
Herman J. Bierens∗
Pennsylvania State University, USA
June 4, 2018

Abstract
In this paper we propose a new estimation approach for the static random effects linear panel data model with fixed time dimension, with instruments based on the integrated conditional moment (ICM) principle, which is in essence an instrumental variable approach with infinitely many instruments. The advantages of this new approach are that it allows to estimate the parameters of time-invariant covariates directly and yields a consistent ICM test of the null hypothesis that the model is a proper multivariate conditional expectation model, including the case where the parameters are time-varying. Under this null hypothesis the parameter estimates are strongly consistent, unbiased and asymptotically normal. Moreover, it will be shown that under some regularity conditions these asymptotic normality results also hold for the fixed effects case. The ICM estimation and testing procedures are checked by a limited Monte Carlo analysis.

∗Professor Emeritus of Economics. Please address comments by email only to hbierens@psu.edu, as I no longer have an office at Penn State.
1 Introduction

In this paper we propose a new estimation approach for the static random effects linear panel data model with fixed time dimension $T$, with instruments based on the integrated conditional moment (ICM) principle, which is in essence an instrumental variable approach with infinite many instruments. The advantages of this new approach are that it allows to estimate the parameters of time-invariant covariates directly and yields a consistent ICM test of the null hypothesis that the model, in the form of a system of linear regressions, is a proper multivariate conditional expectation model, including the case where the parameters are time-varying. Under this null hypothesis the parameter estimates are strongly consistent, unbiased and asymptotically normal. Moreover, it will be shown that under some regularity conditions these asymptotic normality results also hold for the fixed effects case.

The ICM estimation approach involved is related to Carrasco and Florens (2000). These authors design an efficient GMM estimator in the case of infinitely many real-valued moment conditions. However, their approach requires to compute eigenvalues and corresponding eigenfunctions of the covariance kernel of the moment functions.

The consistent ICM tests involved have a non-pivotal asymptotic null distribution, as is the case for the ICM tests considered in Bierens (1982) and Bierens and Ploberger (1997), and therefore their critical values have to be approximated by a bootstrap method. The accuracy of the bootstrap approach will be checked by a Monte Carlo analysis.

2 The static random effects linear panel data model

The standard static random effects linear panel data model takes the form

$$Y_{i,t} = \alpha_0 + \beta_0'X_{i,t} + V_i + U_{i,t}, \ i = 1, 2, ..., N, \ t = 1, 2, ..., T, \quad (1)$$

where for each $i$ and $t$,

- $Y_{i,t}$ is the dependent variable,
- $X_{i,t} \in \mathbb{R}^p$ is a vector of stochastic covariates,
• $V_i$ is the unobserved random effect, which does not depend on $t$, and
• $U_{i,t}$ is an error term.

The standard approaches to estimate $\beta_0$ is either to take all variables in first differences, or in difference of their time averages, and then use GMM, or assume independence and normality of $V_i$ and the $U_{i,t}$’s and then conduct maximum likelihood (ML) estimation. However, these differencing methods not only wipe out the constant term and the random effects $V_i$, but also any component of $X_{i,t}$ that does not depend on $t$, such as race and gender indicators. Moreover, the ML approach requires too restrictive conditions, such as independence and normality of $V_i$ and the $U_{i,t}$’s with zero expectations and constant variances. See for example Baltagi (2005, sections 2.3 and 2.4).

In this paper we propose a different approach by taking first all the variables involved in deviation of their cross-section means, and then using an integrated conditional moment (ICM) estimation approach to estimate $\beta_0$ consistently and asymptotic normally, regardless whether some components of $X_{i,t}$ are time invariant or not. The main reason for taking all the variables involved in deviation of their cross-section means is merely for the ease in proving the consistency and asymptotic normality of the proposed ICM estimator of $\beta_0$.

Moreover, this approach also allows the constant $\alpha_0$ in model (1) to be time-dependent, i.e.,

$$Y_{i,t} = \alpha_{0,t} + \beta_0 X_{i,t} + V_i + U_{i,t}, \quad i = 1, 2, ..., N, \quad t = 1, 2, ..., T,$$

(2)

because the $\alpha_{0,t}$’s are wiped out after taking all the variables in deviation of their cross-section means. Consequently, we may without loss of generality assume that

$$E[V_i] = 0 \text{ and } E[U_{i,t}] = 0 \text{ for all } i \text{ and } t,$$

(3)

while allowing for heterogeneity of the errors $U_{i,t}$’s across time, provided of course that these expectations are well defined.

3 Assumptions

As to the data, it will be assumed that:
Assumption 1. Jointly in all $t$, all the random variables in model (2) are i.i.d. across individuals $i = 1, 2, ..., N$ and have finite second moments. In particular, for each $t$, $X_{i,t}$ is i.i.d. as $X_t$, $U_{i,t}$ is i.i.d. as $U_t$ and $V_i$ is i.i.d. as $V$. Moreover, for each $t$, $\text{Var}(X_t)$ is nonsingular.\footnote{As otherwise $\beta_0$ is not identified.}

As to further conditions on $V_i$, it is not unreasonable to assume that the random effect $V_i$ represents the net effect of unobserved time-invariant covariates, represented by a single random variable $W_i$, say, ”net” in the sense that the effect of the $X_{i,t}$’s on $W_i$ is eliminated in some sense. In particular, interpret $V_i$ as

$$V_i = W_i - E[W_i|X_{i,1}, X_{i,2}, ..., X_{i,T}],$$

so that

$$E[V_i|X_{i,1}, X_{i,2}, ..., X_{i,T}] = 0 \ a.s.$$ \hspace{1cm} (5)

Note that the latter is much weaker a condition than in the ML case, which usually requires that $V_i$ is independent of $X_{i,1}, X_{i,2}, ..., X_{i,T}$ and is $\mathcal{N}[0, \sigma^2_V]$ distributed. See for example Baltagi (2005, sections 2.3 and 2.4).

Therefore, it will be assumed that:

Assumption 2. For each $i$, $E[V_i|G_{i,T}] = 0$ a.s., where $G_{i,T} = \sigma(\{X_{i,t}\}_{t=1}^T)$ is the $\sigma$-algebra generated by the sequence $\{X_{i,t}\}_{t=1}^T$.

To motivate conditional expectation conditions on the $U_{i,t}$’s, write model (2) in the form of a system regression equations:

$$Y^T_i = \begin{pmatrix} Y_{i,1} \\ \vdots \\ Y_{i,T} \end{pmatrix} = \begin{pmatrix} \alpha_{0,1} \\ \vdots \\ \alpha_{0,T} \end{pmatrix} + \begin{pmatrix} X'_{i,1} \\ \vdots \\ X'_{i,T} \end{pmatrix} \beta_0 + V_i \begin{pmatrix} 1 \\ \vdots \\ 1 \end{pmatrix} + U_i \begin{pmatrix} U_{i,1} \\ \vdots \\ U_{i,T} \end{pmatrix} \hspace{1cm} (6)$$

say, so that given Assumption 2, $E[Y^T_i|X^T_i] = \alpha^T_0 + X^T_i \beta_0 + E[U^T_i|X^T_i]$. Therefore, for $Y^T_i = \alpha^T_0 + X^T_i \beta_0 + V_i e^T + U^T_i$ to be correctly specified as
a multivariate conditional expectation model one should require, next to Assumption 2, that \( E[U_T^i|X_T^i] = 0 \) a.s. as well, i.e.,

**Assumption 3.** For each \( t \) and \( i \), \( E[U_{i,t}|G_{i,T}] = 0 \) a.s., where \( G_{i,T} \) is defined in Assumption 2.

**Remark 1.** As said before, this condition leaves the possibility open that the \( U_{i,t} \)'s are serially correlated. Moreover, \( V_i \) may be correlated with the \( U_{i,t} \)'s as well.

**Remark 2.** If the random effects \( V_i \) were observable then a more appropriate condition on the \( U_{i,t} \)'s would be

\[
E[U_{i,t}|F_{i,T}] = 0 \text{ a.s. for } t = 1, 2, ..., T,
\]

where \( F_{i,T} \) is the \( \sigma \)-algebra generated by \( (V_i, X_{i,1}', X_{i,2}', ..., X_{i,T}')' \). In this case \( E[V_i U_{i,t}|F_{i,T}] = V_i E[U_{i,t}|F_{i,T}] = 0 \) a.s., hence \( E[V_i U_{i,t}] = E(E[V_i U_{i,t}|F_{i,T}]) = 0 \), so that \( V_i \) is uncorrelated with the \( U_{i,t} \)'s, and

\[
E[U_{i,t}|G_{i,T}] = E(E[U_{i,t}|F_{i,T}]|G_{i,T}) = 0 \text{ a.s.}
\]

However, in our proposed ICM estimation approach uncorrelatedness of \( V_i \) and the \( U_{i,t} \)'s is of no particular use, although it will not harm either. Therefore, we will stick with Assumption 3.

