u}_{nt}' u_{nti} \tilde{w}_{nt} = o_p(a_n^{-3/2}) \quad (5.126)$$

Since

$$\hat{\tilde{\Sigma}}_{uun} = o_p(a_n^{-1}) \quad , \quad (5.127)$$

it follows from Assumptions 5.4a and 5.28 and Lemma 5.2 that for

$i = 1, 2, \dots, q$ ,

$$b_n^{-1} \sum_{t=1}^{b_n} \text{tr}[\tilde{f}_{xxix}(\tilde{x}_t^*; \beta^*) (\hat{x}_{ti} - x_{ti}^0) \hat{\tilde{\Sigma}}_{uun}] \tilde{w}_{nt} = o_p(a_n^{-3/2}) \quad (5.128)$$

By (5.100), (5.127), and Assumptions 5.4a and 5.28, for  $j = 1, 2, \dots, k$ ,

$$b_n^{-1} \sum_{t=1}^{b_n} \text{tr}[\tilde{f}_{xxj\beta}(\tilde{x}_t^*; \beta^*) (\hat{\beta}_{nj} - \beta_j^0) \hat{\tilde{\Sigma}}_{uun}] \tilde{w}_{nt} = o_p(\max[n^{-1/2} a_n^{-1}, a_n^{-2}]) \quad (5.129)$$

Since by Assumption 5.26,

$$V\{b_n^{-1} \sum_{t=1}^{b_n} v_{nt}^* \tilde{W}_{nt}\} = b_n^{-2} \sum_{t=1}^{b_n} \sigma_{vnt}^2 \tilde{W}_{nt}' \tilde{W}_{nt}$$

$$= O(n^{-1}) \quad ,$$

We have

$$b_n^{-1} \sum_{t=1}^{b_n} v_{nt}^* \tilde{W}_{nt} = O_p(n^{-1/2}) \quad . \quad (5.130)$$

It follows from (5.120), (5.123) through (5.126), (5.128), (5.129), and (5.130) that

$$b_n^{-1} \sum_{t=1}^{b_n} [Y_{nt}^* - f(\tilde{X}_{nt}; \beta^0)] \tilde{W}_{nt} = b_n^{-1} \sum_{t=1}^{b_n} v_{nt}^* \tilde{W}_{nt} + O_p(\max[n^{-1/2} a_n^{-1/2}, a_n^{-3/2}])$$

$$= b_n^{-1} \sum_{t=1}^{b_n} v_{nt}^* \tilde{W}_{nt} + o_p(n^{-1/2})$$

$$= O_p(n^{-1/2}) \quad . \quad (5.131)$$

We observe that

$$\begin{aligned}
 b_n^{-1} \sum_{t=1}^{b_n} [Y_{nt}^* - f(\tilde{x}_{nt}; \beta^*)] \tilde{w}_{nt} &= b_n^{-1} \sum_{t=1}^{b_n} [Y_{nt}^* - f(\tilde{x}_{nt}; \beta^0)] \tilde{w}_{nt} \\
 &\quad - (\beta_n^* - \beta^0)' b_n^{-1} \sum_{t=1}^{b_n} \tilde{f}_\beta(\tilde{x}_{nt}; \beta^{**}) \tilde{w}_{nt} ,
 \end{aligned} \tag{5.132}$$

where  $\beta_n^{**}$  is on the segment joining  $\beta^0$  and  $\beta_n^*$ . By Assumptions 5.1a, 5.4a, 5.9, and 5.13,

$$\begin{aligned}
 b_n^{-1} \sum_{t=1}^{b_n} \tilde{f}_\beta(\tilde{x}_{nt}; \beta_n^{**}) \tilde{w}_{nt} &= b_n^{-1} \sum_{t=1}^{b_n} [\tilde{f}_\beta(\tilde{x}_t^0; \beta^{**}) \tilde{w}_{nt} + \tilde{u}_{nt} \tilde{f}_{x\beta}(\tilde{x}_{nt}^*; \beta_n^{**}) \tilde{w}_{nt}] \\
 &= o_p(1) ,
 \end{aligned} \tag{5.133}$$

where  $\tilde{x}_{nt}^*$  is on the segment joining  $\tilde{x}_t^0$  and  $\tilde{x}_{nt}$ . Therefore, it follows from (5.100), (5.131), (5.132), and (5.133) that

$$b_n^{-1} \sum_{t=1}^{b_n} [Y_{nt}^* - f(\tilde{x}_{nt}; \beta_n^*)] \tilde{w}_{nt} = o_p(\max[n^{-1/2}, a_n^{-1}]) . \tag{5.134}$$

Using Assumptions 5.4a and 5.13, we have

$$b_n^{-1} \sum_{s=1}^{b_n} \tilde{W}'_{ns} \tilde{f}_{\beta i}(\tilde{x}_{ns}; \beta_n^*) = o_p(1) . \quad (5.135)$$

By the argument used to obtain (5.44) and (5.45) in the proof of Lemma 5.1,

$$b_n^{-1} \sum_{t=1}^{b_n} \tilde{f}_{\beta}(\tilde{x}_{nt}; \beta_n^*) \tilde{W}_{nt} = \tilde{h}^0 + o_p(1) , \quad (5.136)$$

$$b_n^{-1} \sum_{t=1}^{b_n} \tilde{f}_{\beta}(\tilde{x}_{nt}; \beta^0) \tilde{W}_{nt} = \tilde{h}^0 + o_p(1) . \quad (5.137)$$

By (5.117) and Assumption 5.26,

$$\hat{\tilde{m}}_n^{-1} = \tilde{m}^{-1} + o_p(1) . \quad (5.138)$$

It follows from (5.121), (5.122), (5.131), and (5.134) through (5.138) that

$$\tilde{\beta}_n - \beta^0 = \{h^0 \bar{m}^{-1} h^{0'}\}^{-1} h^0 \bar{m}^{-1} \{b_n^{-1} \sum_{t=1}^{b_n} \tilde{w}_{nt}' v_{nt}^*\} + o_p(n^{-1/2}) . \quad (5.139)$$