**Remark 3.** Assumptions 1-3 are similar to the conditions in Arrellano and Honore (2001) for the static random effect panel data model, except that instead of the conditional expectation conditions in Assumptions 2 and 3 these authors formulate their conditions in terms of linear projections.

Given Assumptions 2 and 3, we may now treat \( V_i + U_{i,t} = R_{i,t} \) as the error term in model (2):

\[
Y_{i,t} = \alpha_{0,t} + \beta_0'X_{i,t} + R_{i,t}, \quad i = 1, 2, ..., N, \quad t = 1, 2, ..., T,
\]

where by the combined Assumptions 2 and 3, this model is correctly specified as a conditional expectation model under the null hypothesis that for \( t = 1, 2, ..., T \), with \( R_{i,t} = V_i + U_{i,t} \),

\[
H_0 : \Pr(E[R_{i,t}|G_{i,T}] = 0) = 1.
\]
In other words, $H_0$ amounts to the hypothesis that Assumption 2 and 3 hold.

As said before, our approach allows for time invariant covariates. To make this explicit, let

$$X_{i,t} = \left( \begin{array}{c} Z_i^* \\ X_{i,t}^* \end{array} \right),$$

where $Z_i^* \in \mathbb{R}^q$ is the vector of time-invariant covariates, and $X_{i,t}^* \in \mathbb{R}^{p-q}$ is the vector of time-varying covariates. Then $\mathcal{G}_{i,T}$ in Assumptions 2 and 3 is the $\sigma$-algebra generated by the random vector

$$Z_i = (Z_i^*, X_{i,1}^*, X_{i,2}^*, \ldots, X_{i,T}^*)' \in \mathbb{R}^\ell, \ \ell = q + (p-q)T.$$ (9)

The ICM estimation procedure proposed in the next sections employs infinitely many weight functions $w(z, \xi)$, indexed by $\xi \in \Xi \subset \mathbb{R}^{\ell}$, where $w(.)$ satisfies the following conditions.

**Assumption 4.** The function $w(u)$ is a real or complex-valued analytical function on $\mathbb{R}$ with higher-order derivatives $w_n(u) = (d/du)^n w(u)$ for $n \in \mathbb{N}$, $w_0(u) = w(u)$, such that the set $\mathcal{N}_0 = \{n \in \mathbb{N} : w_n(0) = 0\}$ is either finite or empty.$^2$

Then the following result holds.

**Lemma 1.** Let $Z \in \mathbb{R}^\ell$ be a bounded random vector and let $U$ be a random variable or vector satisfying $E[||U||] < \infty$ and $\Pr[E[U|Z] = 0] < 1$. Then under Assumption 4 with $\mathcal{N}_0 = \emptyset$, the set

$$S = \{\xi \in \mathbb{R}^\ell : E[U \cdot w(\xi'Z) = 0]\}$$

has Lebesgue measure zero and is nowhere dense in $\mathbb{R}^\ell$. If $\mathcal{N}_0 \neq \emptyset$ but finite this result carries over, provided that one component of $Z$, say $Z_1$, satisfies $\Pr[Z_1 \neq 0] = 1$.

**Proof.** See Theorem 2.2 in Bierens (2017, p.309). ■

Examples of suitable functions $w(u)$ are the complex $\exp(.)$ function $w(u) = \exp(i.u) = \cos(u) + i \sin(u)$, where $i = \sqrt{-1}$, used in Bierens (1982,

---

$^2$Note that Assumption 4 encompasses Assumptions 2.3-2.4 and 2.6 in Bierens (2017, pp. 307-309).
the real \( \exp(.) \) function \( w(u) = \exp(u) \) used in Bierens (1990), and \( w(u) = \cos(u) + \sin(u) \), for example, among many other options for \( w(u) \). In each of these three cases the set \( N_0 \) in Assumption 4 is empty.

On the other hand, the functions \( \cos(u) \) and \( \sin(u) \) separately are not suitable because in these cases the set \( N_0 \) in Assumption 4 is infinite, although these functions are non-polynomial and analytical. This is the case considered by Stinchcombe and White (1998). They only require that \( w(u) \) is non-polynomial and analytical. In their case the result in Lemma 1 reads: The set \( \{ (\xi, \xi')' \in \mathbb{R}^{\ell+1} : E[U.w(\xi + \xi'Z) = 0] \} \) has Lebesgue measure zero and is nowhere dense in \( \mathbb{R}^{\ell+1} \), provided that in this case \( Z \) is bounded as well. See Theorem 2.3 in Stinchcombe and White (1998) and Theorem 2.3 in Bierens (2017, p. 317).3

If \( Z \) in Lemma 1 is not bounded we may without loss of generality replace \( Z \) with \( \Phi(Z) \), where \( \Phi \) is a bounded one-to-one mapping from \( \mathbb{R}^{\ell} \) into \( \mathbb{R}^{\ell} \) with Borel measurable inverse, because then \( Z \) and \( \Phi(Z) \) generate the same \( \sigma \)-algebra, so that \( E[U|Z] = E[U|\Phi(Z)] \), hence under the conditions of Lemma 1, the set \( \{ \xi \in \mathbb{R}^\ell : E[U.w(\xi'\Phi(Z)) = 0] \} \) has Lebesgue measure zero and is nowhere dense in \( \mathbb{R}^{\ell} \).

4 Identification of \( \beta_0 \) by infinitely many moment conditions

In the present case the vector of conditional variables for each cross-section \( i \) is the random vector \( Z_i \in \mathbb{R}^{\ell} \) in (9), with \( \ell = q + (p - q)T \). Suppose for the time being that the \( Z_i \)'s are already bounded.

For any \( \beta \in \mathbb{R}^p \), and each \( i \) and \( t \), denote

\[
\psi_{i,t}(\beta) = Y_{i,t} - E[Y_{1,t}] - \beta'(X_{i,t} - E[X_{1,t}]) = R_{i,t} - (\beta - \beta_0)'(X_{i,t} - E[X_{1,t}])
\]

Moreover let \( f(\xi) \) a continuous density with support \( \Xi \subset \mathbb{R}^{\ell} \). Then given that \( \Xi \) has positive Lebesgue measure, it follows straightforwardly from Lemma 1 and the observation that

\[
\Pr(E[X_{1,t} - E[X_{1,t}]|Z_1] = 0) < 1
\]

\[3\]The proof of the latter theorem is much easier and more transparent than the proof in Stinchcombe and White (1998).
that under Assumptions 1-4,

\[ Q(\beta) = \sum_{t=1}^{T} \int_{\Xi} \left| E[\psi_{i,t}(\beta)w(\xi'Z_{i})] \right|^2 f(\xi) d\xi \]

\[ = \sum_{t=1}^{T} \int_{\Xi} \left| (\beta - \beta_0)' E [(X_{1,t} - E[X_{1,t}]) w(\xi'Z_{1})] \right|^2 f(\xi) d\xi \]

\[ = (\beta - \beta_0)' \left( \sum_{t=1}^{T} \Sigma_t \right) (\beta - \beta_0) \]

\[ = 0 \quad \text{for} \quad \beta = \beta_0, \]

\[ > 0 \quad \text{for} \quad \beta \neq \beta_0, \]

provided that \( \sum_{t=1}^{T} \Sigma_t \) is nonsingular, where in the real \( w \) case,

\[ \Sigma_t = \int_{\Xi} E[(X_{1,t} - E[X_{1,t}])w(\xi'Z_{1})] \]

\[ \times E[(X_{1,t} - E[X_{1,t}])'w(\xi'Z_{1})] f(\xi) d\xi, \] (10)

and in the complex \( w \) case,

\[ \Sigma_t = \int_{\Xi} E[(X_{1,t} - E[X_{1,t}])w(\xi'Z_{1})] \]

\[ \times E[(X_{1,t} - E[X_{1,t}])'\overline{w(\xi'Z_{1})}] f(\xi) d\xi, \] (11)

where the bar denotes the complex conjugate. Note that in the latter case \( \Sigma_t \) is real-valued as well.

In either case \( \Sigma_t \) is nonsingular, due to the last condition in Assumption 1. More precisely, we have the following result.

**Lemma 2.** Suppose that the vectors \( Z_i \in \mathbb{R}^d \) of conditioning variables has bounded support and satisfy \( E[X_{i,t}|Z_i] = X_{i,t} \) a.s.\(^4\) for \( t = 1, 2, \ldots, T \). Let \( f(\xi) \) be a continuous density with support \( \Xi \subset \mathbb{R}^d \), where \( \Xi \) has positive Lebesgue measure. Then under Assumptions 1 and 4 the matrices \( \Sigma_t \) in (11) and (12) are nonsingular.

\(^4\)The latter is the case if \( X_{i,t} \) is measurable with respect to the \( \sigma \)-algebra generated by \( Z_i \). C.f. (9).
Proof. Suppose that there exists a nonzero vector $\gamma \in \mathbb{R}^p$ such that $\gamma' \Sigma_t \gamma = 0$. Then in the case (12), $\int_\Xi |E[\gamma'(X_{1,t} - E[X_{1,t}])w(\xi' Z_1)]|^2 f(\xi)d\xi = 0$, which implies that $E[\gamma'(X_{1,t} - E[X_{1,t}])w(\xi' Z_1)] = 0$ a.e. on $\Xi$. In its turn the latter implies, by Lemma 1, that $E[\gamma'(X_{1,t} - E[X_{1,t}])|Z_1] = 0$ a.s., and since $\gamma'(X_{1,t} - E[X_{1,t}])$ is measurable with respect to the $\sigma$-algebra generated by $Z_1$ it follows now that $\gamma'(X_{1,t} - E[X_{1,t}]) = E[\gamma'(X_{1,t} - E[X_{1,t}])|Z_1] = 0$ a.s. But then $\gamma'(X_{1,t} - E[X_{1,t}])'(X_{1,t} - E[X_{1,t}]) = 0$ a.s., and taking the expectation yields $\gamma' \text{Var}(X_{1,t})\gamma = 0$. However, the latter is excluded by the last condition in Assumption 1.

For notational convenience, we will in the sequel denote

$$w_i(\xi) = w(\xi' Z_i),$$

where $Z_i$ satisfies the conditions of Lemma 2.

Now suppose that $H_0$ is false, i.e.,

$$H_1 : \Pr( E[R_{i,t}|G_{i,T}] = 0) < 1,$$

which is equivalent to the hypothesis that Assumptions 2 and/or 3 do not hold. Then under Assumption 4, $E[R_{i,t}w_i(\xi)] \neq 0$ a.e. on $\Xi$, whereas under $H_0$, $E[R_{i,t}w_i(\xi)] \equiv 0$ on $\Xi$. These results will be used to devise a consistent test of the validity of the null hypothesis (8).

Throughout we derive our results for the case that $w_i(\xi)$ is complex-valued, in particular the case where $w(u) = \exp(iu)$. The real-valued cases then follow trivially by replacing the complex conjugate $\bar{\omega}_i(\xi)$ with $\omega_i(\xi)$ itself. Moreover, throughout it will be assumed that:

**Assumption 5.** The integration range $\Xi$ is compact, with positive Lebesgue measure.\(^5\)

The compactness of $\Xi$ is not strictly necessary, but this condition enables us to apply Jennrich's (1969) uniform strong law of large numbers [USLLN hereafter] in deriving the strong consistency of the ICM estimator of $\beta_0$ below.

### 5  ICM estimation: The single $t$ case

Denote

\(^5\)The latter condition excludes the case that $\Xi$ is a singleton.
\[ Y_{N,t} = N^{-1} \sum_{i=1}^{N} Y_{i,t}, \quad \tilde{Y}_{i,t} = Y_{i,t} - \bar{Y}_{N,t}, \]
\[ X_{N,t} = N^{-1} \sum_{i=1}^{N} X_{i,t}, \quad \tilde{X}_{i,t} = X_{i,t} - \bar{X}_{N,t}, \]
\[ R_{N,t} = N^{-1} \sum_{i=1}^{N} R_{i,t}, \quad \tilde{R}_{i,t} = R_{i,t} - \bar{R}_{N,t}, \]

so that model (7) reads
\[ \tilde{Y}_{i,t} = \beta_0' \tilde{X}_{i,t} + \tilde{R}_{i,t}, \quad i = 1, 2, ..., N, \quad t = 1, 2, ..., T. \tag{15} \]

For any \( \beta \in \mathbb{R}^p \), and each \( i \) and \( t \), denote
\[ \tilde{\psi}_{i,t}(\beta) = \tilde{Y}_{i,t} - \beta' \tilde{X}_{i,t} = \tilde{R}_{i,t} - (\beta - \beta_0)' \tilde{X}_{i,t} \]
and for each \( t \) consider the ICM objective function
\[ \hat{Q}_{t,N}(\beta) = \int_{\Xi} \left| \frac{1}{N} \sum_{i=1}^{N} \tilde{\psi}_{i,t}(\beta) \omega_i(\xi) \right|^2 f(\xi) d\xi, \tag{16} \]

where \( f(\xi) \) and \( \Xi \) satisfy the conditions in Lemma 2 and Assumption 5. Note that
\[ \hat{Q}_{t,N}(\beta) = \int_{\Xi} \left( \frac{1}{N} \sum_{i=1}^{N} \left( \bar{R}_{i,t} - (\beta - \beta_0)' \bar{X}_{i,t} \right) \omega_i(\xi) \right) \times \left( \frac{1}{N} \sum_{j=1}^{N} \left( \bar{R}_{j,t} - (\beta - \beta_0)' \bar{X}_{j,t} \right) \omega_j(\xi) \right) f(\xi) d\xi \]
\[ = \frac{1}{N^2} \sum_{i=1}^{N} \sum_{j=1}^{N} \bar{X}_{i,t} \tilde{X}_{j,t}^t \int_{\Xi} \omega_i(\xi) \omega_j(\xi) f(\xi) d\xi \]
\[- (\beta - \beta_0)' \frac{1}{N^2} \sum_{i=1}^{N} \sum_{j=1}^{N} \bar{X}_{i,t} \bar{R}_{j,t} \int_{\Xi} \omega_i(\xi) \omega_j(\xi) f(\xi) d\xi \]
\[- (\beta - \beta_0)' \frac{1}{N^2} \sum_{j=1}^{N} \sum_{i=1}^{N} \bar{X}_{i,t} \bar{R}_{j,t} \int_{\Xi} \omega_i(\xi) \omega_j(\xi) f(\xi) d\xi \]
\[+ \frac{1}{N^2} \sum_{i=1}^{N} \sum_{j=1}^{N} \bar{R}_{i,t} \bar{R}_{j,t} \int_{\Xi} \omega_i(\xi) \omega_j(\xi) f(\xi) d\xi \]
\[
\frac{1}{N^2} \sum_{i=1}^{N} \sum_{j=1}^{N} \bar{X}_{i,t} \bar{X}_{j,t} \int_{\Xi} \omega_i(\xi) \omega_j(\xi) f(\xi) d\xi - 2(\beta - \beta_0)' \frac{1}{N^2} \sum_{i=1}^{N} \sum_{j=1}^{N} \bar{X}_{i,t} \bar{R}_{j,t} \int_{\Xi} \text{Re}[\omega_i(\xi) \omega_j(\xi)] f(\xi) d\xi + \frac{1}{N^2} \sum_{i=1}^{N} \sum_{j=1}^{N} \bar{R}_{i,t} \bar{R}_{j,t} \int_{\Xi} \omega_i(\xi) \omega_j(\xi) f(\xi) d\xi,
\]

where the last equality follows from the trivial equalities \((a + i.b)(c - i.d) = (a.c + b.d) + i.(b.c - a.d)\) and \((a - i.b)(c + i.d) = (a.c + b.d) - i.(b.c - a.d)\).