By Assumption 5.16a, Theorem 2.12, and the argument used to obtain (5.55),

$$n^{1/2} b_n^{-1} \sum_{t=1}^{b_n} \tilde{w}_{nt}' v_{nt}^* \xrightarrow{L} N(0, \bar{\Sigma}) . \quad (5.140)$$

Hence, the result follows from (5.139) and (5.140).  $\square$

As we noted earlier, the modified instrumental variable estimator  $\tilde{\beta}_n$  takes into account the differences among  $\sigma_{vnt}^2$ , which the instrumental variable estimator  $\hat{\beta}_n$  ignores. The following corollary shows that  $\tilde{\beta}_n$  is asymptotically more efficient than  $\hat{\beta}_n$  in the sense that the difference of the two limiting covariance matrices is positive semidefinite.

Corollary 5.11.1. Let Assumptions 5.1a, 5.2, 5.4a, 5.5 through 5.9, 5.10a, 5.11 through 5.15, 5.16a, 5.17, 5.19, 5.20a, and 5.21 through 5.28 hold. If  $a_n^{-1} = o(n^{-1/3})$ , then

$$n^{1/2} (\tilde{\beta}_n - \beta^0) \xrightarrow{L} N(0, \tilde{V}^*) ,$$

and if  $a_n^{-1} = o(n^{-1/2})$ , then

$$n^{1/2} (\hat{\beta}_n - \beta^0) \xrightarrow{L} N(Q, \underline{V}) ,$$

where  $\underline{V}^*$  is given in Theorem 5.11, and  $\underline{V}$  is given in (5.99).

Furthermore, the difference  $[\underline{V} - \underline{V}^*]$  is positive semidefinite.

Proof. The limiting distributions are given in Theorem 5.11 and (5.99).

We observe that

$$\underline{V} - \underline{V}^* = \underline{R}' \underline{\bar{m}}^{-1} \underline{R} ,$$

where

$$\underline{R} = \underline{\bar{m}} \underline{h}^0, \{ \underline{h}^0 \underline{h}^{0'} \}^{-1} - \underline{h}^0, \{ \underline{h}^0 \underline{\bar{m}}^{-1} \underline{h}^{0'} \}^{-1} .$$

Since  $\underline{\bar{m}}$  is positive definite by Assumption 5.26,  $\underline{\bar{m}}^{-1}$  is also positive definite. Thus, the result follows.  $\square$

The next two theorems provide two consistent estimators of the covariance matrix of the limiting distribution of  $\tilde{\beta}_n$ .

Theorem 5.12. Let Assumptions 5.1a, 5.2, 5.4a, 5.5 through 5.9, 5.10a, 5.11 through 5.15, 5.16a, 5.17, 5.19a, 5.20a, and 5.21 through 5.28 hold. Also let,

$$\hat{\underline{V}}(\tilde{\beta}_n) = \{ \tilde{h} \tilde{m}^{-1} \tilde{h}' \}^{-1} ,$$

where

$$\tilde{h} = b_n^{-1} \sum_{t=1}^{b_n} \tilde{f}_{\beta}(\hat{x}_t; \tilde{\beta}_n) \tilde{w}_{nt} ,$$

$$\tilde{m} = b_n^{-2} \sum_{t=1}^{b_n} \hat{\sigma}_{vnt}^2 \tilde{w}_{nt}' \tilde{w}_{nt} .$$

Then,

$$n \tilde{y}(\hat{\beta}_n) \xrightarrow{P} y^* ,$$

where  $y^*$  is given in Theorem 5.11.

Proof. We observe that

$$\begin{aligned} \tilde{h} &= b_n^{-1} \sum_{t=1}^{b_n} \tilde{f}_{\beta}^0 \tilde{w}_{nt} + b_n^{-1} \sum_{t=1}^{b_n} \tilde{f}_{\beta x}(\tilde{x}_t^*; \tilde{\beta}_n^*)(\hat{x}_t - \tilde{x}_t^0)' \tilde{w}_{nt} \\ &\quad + b_n^{-1} \sum_{t=1}^{b_n} \tilde{f}_{\beta \beta}(\tilde{x}_t^*; \tilde{\beta}_n^*)(\tilde{\beta}_n - \tilde{\beta}^0) \tilde{w}_{nt} , \end{aligned}$$

where  $\tilde{x}_t^*$  is on the segment joining  $\tilde{x}_t^0$  and  $\hat{x}_t$ , and  $\tilde{\beta}_n^*$  is on the segment joining  $\tilde{\beta}^0$  and  $\tilde{\beta}_n$ . Thus, by (5.137), Lemma 5.2,

Assumptions 5.4a and 5.13,

$$\tilde{h} = h^0 + o_p(1) \quad .$$

Also, by (5.117),

$$\begin{aligned} n \tilde{m} &= \hat{m}_n \\ &= \bar{m} + o_p(1) \quad . \end{aligned}$$

Hence, the result follows.  $\square$

Theorem 5.13. Let Assumptions 5.1a, 5.2, 5.4a, 5.5 through 5.9, 5.10a, 5.11 through 5.15, 5.16a, 5.17, 5.19, 5.20a, and 5.21 through 5.28 hold.

Also, let

$$\tilde{V}(\tilde{\beta}_n) = \{\tilde{h} \tilde{V}_o^{-1} \tilde{h}'\}^{-1} \quad ,$$

where

$$\tilde{V}_o = b_n^{-2} \sum_{t=1}^{b_n} \tilde{w}_{nt}' \tilde{w}_{nt} [y_{nt}^* - f(\tilde{x}_{nt}; \tilde{\beta}_n)]^2 \quad ,$$

and  $\tilde{h}$  is defined in Theorem 5.12. Then,

$$n \tilde{V}(\tilde{\beta}_n) \xrightarrow{P} V^* \quad ,$$

where  $\tilde{y}^*$  is given in Theorem 5.11.