Hence

\[
\widehat{Q}_{t,N}(\beta) = (\beta - \beta_0)' \widehat{\Sigma}_{t,N}(\beta - \beta_0) - 2(\beta - \beta_0)' \widehat{\eta}_{t,N} + \widehat{\rho}_{t,N},
\]  

where

\[
\widehat{\Sigma}_{t,N} = \int_{\Xi} \left( \frac{1}{N} \sum_{i=1}^{N} \bar{X}_{i,t} \omega_i(\xi) \right) \left( \frac{1}{N} \sum_{j=1}^{N} \bar{X}_{j,t} \omega_j(\xi) \right) f(\xi) d\xi = \frac{1}{N^2} \sum_{i=1}^{N} \sum_{j=1}^{N} \bar{X}_{i,t} \bar{X}_{j,t} \int_{\Xi} \omega_i(\xi) \omega_j(\xi) f(\xi) d\xi;
\]

\[
\widehat{\eta}_{t,N} = \int_{\Xi} \text{Re} \left[ \left( \frac{1}{N} \sum_{i=1}^{N} \bar{X}_{i,t} \omega_i(\xi) \right) \left( \frac{1}{N} \sum_{j=1}^{N} \bar{R}_{j,t} \omega_j(\xi) \right) \right] f(\xi) d\xi = \frac{1}{N^2} \sum_{i=1}^{N} \sum_{j=1}^{N} \bar{X}_{i,t} \bar{R}_{j,t} \int_{\Xi} \text{Re}[\omega_i(\xi) \omega_j(\xi)] f(\xi) d\xi;
\]

\[
\widehat{\rho}_{t,N} = \int_{\Xi} \left| \frac{1}{N} \sum_{i=1}^{N} \bar{R}_{i,t} \omega_i(\xi) \right|^2 f(\xi) d\xi,
\]

Thus,

\[
\beta_{t,N} = \arg \min_{\beta \in \mathbb{R}^p} \widehat{Q}_{t,N}(\beta) = \beta_0 + \widehat{\Sigma}_{t,N}^{-1} \widehat{\eta}_{t,N},
\]

provided that \(\widehat{\Sigma}_{t,N}\) is nonsingular.

Of course, this estimator is infeasible because \(\beta_0\) is unknown. To derive a feasible expression for \(\beta_{t,N}\), expand \(\widehat{Q}_{t,N}(\beta)\) in a slightly different way than (17), as

\[
\widehat{Q}_{t,N}(\beta) = \beta' \widehat{\Sigma}_{t,N} \beta - 2\beta' \widehat{\mu}_{t,N} + \widehat{\gamma}_{t,N}
\]
where $\tilde{\Sigma}_{t,N}$ is the same as before, and

$$
\tilde{\mu}_{t,N} = \frac{1}{N^2} \sum_{i=1}^{N} \sum_{j=1}^{N} \tilde{X}_{i,t} \bar{Y}_{j,t} \int_{\Xi} \text{Re}[\omega_i(\xi) \bar{\omega}_j(\xi)] f(\xi) d\xi,
$$

(22)

$$
\tilde{\gamma}_{t,N} = \frac{1}{N^2} \sum_{i=1}^{N} \sum_{j=1}^{N} \tilde{Y}_{i,t} \bar{Y}_{j,t} \int_{\Xi} \omega_i(\xi) \bar{\omega}_j(\xi) f(\xi) d\xi
= \int_{\Xi} \left| \frac{1}{N} \sum_{i=1}^{N} \tilde{Y}_{i,t} \omega_i(\xi) \right|^2 f(\xi) d\xi,
$$

so that

$$
\tilde{\beta}_{t,N} = \arg\min_{\beta \in \mathbb{R}^p} \tilde{Q}_{t,N}(\beta) = \tilde{\Sigma}_{t,N}^{-1} \tilde{\mu}_{t,N}.
$$

### 5.1 Strong consistency and unbiasedness

#### 5.1.1 Strong consistency

We will shown now that $\tilde{\Sigma}_{t,N} \xrightarrow{a.s.} \Sigma_t$, where $\Sigma_t$ is the same as in (12), and $\tilde{\eta}_{t,N} \xrightarrow{a.s.} 0$, hence $\tilde{\beta}_{t,N} \xrightarrow{a.s.} \beta_0$. First, it is a standard exercise to verify that under Assumptions 1-4, and for each $t$ and $\xi \in \Xi$,

$$
\frac{1}{N} \sum_{i=1}^{N} \tilde{X}_{i,t} \omega_i(\xi) = \frac{1}{N} \sum_{i=1}^{N} (X_{i,t} - E[X_{1,t}]) \omega_i(\xi)
- (X_{N,t} - E[X_{1,t}]) \frac{1}{N} \sum_{i=1}^{N} \omega_i(\xi)
$$

(23)

$$
\frac{1}{N} \sum_{i=1}^{N} \tilde{R}_{i,t} \omega_i(\xi) = \frac{1}{N} \sum_{i=1}^{N} R_{i,t} \omega_i(\xi) - R_{N,t} \frac{1}{N} \sum_{i=1}^{N} \omega_i(\xi).
$$

(24)

Moreover, it follows from the USLLN that under Assumptions 1-5,

$$
\sup_{\xi \in \Xi} \left\| \frac{1}{N} \sum_{i=1}^{N} (X_{i,t} - E[X_{i,t}]) \text{Re}[\omega_i(\xi)] - E[(X_{1,t} - E[X_{1,t}]) \text{Re}[\omega_1(\xi)] \right\| \xrightarrow{a.s.} 0,
$$

$$
\sup_{\xi \in \Xi} \left\| \frac{1}{N} \sum_{i=1}^{N} (X_{i,t} - E[X_{i,t}]) \text{Im}[\omega_i(\xi)] - E[(X_{1,t} - E[X_{1,t}]) \text{Im}[\omega_1(\xi)] \right\| \xrightarrow{a.s.} 0,
$$

12
\[
\sup_{\xi \in \Xi} \left| \frac{1}{N} \sum_{i=1}^{N} \omega_i(\xi) - E[\omega_1(\xi)] \right| \overset{a.s.}{\rightarrow} 0,
\]
\[
\sup_{\xi \in \Xi} \left| \frac{1}{N} \sum_{i=1}^{N} R_{i,t} \omega_i(\xi) \right| \overset{a.s.}{\rightarrow} 0.
\]
whereas by Kolmogorov’s strong law of large numbers, \( \overline{X}_{N,t} \overset{a.s.}{\rightarrow} E[X_{1,t}] \), \( \overline{R}_{N,t} \overset{a.s.}{\rightarrow} 0 \). Hence,
\[
\hat{\eta}_{t,N} \overset{a.s.}{\rightarrow} 0, \text{ and}
\hat{\Sigma}_{t,N} \overset{a.s.}{\rightarrow} \Sigma_t, \text{ where}
\]
\[
\Sigma_t = \int_{\Xi} E[(X_{1,t} - E[X_{1,t}]) \omega_1(\xi)] \cdot E[(X_{1,t} - E[X_{1,t}]) \omega_1(\xi)] \times f(\xi) d\xi.
\]
Note that \( \Sigma_t \) is the same as in (12).

It follows now straightforwardly that \( \hat{\beta}_{t,N} \overset{a.s.}{\rightarrow} \beta_0 \), using the fact that by Lemma 2, \( \Sigma_t \) is nonsingular, so that \( \hat{\Sigma}_{t,N}^{-1} \overset{a.s.}{\rightarrow} \Sigma_t^{-1} \).

### 5.1.2 Unbiasedness

But we have more, namely, \( E[\hat{\beta}_{t,N}] = \beta_0 \) for sufficiently large \( N \). To see this, note that
\[
E \left[ \hat{\Sigma}_{t,N}^{-1} \hat{\eta}_{t,N} | X_{1,t}, X_{2,t}, ..., X_{N,t} \right] = \hat{\Sigma}_{t,N}^{-1} \left( E \left[ \hat{\eta}_{t,N} | X_{1,t}, X_{2,t}, ..., X_{N,t} \right] \right),
\]
provided that \( \hat{\Sigma}_{t,N} \) is nonsingular, and
\[
\hat{\eta}_{t,N} = \frac{1}{N^2} \sum_{i=1}^{N} \sum_{j=1}^{N} \tilde{X}_{i,t} \int_{\Xi} \text{Re} \left[ \omega_i(\xi) \overline{\omega_j(\xi)} \right] f(\xi) d\xi \tilde{R}_{i,t}
\]
\[
= \frac{1}{N^2} \sum_{i=1}^{N} \sum_{j=1}^{N} \tilde{X}_{i,t} \int_{\Xi} \text{Re} \left[ \omega_i(\xi) \overline{\omega_j(\xi)} \right] f(\xi) d\xi R_{i,t}
\]
\[
- \frac{1}{N^2} \sum_{i=1}^{N} \sum_{j=1}^{N} \tilde{X}_{i,t} \int_{\Xi} \text{Re} \left[ \omega_i(\xi) \overline{\omega_j(\xi)} \right] f(\xi) d\xi \frac{1}{N} \sum_{m=1}^{N} R_{m,t}
\]
so that
\[
E \left[ \tilde{\eta}_{t,N} | X_{1,t}, X_{2,t}, \ldots, X_{N,t} \right]
= \frac{1}{N^2} \sum_{i=1}^{N} \sum_{j=1}^{N} \tilde{X}_{i,t} \int_{\mathbb{R}} \text{Re}[\omega_i(\xi)\omega_j(\xi)] f(\xi) d\xi 
\times \left( E \left[ R_{i,t} | X_{1,t}, X_{2,t}, \ldots, X_{N,t} \right] - \frac{1}{N} \sum_{m=1}^{N} E \left[ R_{m,t} | X_{1,t}, X_{2,t}, \ldots, X_{N,t} \right] \right)
= 0 
\]
because by cross-sectional independence and Assumptions 2 and 3,
\[
E \left[ R_{i,t} | X_{1,t}, X_{2,t}, \ldots, X_{N,t} \right] = E \left[ R_{i,t} | X_{i,t} \right] = E \left( E \left[ R_{i,t} | G_{i,T} \right] | X_{i,t} \right) = 0.
\]
Thus,
\[
E [\tilde{\beta}_{t,N}] - \beta_0 = E \left[ \tilde{\Sigma}_{t,N}^{-1} \tilde{\eta}_{t,N} \right] = 0,
\]
for \(N\) so large that \(\tilde{\Sigma}_{t,N}\) is nonsingular.

Since \(\tilde{\Sigma}_{t,N} \xrightarrow{a.s.} \Sigma_t\) it follows by the continuity of the determinant of a square matrix in the elements of this matrix that \(\text{det}(\tilde{\Sigma}_{t,N}) \xrightarrow{a.s.} \text{det}(\Sigma_t)\), which is equivalent\(^6\) to the statement that for all \(\varepsilon > 0\),
\[
\lim_{K \to \infty} \Pr \left[ \sup_{N \geq K} \left| \text{det}(\tilde{\Sigma}_{t,N}) - \text{det}(\Sigma_t) \right| \leq \varepsilon \right] = 1.
\]
In its turn this implies that for \(\varepsilon < \text{det}(\Sigma_t)\),
\[
\lim_{K \to \infty} \Pr \left( \inf_{N \geq K} \text{det}(\tilde{\Sigma}_{t,N}) \geq \text{det}(\Sigma_t) - \varepsilon \right) = 1
\]
and thus
\[
\lim_{K \to \infty} \Pr \left( \inf_{N \geq K} \text{det}(\tilde{\Sigma}_{t,N}) > 0 \right) = 1.
\]
But \(\inf_{N \geq K} \text{det}(\tilde{\Sigma}_{t,N}) > 0\) implies \(E[\tilde{\beta}_{t,N} | X_{1,t}, X_{2,t}, \ldots, X_{N,t}] = \beta_0\) for all \(N \geq K\), hence
\[
\lim_{K \to \infty} \Pr \left( E[\tilde{\beta}_{t,N} | X_{1,t}, X_{2,t}, \ldots, X_{N,t}] = \beta_0 \right. \text{ for all } N \geq K = 1.
\]
Moreover, \( E[\widehat{\beta}_{t,N}|X_{1,t}, X_{2,t}, \ldots, X_{N,t}] = \beta_0 \) for all \( N \geq K \) implies that \( E[\widehat{\beta}_{t,N}] = \beta_0 \) for all \( N \geq K \), where the latter event has either "probability" 1 or 0. Thus

\[
1 = \lim_{K \to \infty} \Pr \left( E[\widehat{\beta}_{t,N}|X_{1,t}, X_{2,t}, \ldots, X_{N,t}] = \beta_0 \text{ for all } N \geq K \right)
\leq \lim_{K \to \infty} I \left( E[\widehat{\beta}_{t,N}] = \beta_0 \text{ for all } N \geq K \right),
\]

where \( I(.) \) is the indicator function. Consequently, there exists a fixed \( K \in \mathbb{N} \) such that \( E[\widehat{\beta}_{t,N}] = \beta_0 \) for all \( N \geq K \).

Summarizing, it has been shown that

**Lemma 3.** Under Assumptions 1-5, \( \widehat{\beta}_{t,N} \stackrel{\text{a.s.}}{\to} \beta_0 \) and for sufficiently large \( N \), \( \widehat{\beta}_{t,N} \) is unbiased, in the sense that there exists a fixed \( K \in \mathbb{N} \) such that \( E[\widehat{\beta}_{t,N}] = \beta_0 \) for all \( N \geq K \).

### 6 Large \( N \) fixed \( T \) asymptotics

#### 6.1 Strong consistency and unbiasedness

In the case that \( T \geq 2 \) is fixed, the actual ICM objective function may be taken as

\[
\widehat{Q}_N(\beta) = \sum_{t=1}^{T} \widehat{Q}_{t,N}(\beta) = (\beta - \beta_0)' \left( \sum_{t=1}^{T} \Sigma_{t,N} \right) (\beta - \beta_0) - 2(\beta - \beta_0)' \sum_{t=1}^{T} \hat{\eta}_{t,N}
\]

\[
+ \sum_{t=1}^{T} \hat{\mu}_{t,N},
\]

hence

\[
\widehat{\beta}_N = \beta_0 + \left( \sum_{t=1}^{T} \Sigma_{t,N} \right)^{-1} \sum_{t=1}^{T} \hat{\eta}_{t,N} = \left( \sum_{t=1}^{T} \Sigma_{t,N} \right)^{-1} \sum_{t=1}^{T} \hat{\mu}_{t,N},
\]

where \( \hat{\mu}_{t,N} \) is defined in (22).
It is now easy to verify, similar to Lemma 3, that

**Theorem 1.** Under Assumptions 1-5, $\hat{\beta}_N \xrightarrow{a.s.} \beta_0$, and for sufficiently large $N$, $\hat{\beta}_N$ is unbiased, in the sense that there exists a fixed $K \in \mathbb{N}$ such that $E[\hat{\beta}_N] = \beta_0$ for all $N \geq K$.

### 6.2 Asymptotic normality

The next step is to show that under Assumptions 1-5, $\sqrt{N}(\hat{\beta}_N - \beta_0) \xrightarrow{d} \mathcal{N}_p(0, \Delta)$ for some asymptotic variance matrix $\Delta$. First, let us verify that $\sqrt{N} \sum_{t=1}^{T} \tilde{\eta}_{t,N} \xrightarrow{d} \mathcal{N}_p(0, \Delta_1)$ for some asymptotic variance matrix $\Delta_1$, as follows. Recall that

$$\sqrt{N} \sum_{t=1}^{T} \tilde{\eta}_{t,N} = \sum_{t=1}^{T} \int_{\Xi} \text{Re} \left[ \left( \frac{1}{N} \sum_{i=1}^{N} \tilde{X}_{i,t} \omega_i(\xi) \right) \left( \frac{1}{\sqrt{N}} \sum_{j=1}^{N} \tilde{R}_{j,t} \overline{\omega_j(\xi)} \right) \right] f(\xi) d\xi$$

$$= \frac{1}{\sqrt{N}} \sum_{i=1}^{N} \int_{\Xi} \text{Re} \left[ \left( \sum_{t=1}^{T} E[X_{1,t} - E[X_{1,t}]] \omega_1(\xi) \right) \right]$$

$$\times R_{i,t} \left( \overline{\omega_i(\xi)} - E[\omega_1(\xi)] \right) f(\xi) d\xi + o_p(1), \quad (28)$$

where the second equality in (28) follows from

$$\frac{1}{\sqrt{N}} \sum_{i=1}^{N} \tilde{R}_{i,t} \omega_i(\xi)$$

$$= \frac{1}{\sqrt{N}} \sum_{i=1}^{N} \tilde{R}_{i,t} \left( \omega_i(\xi) - E[\omega_1(\xi)] \right)$$

$$= \frac{1}{\sqrt{N}} \sum_{i=1}^{N} R_{i,t} \left( \omega_i(\xi) - E[\omega_1(\xi)] \right)$$