Proof. By Theorem 5.11 and the argument used to obtain (5.65) in the proof of Theorem 5.3,

$$\begin{aligned} n \tilde{y}_o &= \tilde{m}_n + o_p(b_n^{-1/2}) \\ &= \bar{m} + o_p(1) . \end{aligned}$$

Hence, the result follows.  $\square$

The quantity  $\tilde{y}_o$  seems to estimate the true variability of

$$b_n^{-1} \sum_{t=1}^{b_n} \tilde{w}_{nt}' v_{nt}^*$$

more efficiently than the quantity  $\tilde{m}$ . Also, the estimator  $\tilde{y}(\tilde{\beta}_n)$  has a form analogous to the estimator of the covariance matrix of the two-stage least squares estimator. Thus, we prefer  $\tilde{y}(\tilde{\beta}_n)$  to  $\hat{y}(\tilde{\beta}_n)$  as an estimator of the limiting covariance matrix of  $\tilde{\beta}_n$ .

## 6. Bias

As shown in (5.101), the instrumental variable estimator  $\hat{\beta}_n$  has a bias of  $o_p(a_n^{-1})$  which is due to the nonlinearity of the functional form. The next theorem shows that the modified instrumental variable estimator  $\tilde{\beta}_n$  does not have such a bias. That is  $\tilde{\beta}_n$  is unbiased up to the terms of  $o_p(a_n^{-1})$ .

Theorem 5.14. Let Assumptions 5.1a, 5.2, 5.4a, 5.5 through 5.9, 5.10a, 5.11 through 5.15, 5.17, 5.19, 5.20a, and 5.21 through 5.28 hold. Then, the expectation of the expansion of  $(\tilde{\beta}_n - \beta^0)$  through terms of  $O_p(a_n^{-1})$  is zero.

Proof. Since

$$\max[n^{-1/2} a_n^{-1/2}, a_n^{-3/2}] = o_p(a_n^{-1})$$

we have from (5.131) that

$$b_n^{-1} \sum_{t=1}^{b_n} [Y_{nt}^* - f(\underline{x}_{nt}; \beta^0)] W_{nt} = b_n^{-1} \sum_{t=1}^{b_n} v_{nt}^* + o_p(a_n^{-1}) .$$

Hence, the result follows from (5.121), (5.122), and (5.134) through (5.138).  $\square$

#### F. Modified Estimators of the Error Covariance

##### Matrix and the True Value

We observe from the expressions (5.121), (5.131), and (5.134) that without the assumption  $a_n^{-1} = o(n^{-1/3})$  the modified instrumental variable estimator  $\tilde{\beta}_n$  satisfies

$$\tilde{\beta}_n - \beta^0 = O_p(\max[n^{-1/2}, a_n^{-3/2}]) . \quad (5.141)$$

As we discussed in the previous section, the estimator  $\tilde{\beta}_n$  has better properties than the instrumental variable estimator  $\hat{\beta}_n$ . In Sections C and D, we introduced the estimation procedures based on  $\hat{\beta}_n$  for the error covariance matrix and the unknown true values. Hence, we may expect to obtain improved estimators of  $\Sigma_n$  and  $x_t^0$  by applying such procedures with  $\tilde{\beta}_n$  replacing  $\hat{\beta}_n$  and  $Y_{nt}^*$  replacing  $Y_{nt}$ . The modified estimator  $\tilde{\Sigma}_n$  of  $\Sigma_n$  is defined by

$$\text{vech } \tilde{\Sigma}_n = \left\{ \sum_{t=1}^{b_n} \tilde{\eta}_{nt}' \tilde{\eta}_{nt} \right\}^{-1} \sum_{t=1}^{b_n} \tilde{\eta}_{nt}' \tilde{v}_{nt}^2, \quad (5.142)$$

where

$$\tilde{\eta}_{nt} = (\tilde{F}_{nt} \otimes \tilde{F}_{nt}) \tilde{\eta}_{q+1},$$

$$\tilde{F}_{nt} = [1, -f_x(\hat{x}_t; \tilde{\beta}_n)],$$

$$\tilde{v}_{nt} = Y_{nt}^* - f(\hat{x}_t; \tilde{\beta}_n).$$

Note that we have replaced  $\hat{x}_{nt}$  in  $\hat{\eta}_{nt}$  of (5.78) by  $\hat{x}_t$ . Since  $\hat{\eta}_{nt}$  or  $\tilde{\eta}_{nt}$  are used as an estimator of  $\eta_t$  in (5.77), we use  $\hat{x}_t$  which improves over  $\hat{x}_{nt}$  as an estimator of  $x_t^0$ . We expect  $\tilde{v}_{nt}$  to be a better estimator of  $v_{nt}^* = e_{nt} - f_{xt}^0 u'_{nt}$  than  $\hat{v}_{nt}$  used in Section C.

We observe that  $\delta_{nt}$  and  $v_{nt}^*$  are uncorrelated, and are independent if  $(e_{nt}, u_{nt})$  are normally distributed. This observation provides an additional intuitive justification for the use of least squares method in (5.142), since errors in  $\tilde{\eta}_{nt}$  are approximately functions of  $\delta_{nt}$ .

The  $\tilde{\eta}_{nt}$  and  $\tilde{v}_{nt}$  satisfy the conditions  $\hat{\eta}_{nt}$  and  $\hat{v}_{nt}$  satisfied in Section C. Thus, the properties of  $\hat{\Sigma}_n$  obtained in Section C are also valid for  $\tilde{\Sigma}$ . Since errors in  $\tilde{\eta}_{nt}$  as an estimator of  $\eta_t$  are still of  $O_p(a_n^{-1/2})$ , the estimator  $\tilde{\Sigma}_n$  is equivalent to  $\hat{\Sigma}_n$  up to the level of approximation considered here.

**Theorem 5.15.** Let Assumptions 5.1a, 5.2, 5.4a, 5.5 through 5.9, 5.10a, 5.11 through 5.15, 5.17, 5.19, 5.20a, and 5.21 through 5.28 hold. Then,

$$\tilde{\Sigma}_n - \Sigma_n = O_p(\max[a_n^{-3/2}, n^{-1/2} a_n^{-1/2}]) .$$

If, in addition,

$$a_n^{-1} = o(n^{-1/3}) ,$$

then

$$n^{1/2} \text{vech}(\tilde{\Sigma}_n - \Sigma_n) \xrightarrow{L} N(0, V_\sigma) ,$$

where  $V_\sigma$  is given in Theorem 5.7.

**Proof.** The results follow from (5.141) and the argument used in the

proofs of Theorem 5.6 and Theorem 5.7.  $\square$

Even though the asymptotic properties of  $\tilde{\Sigma}_n$  and  $\hat{\Sigma}_n$  are the same, we expect  $\tilde{\Sigma}_n$  to be a better estimator of  $\Sigma_n$  than  $\hat{\Sigma}_n$ . This is because  $\tilde{\eta}_{nt}$  and  $\tilde{v}_{nt}$  are superior to  $\hat{\eta}_{nt}$  and  $\hat{v}_{nt}$  as estimators of  $\eta_t$  and  $v_{nt}^*$  respectively.

Using the estimator  $\tilde{\Sigma}_n$ , we define the modified estimator  $\tilde{x}_t$  of  $x_t^0$  to be the value of  $x_t$  contained in  $I$  that minimizes

$$D_n^*(x_t) = [Y_{nt}^* - f(x_t; \tilde{\beta}_n), x_{nt} - x_t] \tilde{\Sigma}_n^{-1} [Y_{nt}^* - f(x_t; \tilde{\beta}_n), x_{nt} - x_t]' . \quad (5.143)$$

As in Chapter IV, the bias of  $\hat{\beta}_n$  due to the curvature may help cancelling the bias in  $\hat{x}_t$ . Also, the adjustment  $d\hat{Y}_{nt}$  in (5.143) may provide a further reduction in the bias for  $\tilde{x}_t$ , as for the adjusted estimator based on the local quadratic approximation discussed in Chapter IV. However the errors in  $a_{n\tilde{\Sigma}_n}^{\hat{\Sigma}_n}$  or  $a_{n\tilde{\Sigma}_n}^{\tilde{\Sigma}_n}$  are of  $O_p(\max[b_n^{-1/2}, a_n^{-1/2}])$ . Thus, the higher order expansions of  $\hat{x}_t$  and  $\tilde{x}_t$  involve the error terms in  $\hat{\Sigma}_n$  and  $\tilde{\Sigma}_n$ , and are not as simple as those of the estimators in Chapter IV. We do not discuss the higher order results for the estimators of  $\Sigma_n$  and  $x_t^0$  in this chapter.

The property of  $\hat{x}_t$  given in Section D also holds for  $\tilde{x}_t$ .

**Theorem 5.16.** Let Assumptions 5.1a, 5.2, 5.4a, 5.5 through 5.9, 5.10a, 5.11 through 5.15, 5.17, 5.19, 5.20a, and 5.21 through 5.28 hold. Then,

$$\tilde{x}_t - x_t^0 = \delta_{nt} + O_p(\max[n^{-1/2}, a_n^{-1}]) ,$$

where  $\delta_{nt}$  is given in Theorem 5.8.

Proof. The adjustment term  $d\hat{Y}_{nt}$  is of  $O_p(a_n^{-1})$ . Hence, the result follows from (5.141), Theorem 5.15, and the argument used in the proof of Theorem 5.8.  $\square$

The estimator of  $y_t^0$  and the local approximation to  $\tilde{x}_t$  can be obtained in a manner similar to those of Section D. Also, results similar to Corollary 5.8.1, Theorem 5.9, and Corollary 5.9.1 can be derived by the argument used in Section D.

#### G. An Iterative Estimation Procedure

The results in Sections E and F suggest that a further improved estimator of  $\beta^0$  be obtained based on the modified estimators  $\tilde{\beta}_n$ ,  $\tilde{\Sigma}_n$ , and  $\tilde{x}_t$ . The modified instrumental variable estimator  $\tilde{\beta}_n$  was obtained by minimizing  $Q_n^*(\beta)$  in (5.104) which involves the estimates  $\hat{\sigma}_{vnt}^2$  and  $d\hat{Y}_{nt}$  of  $\sigma_{vnt}^2$  and the bias adjustment respectively. We may expect to estimate  $\sigma_{vnt}^2$  and the bias adjustment more efficiently than  $\hat{\sigma}_{vnt}^2$  and  $d\hat{Y}_{nt}$  by

$$\tilde{\sigma}_{vnt}^2 = [1, -f_x(\tilde{x}_t; \tilde{\beta}_n)] \tilde{\Sigma}_n [1, -f_x(\tilde{x}_t; \tilde{\beta}_n)]',$$

$$d\tilde{Y}_{nt} = \frac{1}{2} \text{tr} [f_{xx}(\tilde{x}_t; \tilde{\beta}_n) \tilde{\Sigma}_{uun}] ,$$

where  $\tilde{\Sigma}_{uun}$  is the portion of  $\tilde{\Sigma}_n$  that corresponds to  $u_{nt}$ . Hence, by repeating the minimization of  $Q_n^*(\beta)$  with  $\tilde{\sigma}_{vnt}^2$  and  $d\tilde{Y}_{nt}$  replacing  $\hat{\sigma}_{vnt}^2$  and  $d\hat{Y}_{nt}$ , we expect to obtain an improved

estimator of  $\beta^0$ . We suggest that in practice such a procedure be iterated to obtain estimates of the parameters.