$$- \sqrt{N} R_{N,t} \frac{1}{N} \sum_{j=1}^{N} \left( \omega_j(\xi) - E[\omega_1(\xi)] \right)$$

$$= \frac{1}{\sqrt{N}} \sum_{i=1}^{N} R_{i,t} \left( \omega_i(\xi) - E[\omega_1(\xi)] \right) + o_p(1) \quad (29)$$
and similar,

\[
\frac{1}{N} \sum_{i=1}^{N} \tilde{X}_{i,t} \omega_i(\xi)
\]

\[
= \frac{1}{N} \sum_{i=1}^{N} (X_{i,t} - E[X_{1,t}]) \left( \omega_i(\xi) - \overline{\omega}_T(\xi) \right)
\]

\[
= \frac{1}{N} \sum_{i=1}^{N} (X_{N,t} - E[X_{1,t}]) \left( \omega_i(\xi) - \overline{\omega}_T(\xi) \right)
\]

\[
= \frac{1}{N} \sum_{i=1}^{N} (X_{i,t} - E[X_{1,t}]) \omega_i(\xi) + o_p(1)
\]

\[
= \frac{1}{N} \sum_{i=1}^{N} (X_{i,t} - E[X_{1,t}]) \omega_i(\xi) + o_p(1)
\]

\[
= E \left[ (X_{1,t} - E[X_{1,t}]) \omega_1(\xi) \right] + o_p(1),
\]

where the \(o_p(1)\) terms in both equations are uniform in \(\xi \in \Xi\), due to the USLLN and the easy facts that \(\sqrt{N R_{N,t}} = O_p(1)\) and \(\sqrt{N (X_{N,t} - E[X_{1,t}])} = O_p(1)\). Thus,

\[
\sqrt{N} \sum_{t=1}^{T} \tilde{\eta}_{t,N} = \frac{1}{\sqrt{N}} \sum_{i=1}^{N} W_i + o_p(1),
\]

where

\[
W_i = \int_{\Xi} \text{Re} \left[ \left( \sum_{t=1}^{T} R_{i,t} E \left[ (X_{1,t} - E[X_{1,t}]) \omega_1(\xi) \right] \right) \left( \overline{\omega}_1(\xi) - E[\overline{\omega}_1(\xi)] \right) \right] \times f(\xi) d\xi
\]

\[
= \sum_{t=1}^{T} R_{i,t} \int_{\Xi} \text{Re} \left[ g_t(\xi) \left( \overline{\omega}_1(\xi) - E[\overline{\omega}_1(\xi)] \right) \right] f(\xi) d\xi
\]

\[
= \sum_{t=1}^{T} R_{i,t} \int_{\Xi} \phi_{i,t}(\xi) f(\xi) d\xi
\]

with

\[
g_t(\xi) = E \left[ (X_{1,t} - E[X_{1,t}]) \omega_1(\xi) \right]
\]

\[
\phi_{i,t}(\xi) = \text{Re} \left[ g_t(\xi) \left( \overline{\omega}_1(\xi) - E[\overline{\omega}_1(\xi)] \right) \right]
\]

17
Clearly, the $W_i$'s are i.i.d., with zero expectation and variance matrix
\[
\Delta_1 = E[W_1W_1'] = \sum_{t_1=1}^{T} \sum_{t_2=2}^{T} E \left[ R_{t_1} R_{t_2} \left( \int \phi_{1,t_1}(\xi)f(\xi)d\xi \right) \times \left( \int \phi_{1,t_2}(\xi)f(\xi)d\xi \right)' \right].
\] (33)

Hence, it follows from the standard multivariate central limit theorem that
\[
\sqrt{N} \sum_{t=1}^{T} \tilde{\eta}_{t,N} = \frac{1}{\sqrt{N}} \sum_{i=1}^{N} W_i + o_p(1) \overset{d}{\rightarrow} N_p(0, \Delta_1).
\] (34)

Finally, denote
\[
\Delta_2 = \sum_{t=1}^{T} \Sigma_t,
\] (35)
where $\Sigma_t$ is given by (12). Then

**Theorem 2.** Under Assumptions 1-5, and with $T$ fixed, $\sqrt{N}(\hat{\beta}_N - \beta_0) \overset{d}{\rightarrow} N_p(0, \Delta_2^{-1}\Delta_1\Delta_2^{-1})$, where $\Delta_1$ and $\Delta_2$ are defined by (33) and (35), respectively.

### 6.3 A consistent estimator of the asymptotic variance matrix

As we have seen before, the matrix $\Delta_2$ can be estimated consistently by
\[
\hat{\Delta}_{2,N} = \sum_{t=1}^{T} \hat{\Sigma}_{t,N}.
\] (36)

As to the matrix $\Delta_1$, recall first that by (30) and (31),
\[
\hat{g}_{N,t}(\xi) = \frac{1}{N} \sum_{i=1}^{N} \tilde{X}_{i,t}\omega_i(\xi) \overset{a.s.}{\rightarrow} E \left[ (X_{1,t} - E[X_{1,t}])\omega_1(\xi) \right] = g_t(\xi),
\] (37)
uniformly on $\Xi$. Moreover, denoting

$$\hat{\phi}_{i,t,N}(\xi) = \text{Re} \left[ \hat{g}_{N,t}(\xi) \left( \frac{\omega_i(\xi)}{N} - \frac{1}{N} \sum_{k=1}^{N} \omega_k(\xi) \right) \right]$$

(38)

$$= \frac{1}{N} \sum_{j=1}^{N} \tilde{X}_{j,t} \text{Re} \left[ \omega_j(\xi) \left( \frac{\omega_i(\xi)}{N} - \frac{1}{N} \sum_{k=1}^{N} \omega_k(\xi) \right) \right]$$

$$= \frac{1}{N} \sum_{j=1}^{N} \tilde{X}_{j,t} \left( \text{Re} \left[ \omega_j(\xi) \omega_i(\xi) \right] - \frac{1}{N} \sum_{k=1}^{N} \text{Re} \left[ \omega_j(\xi) \omega_k(\xi) \right] \right)$$

it is not hard to verify from (37), the USLLN and $\sup_{i \in \mathbb{N}} \sup_{\xi \in \Xi} |\omega_i(\xi)| < \infty$ that

$$\sup_{i \in \mathbb{N}} \sup_{\xi \in \Xi} \left\| \hat{\phi}_{i,t,N}(\xi) - \phi_{i,t}(\xi) \right\|_{a.s.} \to 0,$$

(39)

where $\phi_{i,t}(\xi)$ is defined by (32).

Now by (33) and (39) it not hard to verify that

$$\Delta_1 = \frac{1}{N} \sum_{i=1}^{N} \sum_{t_1=1}^{T} \sum_{t_2=1}^{T} R_{i,t_1} R_{i,t_2} \int_{\Xi} \int_{\Xi} \tilde{\phi}_{i,t_1,N}(\xi_1) \tilde{\phi}_{i,t_2,N}(\xi_2) f(\xi_1) f(\xi_2) d\xi_1 d\xi_2$$

$$+ o_p(1)$$

$$= \frac{1}{N} \sum_{i=1}^{N} \sum_{t_1=1}^{T} \sum_{t_2=1}^{T} R_{i,t_1} R_{i,t_2} \left( \int_{\Xi} \tilde{\phi}_{i,t_1,N}(\xi) f(\xi) d\xi \right) \left( \int_{\Xi} \tilde{\phi}_{i,t_2,N}(\xi) f(\xi) d\xi \right)'$$

$$+ o_p(1)$$

$$= \frac{1}{N} \sum_{i=1}^{N} \sum_{t_1=1}^{T} \sum_{t_2=1}^{T} \tilde{R}_{i,t_1} \tilde{R}_{i,t_2} \left( \int_{\Xi} \tilde{\phi}_{i,t_1,N}(\xi) f(\xi) d\xi \right) \left( \int_{\Xi} \tilde{\phi}_{i,t_2,N}(\xi) f(\xi) d\xi \right)'$$

$$+ o_p(1)$$

$$= \hat{\Delta}_{1,N} + o_p(1),$$

where

$$\hat{\Delta}_{1,N} = \frac{1}{N} \sum_{i=1}^{N} \sum_{t_1=1}^{T} \sum_{t_2=1}^{T} \left( \tilde{Y}_{i,t_1} - \tilde{\beta}'_{N} X_{i,t_1} \right) \left( \tilde{Y}_{i,t_2} - \tilde{\beta}'_{N} X_{i,t_2} \right)$$

$$\times \left( \int_{\Xi} \tilde{\phi}_{i,t_1,N}(\xi) f(\xi) d\xi \right) \left( \int_{\Xi} \tilde{\phi}_{i,t_2,N}(\xi) f(\xi) d\xi \right)'$$