In the theoretical development, we have used sequences  $\{a_n\}$  and  $\{b_n\}$  to investigate various asymptotic properties. We assumed that the number  $b_n$  of data points is known but the value of  $a_n$  is unknown. Thus, we apply our estimation procedure to situations without knowledge of the decreasing error variances. When either the number of observations  $N$  is large or the error variances can be considered to be small relative to total variation, the asymptotic results for our estimation procedure can be used as approximations.

We introduce an iterative estimation procedure for practical use. In the following procedure, we let  $N$  be the number of observations, and denote the  $i$ -th stage estimates of  $\beta^0$ ,  $\tilde{x}_t^0$ , and the covariance matrix  $\Sigma$  of errors in  $(Y_t, \tilde{x}_t)$  by  $\beta^{(i)}$ ,  $\tilde{x}_t^{(i)}$ , and  $\Sigma^{(i)}$ , respectively. Also, let

$$\Sigma^{(i)} = \begin{bmatrix} \sigma_{ee}^{(i)} & \Sigma_{eu}^{(i)} \\ \Sigma_{ue}^{(i)} & \Sigma_{uu}^{(i)} \end{bmatrix}.$$

Let  $I_{\max}$  be a predetermined maximum number of iterations. Define  $m_{ZZ.W}$  to be the residual mean square matrix of  $Z_t = (Y_t, \tilde{x}_t)$  obtained by regressing  $Z_t$  on  $W_t$ . In our iterative procedure, we introduce certain modifications to the instrumental variable estimator and the modified instrumental variable estimator. Such modifications are

associated with a predetermined positive constant  $\alpha$ . The discussion of the  $\alpha$ -modification will be given after the presentation of the procedure.

Procedure:

1. Find the  $\beta$  that minimizes

$$Q(\beta) = \left\{ \sum_{t=1}^N [Y_t - f(\underline{x}_t; \beta)] \underline{w}_t' \right\} \left\{ \sum_{t=1}^N \underline{w}_t' \underline{w}_t \right\}^{-1} \left\{ \sum_{t=1}^N \underline{w}_t' [Y_t - f(\underline{x}_t; \beta)] \right\} \\ + (N-p-\alpha)^{-1} \alpha \sum_{t=1}^N [Y_t - f(\underline{x}_t; \beta)]^2 .$$

The resulting  $\beta$  is  $\beta^{(0)}$ . Let  $\underline{x}_t^{(0)} = \underline{x}_t$ ,  $Y_t^{(0)} = Y_t$ , and  $i = 1$ .

2. Obtain  $\underline{\Sigma}^{(i)}$  by the ordinary least squares estimation

$$\text{vech } \underline{\Sigma}^{(i)} = \left\{ \sum_{t=1}^N \underline{\eta}_t^{(i-1)} \underline{\eta}_t^{(i-1)'} \right\}^{-1} \sum_{t=1}^N \underline{\eta}_t^{(i-1)} [v_t^{(i-1)}]^2 ,$$

where

$$\underline{\eta}_t^{(i-1)} = \{ \underline{F}_t^{(i-1)} \otimes \underline{F}_t^{(i-1)} \} \underline{\Phi}_{q+1} ,$$

$$\underline{F}_t^{(i-1)} = [1, -\underline{f}_{\underline{x}}(\underline{x}_t^{(i-1)}; \beta^{(i-1)})] ,$$

$$v_t^{(i-1)} = y_t^{(i-1)} - f(\tilde{x}_t; \beta^{(i-1)}) .$$

3. Let  $\lambda_j^{(i)}$ ,  $j = 1, 2, \dots, q+1$ , be the eigenvalues of  $\tilde{m}_{ZZ.W}^{-1/2} \tilde{\Sigma}^{(i)} \tilde{m}_{ZZ.W}^{-1/2}$ , and let  $\tilde{P}^{(i)}$  be a matrix of the corresponding eigenvectors. For  $j = 1, 2, \dots, q+1$ , let

$$\begin{aligned} \lambda_j^{(i)} &= 0, & \text{if } \lambda_j^{(i)} < 0, \\ &= \lambda_j^{(i)}, & \text{if } 0 < \lambda_j^{(i)} < 1, \\ &= 1, & \text{if } \lambda_j^{(i)} > 1. \end{aligned}$$

4. Compute

$$\tilde{\Sigma}^{(i)} = \tilde{P}^{(i)} \tilde{\Lambda}^{(i)} \tilde{P}^{(i)'} ,$$

where

$$\tilde{\Lambda}^{(i)} = \text{diag}\{\lambda_1^{(i)}, \lambda_2^{(i)}, \dots, \lambda_{q+1}^{(i)}\} .$$

5. For each  $t$ , let

$$\tilde{x}_t^{(i)} = \tilde{x}_t - v_t^{(i-1)} [\sigma_{vvt}^{(i)}]^{-1} \tilde{\Sigma}_{vut}^{(i)} ,$$

$$y_t^{(i)} = y_t + \frac{1}{2} \text{tr} [f_{\tilde{xx}}(\tilde{x}_t^{(i)}; \beta^{(i-1)}) \tilde{\Sigma}_{uu}^{(i)}] ,$$

where

$$\sigma_{vvt}^{(i)} = F_t^{(i-1)} \Sigma_t^{(i)} F_t^{(i-1)'},$$

$$\Sigma_{vut}^{(i)} = F_t^{(i-1)} [\Sigma_{ue}^{(i)}, \Sigma_{uu}^{(i)}],$$

6. Find the  $\beta$  that minimizes

$$Q^*(\beta) = \left\{ \sum_{t=1}^N [Y_t^{(i)} - f(X_t; \beta)] W_t \right\} \left\{ \sum_{t=1}^N \sigma_{vvt}^{(i)} W_t' W_t \right\}^{-1} \left\{ \sum_{t=1}^N W_t' [Y_t^{(i)} - f(X_t; \beta)] \right\} \\ + (N-p-\alpha)^{-1} \alpha \sum_{t=1}^N [\sigma_{vvt}^{(i)}]^{-1} [Y_t^{(i)} - f(X_t; \beta)]^2.$$

The resulting  $\beta$  is  $\beta^{(i)}$ . If  $i > 1$  and  $|\beta^{(i)} - \beta^{(i-1)}|$  is less than a predetermined criterion for convergence, or if  $i = I_{\max}$ , then stop. Otherwise, set  $i = i+1$ , and go to 2.

The modification of  $\Sigma^{(i)}$  introduced in the steps 3 and 4 ensures that the matrices  $\Sigma^{(i)}$  and  $[m_{ZZ.W} - \Sigma^{(i)}]$  are nonnegative definite. Note that with the zero correlation between  $W_t$  and  $(e_t, u_t)$  the residual mean square matrix  $m_{ZZ.W}$  estimates  $\Sigma_{ZZ.W} = \Sigma_{ZZ.W} + \Sigma$ , where  $\Sigma_{ZZ.W}$  is a nonnegative definite matrix.

The modification associated with the value  $\alpha$  in  $Q(\beta)$  and  $Q^*(\beta)$  are analogues of the modifications considered by Fuller (1977).

Fuller introduced certain modifications to the two-stage least squares estimator and the limited information maximum likelihood estimator in the linear simultaneous equation model. The modification used in  $Q(\beta)$  and  $Q^*(\beta)$  is the addition of a multiple of the residual sum of squares. Such a modification can be shown to be equivalent to Fuller's modification to the two-stage least squares estimator. Fuller showed that his modified two-stage least squares estimator possesses finite moments of order depending on the number of observations. We expect the minimizations of  $Q(\beta)$  and  $Q^*(\beta)$  to produce more stable solutions than those without modifications. For the linear simultaneous equation model, Fuller (1977) showed that the modification using  $\alpha = 1$  produces an almost unbiased two-stage least squares estimator and limited information maximum likelihood estimator, and that the modification using  $\alpha = 4$  produces a smaller mean squared error of the limited information estimator than any positive  $\alpha$  smaller than 4. For the nonlinear functional relationship model with instrumental variables, no theoretical investigation nor Monte Carlo study for the modifications has been made. We suggest that a value of  $\alpha$  satisfying  $1 \leq \alpha \leq 4$  be used.

We point out that our iterative procedure is not guaranteed to converge. At each stage, the function  $Q^*(\beta)$  to be minimized has different weights  $\sigma_{vvt}^{(i)}$  based on  $\sum^{(i)}$ . Hence, we recommend that a moderate number be used for  $I_{\max}$ , the maximum number of iteration.

A finite number of iterations and the modification associated with the value  $\alpha$  do not change the limiting distributions of the

estimates. Thus, by Theorem 5.13, the covariance matrix of the approximate distribution of the final estimate  $\hat{\beta}$  from the iterative procedure is estimated by

$$\hat{V}(\hat{\beta}) = \{ \hat{h} \hat{V}_0^{-1} \hat{h}' \}^{-1} ,$$

where

$$\hat{h} = \sum_{t=1}^N \hat{f}_{\beta}(\hat{x}_t; \hat{\beta}) \hat{w}_t ,$$

$$\hat{V}_0 = (N-k)^{-1} N \sum_{t=1}^N \hat{w}_t \hat{w}_t' [\hat{y}_t - f(\hat{x}_t; \hat{\beta})]^2 ,$$

and  $\hat{x}_t$  and  $\hat{y}_t$  are the values of  $x_t^{(i)}$  and  $y_t^{(i)}$  at the final stage. Also, the covariance matrix of the approximate distribution of  $\hat{x}_t$  is estimated by

$$\hat{V}(\hat{x}_t) = \hat{\Sigma}_{uu} - \hat{\Sigma}_{uvt} (\hat{\sigma}_{vvt})^{-1} \hat{\Sigma}_{vut} ,$$

where  $\hat{\Sigma}_{uu}$ ,  $\hat{\Sigma}_{uvt}$ , and  $\hat{\sigma}_{vvt}$  are the final estimates of the respective quantities in the iterative procedure.

## VI. BIBLIOGRAPHY

- Adcock, R. J. 1877. Note on the method of least squares. *Analyst* 4:183-184.
- Adcock, R. J. 1878. A problem in least squares. *Analyst* 5:53-54.
- Amemiya, T. 1974. The nonlinear two-stage least-squares estimator. *Journal of Econometrics* 2:105-110.
- Anderson, D. A. 1981. The circular structural model. *J. Roy. Statist. Soc. B* 43:131-141.
- Anderson, T. W. 1951. Estimating linear restrictions on regression coefficients for multivariate normal distributions. *Ann. Math. Statist.* 22:327-351.
- Anderson, T. W. 1958. *Introduction to Multivariate Statistical Analysis*. Wiley, New York. 374 pp.
- Anderson, T. W. 1973. Asymptotically efficient estimation of covariance matrices with linear structure. *Ann. Statist.* 1:135-141.
- Anderson, T. W. 1976. Estimation of linear functional relationships: approximate distributions and connections with simultaneous equations in econometrics. *J. Roy. Statist. Soc. B* 38:1-20.
- Anderson, T. W., and H. Rubin. 1956. Statistical inference in factor analysis. *Proc. Third Berkeley Symposium*. Univ. of Calif. Press, Berkeley.
- Bellman, R. 1970. *Introduction to Matrix Analysis*. 2nd ed. McGraw-Hill, New York. 403 pp.
- Berkson, J. 1950. Are there two regressions? *J. Amer. Statist. Assoc.* 45:164-180.
- Bhargava, A. K. 1977. Maximum likelihood estimation in a multivariate "errors in variables" regression model with unknown error covariance matrix. *Commun. Statist. A* 6:587-601.
- Bhargava, A. K. 1979. Estimation of a linear transformation and an associated distributional problem. *J. Statist. Planning and Inference* 3:19-26.
- Birch, M. W. 1964. A note on the maximum likelihood estimation of a linear structural relationship. *J. Amer. Statist. Assoc.* 59:1175-1178.