(40)
with
\[
\int_{\Xi} \hat{\phi}_{i,t,N}(\xi)f(\xi)d\xi = \frac{1}{N} \sum_{j=1}^{N} \hat{X}_{j,t} \int_{\Xi} \text{Re} \left[ \omega_j(\xi)\overline{\omega_i(\xi)} \right] f(\xi)d\xi \\
- \frac{1}{N} \sum_{j=1}^{N} \hat{X}_{j,t} \frac{1}{N} \sum_{k=1}^{N} \int_{\Xi} \text{Re} \left[ \omega_j(\xi)\overline{\omega_k(\xi)} \right] f(\xi)d\xi. \tag{41}
\]

Thus,

**Theorem 3.** Under Assumptions 1-5, and with \( T \) fixed, we have,
\[
p \lim_{N \to \infty} \hat{\Delta}_{2,N}^{-1} \hat{\Delta}_{1,N} \hat{\Delta}_{2,N}^{-1} = \Delta_2^{-1} \Delta_1 \Delta_2^{-1},
\]
where \( \hat{\Delta}_{2,N} \) and \( \hat{\Delta}_{1,N} \) are defined by (36) and (40), respectively.

**Remark 4.** Obviously, the results in Theorems 1-3 hold for the case \( T = 1 \) as well. Consequently, similar to Theorems 2 and 3, we can augment Lemma 3 with the following asymptotic normality results.

**Lemma 4.** Under the conditions of Lemma 3,
\[
\sqrt{N}(\hat{\beta}_{t,N} - \beta_0) \overset{d}{\to} N_p 0, \Sigma_t^{-1} \Delta_t \Sigma_t^{-1},
\]
where \( \Sigma_t \) is defined by (12) and
\[
\Delta_t = E \left[ R_{1,t}^2 \left( \int_{\Xi} \phi_{1,t}(\xi)f(\xi)d\xi \right) \left( \int_{\Xi} \phi_{1,t}(\xi)f(\xi)d\xi \right) ^\top \right]. \tag{42}
\]

with \( \phi_{1,t}(\xi) \) defined by (32). Moreover, \( \hat{\Sigma}_{t,N}^{-1} \hat{\Delta}_{t,N} \hat{\Sigma}_{t,N}^{-1} \overset{p}{\to} \Sigma_t^{-1} \Delta_t \Sigma_t^{-1}, \) where \( \hat{\Sigma}_{t,N} \) is defined by (12) and
\[
\hat{\Delta}_{t,N} = \frac{1}{N} \sum_{i=1}^{N} \left( \hat{Y}_{i,t} - \hat{\beta}'_{t,N} X_{i,t} \right)^2 \left( \int_{\Xi} \hat{\phi}_{i,t,N}(\xi)f(\xi)d\xi \right) \left( \int_{\Xi} \hat{\phi}_{i,t,N}(\xi)f(\xi)d\xi \right) ^\top,
\]
where the integrals involved are the same as in (41).
7 Fixed effects

Suppose that the $V_i$'s are fixed effects, i.e., $\Pr[V_i = v_i] = 1$, where the $v_i$'s are nonrandom values. Denote for $v \in \mathbb{R}$,

$$G_N(v) = \frac{1}{N} \sum_{i=1}^{N} I(v_i \leq v),$$

where $I(.)$ is the indicator function. Now suppose that

**Assumption 6.** There exists a proper distribution function $G(v)$ on $\mathbb{R}$ such that $\lim_{N \to \infty} G_N(v) = G(v)$ pointwise in the continuity points of $G(v)$. In addition, let $\sup_{i \in \mathbb{N}} |v_i| < \infty$.

Then by bounded convergence,

$$\lim_{N \to \infty} \frac{1}{N} \sum_{i=1}^{N} v_i = \lim_{N \to \infty} \int v dG_N(v) = \int v dG(v),$$

$$\lim_{N \to \infty} \frac{1}{N} \sum_{i=1}^{N} v_i^2 = \lim_{N \to \infty} \int v^2 dG_N(v) = \int v^2 dG(v),$$

where the first equalities follow from the definition of the integrals $\int v dG_N(v)$ and $\int v^2 dG_N(v)$.

More generally, for $\tau \in \mathbb{R}$ and $i = \sqrt{-1}$,

$$\lim_{N \to \infty} \frac{1}{N} \sum_{i=1}^{N} \exp(i.\tau.v_i) = \lim_{N \to \infty} \int \exp(i.\tau.v) dG_N(v) = \int \exp(i.\tau.v) dG(v),$$

where the latter is the characteristic function of $G$.

Moreover, since the fixed effects are also taken in deviation of their sample means, we may without loss of generality assume that $\int v dG(v) = 0$, so that $\lim_{N \to \infty} \frac{1}{N} \sum_{i=1}^{N} v_i = 0$.

Next, denote $\overline{v}_N = \frac{1}{N} \sum_{i=1}^{N} v_i$ and note that

$$\frac{1}{\sqrt{N}} \sum_{j=1}^{N} (v_j - \overline{v}_N) \omega_j(\xi) = \frac{1}{\sqrt{N}} \sum_{j=1}^{N} (v_j - \overline{v}_N) (\omega_j(\xi) - E[\omega_1(\xi)])$$

\footnote{See for example Bierens (2004, Section 2.3).}
\[
\frac{1}{\sqrt{N}} \sum_{j=1}^{N} \nu_j (\omega_j(\xi) - E[\omega_1(\xi)])
\]

\[
- \tau_N \frac{1}{\sqrt{N}} \sum_{j=1}^{N} (\omega_j(\xi) - E[\omega_1(\xi)])
\]

It can be shown that, with \( \Rightarrow \) denoting weak convergence [c.f. Billingsley (1968)],

\[
W_N(\xi) = \frac{1}{\sqrt{N}} \sum_{j=1}^{N} (\omega_j(\xi) - E[\omega_1(\xi)]) \Rightarrow W(\xi)
\]  \(43\)

where \( W(\xi) \) is a complex-valued zero-mean Gaussian process on \( \Xi \), with covariance function

\[
\Gamma(\xi_1, \xi_2) = E \left[ (\omega_1(\xi_1) - E[\omega_1(\xi)]) (\overline{\omega_1(\xi_2)} - E[\omega_1(\xi)]) \right], \quad 44
\]

hence by the continuous mapping theorem,

\[
\sup_{\xi \in \Xi} \left| \frac{1}{\sqrt{N}} \sum_{j=1}^{N} (\omega_j(\xi) - E[\omega_1(\xi)]) \right| \xrightarrow{d} \sup_{\xi \in \Xi} |W(\xi)|
\]

and thus, since \( \lim_{N \to \infty} \tau_N = 0 \),

\[
\tau_N \frac{1}{\sqrt{N}} \sum_{j=1}^{N} (\omega_j(\xi) - E[\omega_1(\xi)]) = o_p(1)
\]

uniformly on \( \Xi \). Moreover,

\[
E \left[ \left( \frac{1}{\sqrt{N}} \sum_{j=1}^{N} \nu_j (\omega_j(\xi_1) - E[\omega_1(\xi_1)]) \right) \left( \frac{1}{\sqrt{N}} \sum_{j=1}^{N} \nu_j (\overline{\omega_j(\xi_2)} - E[\omega_1(\xi_2)]) \right) \right]
\]

\[
= \frac{1}{N} \sum_{j=1}^{N} v_j^2 E \left[ (\omega_1(\xi_1) - \overline{\sigma_T(\xi_1)}) (\overline{\omega_1(\xi_2)} - E[\omega_1(\xi_2)]) \right]
\]

\[
= \frac{1}{N} \sum_{j=1}^{N} v_j^2 \Gamma(\xi_1, \xi_2) \sim \int v^2 dG(v) \times \Gamma(\xi_1, \xi_2)
\]
This suggests that similar to (43),

\[ W_N^*(\xi) = \frac{1}{\sqrt{N}} \sum_{j=1}^{N} v_j (\omega_j(\xi) - E[\omega_1(\xi)]) \Rightarrow W^*(\xi) \quad (45) \]

where \( W^*(\xi) \) is a zero-mean Gaussian process with covariance function

\[ \Gamma^*(\xi_1, \xi_2) = \int v^2 dG(v) \times \Gamma(\xi_1, \xi_2). \quad (46) \]