- Bock, R. D., and R. E. Bargmann. 1966. Analysis of covariance structures. *Psychometrika* 31:507-534.
- Britt, H. I., and R. H. Luecke. 1973. The estimation of parameters in nonlinear implicit models. *Technometrics* 15:233-247.
- Browne, M. W. 1974. Generalized least squares estimators in the analysis of covariance structures. *South African Statistical Journal* 8:1-24.
- Carter, R. L. 1976. Instrumental variable estimation of the simple errors in variables model. Ph.D. Thesis. Iowa State University. Ames, Iowa. 137 pp.
- Carter, R. L., and W. A. Fuller. 1980. Instrumental variable estimation of the simple errors in variables model. *J. Amer. Statist. Assoc.* 75:687-692.
- Chan, N. N. 1965. On circular functional relationships. *J. Roy. Statist. Soc. B* 27:45-56.
- Chan, N. N. 1980. Estimating linear functional relationships. In K. Matusita, ed. *Recent developments in statistical inference and data analysis*. North-Holland, Amsterdam.
- Chung, K. L. 1974. *A Course in Probability Theory*. 2nd ed. Academic Press, New York. 365 pp.
- Clutton-Brock, M. 1967. Likelihood distributions for estimating functions when both variables are subject to error. *Technometrics* 9:261-269.
- Cochran, W. G. 1968. Errors of measurements in statistics. *Technometrics* 10:637-666.
- Cook, W. R. 1931. On curve fitting by means of least squares. *Philosophical Magazine*, Ser. 7, 12:1025-1039.
- Creasy, M. A. 1956. Confidence limits for the gradient in the linear functional relationship. *J. Roy. Statist. Soc. B* 18:65-69.
- Dahm, P. F. 1979. Estimation of the parameters of the multivariate linear errors in variables model. Ph.D. Thesis. Iowa State University. Ames, Iowa. 165 pp.
- Dahm, P. F., and W. A. Fuller. 1981. Generalized least squares estimation of the functional multivariate linear errors in variables model. Report to the U. S. Census Bureau on J. S. A. 81-14.

- DeGracie, J. S., and W. A. Fuller. 1972. Estimation of the slope and analysis of covariance when the concomitant variable is measured with error. *J. Amer. Statist. Assoc.* 67:930-937.
- Deming, W. E. 1931. The application of least squares. *Philosophical Magazine*, Ser. 7, 11:146-158.
- Deming, W. E. 1943. *Statistical Adjustment of Data*. Wiley, New York. 261 pp.
- Dolby, G. R. 1972. Generalized least squares and maximum likelihood estimation of nonlinear functional relationships. *J. Roy. Statist. Soc. B* 34:393-400.
- Dolby, G. R. 1976. The ultra-structural relation: a synthesis of the functional and the structural relations. *Biometrika* 63:39-50.
- Dolby, G. R., and S. Lipton. 1972. Maximum likelihood estimation of the general nonlinear functional relationship with replicated observations and correlated errors. *Biometrika* 59:121-129.
- Dorff, N., and J. Gurland. 1961. Small sample behavior of slope estimators in a linear functional relation. *Biometrics* 17:283-298.
- Egerton, M. F., and P. J. Laycock. 1979. Maximum likelihood estimation of multivariate non-linear functional relationships. *Math. Operationsforsch. Statist. Ser. Statist.* 10:273-280.
- Fedorov, V. V. 1974. Regression problems with controllable variables subject to error. *Biometrika* 61:49-56.
- Fuller, W. A. 1976. *Introduction to Statistical Time Series*. Wiley, New York. 469 pp.
- Fuller, W. A. 1977. Some properties of a modification of the limited information estimator. *Econometrica* 45:939-953.
- Fuller, W. A. 1980. Properties of some estimators for the errors-in-variables model. *Ann. Statist.* 8:407-422.
- Fuller, W. A. 1982. Measurement error models with heterogeneous error variances. To appear in *Proceedings of the Canadian Conference on Applied Statistics*. Marcel Dekker, New York.
- Fuller, W. A., and M. A. Hidiroglou. 1978. Regression estimation after correcting for attenuation. *J. Amer. Statist. Assoc.* 73:99-104.
- Geary, R. C. 1948. Studies in relations between economic time series. *J. Roy. Statist. Soc. B* 10:140-158.

- Gleser, L. J. 1981. Estimation in a multivariate "errors-in-variables" regression model: large sample results. *Ann. Statist.* 9:24-44.
- Gleser, L. J., and G. S. Watson. 1973. Estimation of a linear transformation. *Biometrika* 60:525-534.
- Griliches, Z., and V. Ringstad. 1970. Errors in variables bias in nonlinear contexts. *Econometrica* 38:368-370.
- Grubbs, F. E. 1948. On testing precision of measuring instruments and product variability. *J. Amer. Statist. Assoc.* 43:243-264.
- Healy, J. D. 1980. Maximum likelihood estimation of a multivariate linear functional relationship. *J. Multivariate Analysis* 10:243-251.
- Hey, E. N., and M. H. Hey. 1960. The statistical estimation of a rectangular hyperbola. *Biometrics* 16:606-617.
- Höschel, H. P. 1978a. Least squares and maximum likelihood estimation of functional relations. In Transactions of the eighth Prague conference on information theory, statistical decision functions, random processes. Academia, Prague.
- Höschel, H. P. 1978b. Generalized least squares estimators of linear functional relations with known error covariance. *Math. Operationsforsch. Statist. Ser. Statist.* 9:9-26.
- Jennrich, R. I. 1969. Asymptotic properties of nonlinear least squares estimators. *Ann. Math. Statist.* 40:633-643.
- Jennrich, R. I., and D. T. Thayer. 1973. A note on Lawley's formulas for standard errors in maximum likelihood factor analysis. *Psychometrika* 38:571-580.
- Jöreskog, K. G. 1967. Some contributions to maximum likelihood factor analysis. *Psychometrika* 32:443-482.
- Jöreskog, K. G. 1970. A general method for analysis of covariance structures. *Biometrika* 57:239-251.
- Jöreskog, K. G. 1973a. A general method for estimating a linear structural equation system. In A. S. Goldberger and O. D. Duncan eds. *Structural equation models in the social sciences*. Seminar Press, New York.
- Jöreskog, K. G. 1973b. Analysis of Covariance Structures. In P. R. Krishnaiah ed. *Multivariate analysis - III*. Academic Press, New York.

- Jöreskog, K. G., and A. S. Goldberger. 1972. Factor analysis by generalized least squares. *Psychometrika* 37:243-260.
- Kendall, M. G., and A. Stuart. 1979. *The Advanced Theory of Statistics*. Vol. 2. 4th ed. Griffin, London. 748 pp.
- Koopmans, T. C. 1937. *Linear Regression Analysis of Economic Time Series*. DeErven F. Bohn, Haarlem, Netherlands. 150 pp.
- Lawley, D. N. 1940. The estimation of factor loadings by the method of maximum likelihood. *Proc. Roy. Soc. Edinburgh A* 60:64-82.
- Lawley, D. N. 1941. Further investigations in factor estimation. *Proc. Roy. Soc. Edinburgh A* 61:176-185.
- Lawley, D. N. 1943. The application of the maximum likelihood method to factor analysis. *Brit. J. Psychol.* 33:172-175.
- Lawley, D. N. 1967. Some new results in maximum likelihood factor analysis. *Proc. Roy. Soc. Edinburgh A* 67:256-264.
- Lawley, D. N. 1976. The inversion of an augmented information matrix occurring in factor analysis. *Proc. Roy. Soc. Edinburgh A* 75:171-178.
- Lawley, D. N., and A. E. Maxwell. 1971. *Factor Analysis as a Statistical Method*. 2nd ed. American Elsevier, New York. 153 pp.
- Lindley, D. V. 1953. Estimation of a functional relationship. *Biometrika* 40:47-49.
- Madansky, W. 1959. The fitting of straight lines when both variables are subject to error. *J. Amer. Statist. Assoc.* 54:173-205.
- Mann, H. B., and A. Wald. 1943. On stochastic limit and order relationships. *Ann. Math. Statist.* 14:217-226.
- Moran, P. A. P. 1971. Estimating structural and functional relationships. *J. Multivariate Analysis* 1:232-255.
- Neyman, J., and E. L. Scott. 1948. Consistent estimates based on partially consistent observations. *Econometrica* 16:1-32.
- Nussbaum, M. 1978. Asymptotic efficiency of estimators of a multivariate linear functional relation. *Akademie der Wissenschaften der DDR. Berlin*.
- O'Neill, M., L. G. Sinclair, and F. J. Smith. 1969. Polynomial curve fitting when abscissas and ordinates are both subject to error. *Computer Journal* 12:52-56.

- Robertson, C. A. 1974. Large sample theory for the linear structural relation. *Biometrika* 61:353-359.
- Robinson, P. M. 1977. The estimation of a multivariate linear relation. *J. Multivariate Analysis* 7:409-423.
- Solari, M. E. 1969. The "maximum likelihood solution" to the problem of estimating a linear functional relationship. *J. Roy. Statist. Soc. B* 31:372-375.
- Sprent, P. 1966. A generalized least-squares approach to linear functional relationships. *J. Roy. Statist. Soc. B* 28:278-297.
- Tintner, G. 1945. A note on rank, multicollinearity, and multiple regression. *Ann. Math. Statist.* 16:304-308.
- Villegas, C. 1969. On the least squares estimation of non-linear relations. *Ann. Math. Statist.* 40:462-466.
- Villegas, C. 1982. Maximum likelihood and least squares estimation in linear and affine functional models. *Ann. Statist.* 10:256-265.
- Wald, A. 1940. Fitting of straight lines if both variables are subject to error. *Ann. Math. Statist.* 11:284-300.
- Whittle, P. 1952. On principal components and least square methods of factor analysis. *Skandinavisk Aktuarietidskrift* 35:223-239.
- Wolter, K. M. 1974. Estimators for a nonlinear functional relationship. Ph.D. Thesis. Iowa State University. Ames, Iowa. 276 pp.
- Wolter, K. M., and W. A. Fuller. 1982a. Estimation of the quadratic errors-in-variables model. *Biometrika* 69:175-182.
- Wolter, K. M., and W. A. Fuller. 1982b. Estimation of nonlinear errors-in-variables models. *Ann. Statist.* 10:539-548.

## VI. ACKNOWLEDGEMENTS

I wish to express my sincere thanks to Professor Wayne A. Fuller. His expert guidance has provided me with a stimulating and enjoyable experience throughout four years of graduate work.

I am grateful to the members of my graduate committee: Professors David Harville, Malay Ghosh, Glen Meeden, and Fred Wright.

Thanks go to Julie Honeick for her excellent typing of this dissertation. Thanks also go to Jane Stowe, who has provided skilled typing of several drafts of the dissertation.

Helen Nelson has frequently worked over time to print several versions of the dissertation. I wish to thank her for her efficient secretarial work as well as her friendship.

Professor Gary Koch and Ms. Rhonda Robinson of University of North Carolina introduced me to the topic of Chapter V. I am grateful for their suggestion.

I am indebted to the Survey Section of the Statistical Laboratory for providing its word processor for the production of the final version of this dissertation.