The proofs of (43) and (45) depend on the choice of \( \omega_j(\xi) \). In particular, if

\[ \omega_i(\xi) = \begin{cases} \exp(i\xi'Z_i) & \text{if } Z_i \text{ is bounded,} \\ \exp(i\xi\Phi(Z_i)) & \text{if } Z_i \text{ is not bounded,} \end{cases} \]

where \( Z_i \) is defined by (9) and \( \Phi : \mathbb{R}^{q+(p-q)T} \to \mathbb{R}^{q+(p-q)T} \) is a bounded one-to-one mapping with Borel measurable inverse, these results follow similar to Lemma 4 in Bierens (1990, Lemma 4) and the argument following this lemma.\(^8\)

Recall that in the random effect case and by (29),

\[
\frac{1}{\sqrt{N}} \sum_{i=1}^{N} \tilde{R}_{i,t} \omega_i(\xi) = \frac{1}{\sqrt{N}} \sum_{i=1}^{N} R_{i,t} (\omega_i(\xi) - E[\omega_1(\xi)]) + o_p(1)
\]

\[
= \frac{1}{\sqrt{N}} \sum_{i=1}^{N} U_{i,t} (\omega_i(\xi) - E[\omega_1(\xi)])
\]

\[
+ \frac{1}{\sqrt{N}} \sum_{i=1}^{N} V_i (\omega_i(\xi) - E[\omega_1(\xi)]) + o_p(1),
\]

whereas in the fixed effect case, with \( R_{i,t} = v_i + U_{i,t} \),

\[
\frac{1}{\sqrt{N}} \sum_{i=1}^{N} \tilde{R}_{i,t} \omega_i(\xi) = \frac{1}{\sqrt{N}} \sum_{i=1}^{N} R_{i,t} (\omega_i(\xi) - E[\omega_1(\xi)]) + o_p(1)
\]

\(^8\)Similar to Lemma 4 in Bierens (1990), using the series expansion of the complex \( \exp() \) function and the boundedness of the \( v_j \)'s, it follows that \( W_N(\xi) \) and \( W_N^*(\xi) \) are tight. Moreover, using the standard (Lindeberg-Levy) central limit theorem in the case \( W_N(\xi) \) and Lyapunov central limit theorem in the case \( W_N^*(\xi) \), applied to their real and imaginary parts jointly, it can be shown that their finite distributions converge to the finite distributions of \( W(\xi) \) and \( W^*(\xi) \). Then the results (43) and (45) follow from Billingsley (1968, p. 47).
\[
\frac{1}{\sqrt{N}} \sum_{i=1}^{N} U_{i,t} (\omega_i(\xi) - E[\omega_1(\xi)]) + \frac{1}{\sqrt{N}} \sum_{i=1}^{N} v_i (\omega_i(\xi) - E[\omega_1(\xi)]) + o_p(1).
\]

Now suppose that the random effects \( V_i \) are random drawings from the distribution function \( G \) in Assumption 5, so that \( V_i \) is independent of the \( U_{i,t} \)'s and the \( X_{i,t} \)'s. Then similar to (45),

\[
\frac{1}{\sqrt{N}} \sum_{j=1}^{N} V_j (\omega_j(\xi) - E[\omega_1(\xi)]) \Rightarrow W^*(\xi)
\]

(47)
as well, with the same covariance function as (46).

Moreover, also

\[
\frac{1}{\sqrt{N}} \sum_{i=1}^{N} U_{i,t} (\omega_i(\xi) - E[\omega_1(\xi)]) \Rightarrow H_t(\xi)
\]

where \( H_t(\xi) \) is a zero-mean Gaussian process, so that

\[
\frac{1}{\sqrt{N}} \sum_{i=1}^{N} U_{i,t} (\omega_i(\xi) - E[\omega_1(\xi)]) = O_p(1)
\]

uniformly on \( \Xi \). Hence,

\[
\sqrt{N} \sum_{t=1}^{T} \tilde{\eta}_{t,N} = \sum_{t=1}^{T} \int_{\Xi} \text{Re} \left[ \frac{1}{N} \sum_{i=1}^{N} \tilde{X}_{i,t} \omega_i(\xi) \frac{1}{\sqrt{N}} \sum_{j=1}^{N} \tilde{R}_{j,t} \omega_j(\xi) \right] f(\xi) d\xi
\]

\[
= \int_{\Xi} \left( \frac{1}{\sqrt{N}} \sum_{i=1}^{N} \sum_{t=1}^{T} U_{i,t} \text{Re} \left[ g_t(\xi) \left( \omega_i(\xi) - E[\omega_1(\xi)] \right) \right] \right) f(\xi) d\xi
\]

\[
+ \int_{\Xi} \text{Re} \left[ \left( \sum_{t=1}^{T} g_t(\xi) \right) \frac{1}{\sqrt{N}} \sum_{i=1}^{N} V_i (\omega_i(\xi) - E[\omega_1(\xi)]) \right] f(\xi) d\xi + o_p(1),
\]

24
where the $o_p(1)$ term follows from (37), and

$$g_t(\xi) = E [(X_{1,t} - E[X_{1,t}])\omega_1(\xi)].$$

Again it can be shown that

$$\left( \frac{1}{\sqrt{N}} \sum_{i=1}^{N} \sum_{t=1}^{T} U_{i,t} \Re \left[ g_t(\xi) \left( \overline{\omega_i(\xi)} - E[\overline{\omega_1(\xi)}] \right) \right] \right) \Rightarrow H(\xi)$$

where $H(\xi)$ is a zero-mean multivariate Gaussian process with covariance function

$$C(\xi_1, \xi_2) = E[H(\xi_1)H(\xi_2)^\prime]$$

$$= E \left[ \sum_{t=1}^{T} U_{1,t} \Re \left[ g_t(\xi_1) \left( \overline{\omega_1(\xi_1)} - E[\overline{\omega_1(\xi_1)}] \right) \right] \right]$$

$$\times E \left[ \sum_{t=1}^{T} U_{1,t} \Re \left[ g_t(\xi_2)^\prime \left( \overline{\omega_1(\xi_2)} - E[\overline{\omega_1(\xi_2)}] \right) \right] \right]$$

hence,

$$\int_{\Xi} \left( \frac{1}{\sqrt{N}} \sum_{i=1}^{N} \sum_{t=1}^{T} U_{i,t} \Re \left[ g_t(\xi) \left( \overline{\omega_i(\xi)} - E[\overline{\omega_1(\xi)}] \right) \right] \right) f(\xi)d\xi d \Rightarrow \int_{\Xi} H(\xi)f(\xi)d\xi,$$

whereas by (47),

$$\int_{\Xi} \Re \left[ \left( \sum_{t=1}^{T} g_t(\xi) \right) \left( \frac{1}{\sqrt{N}} \sum_{i=1}^{N} V_i \left( \overline{\omega_i(\xi)} - E[\overline{\omega_1(\xi)}] \right) \right) \right] f(\xi)d\xi$$

$$\Rightarrow \int_{\Xi} \Re \left[ \left( \sum_{t=1}^{T} g_t(\xi) \right) W^*(\xi) \right] f(\xi)d\xi$$

where both right-hand integrals are zero-mean normal random vectors. Moreover, these normal random vectors are independent, because the $V_i$’s are independent of the $U_{i,t}$’s and the $X_{j,t}$’s. Furthermore,

$$\text{Var} \left( \int_{\Xi} H(\xi)f(\xi)d\xi \right)$$
\[
\begin{align*}
&= \int_{\mathbb{R}^2} E[H(\xi_1)H(\xi_2)] f(\xi_1) f(\xi_2) \, d\xi_1 \, d\xi_2 \\
&= E \left[ \int_{\mathbb{R}^2} \sum_{t=1}^T U_{1,t} \Re \left[ g_t(\xi_1) \left( \overline{\omega_1(\xi_1)} - E[\omega_1(\xi_1)] \right) \right] f(\xi_1) \, d\xi_1 \\
&\quad \times \int_{\mathbb{R}^2} \sum_{t=1}^T U_{1,t} \Re \left[ g_t(\xi_2) \left( \overline{\omega_1(\xi_2)} - E[\omega_1(\xi_2)] \right) \right] f(\xi_2) \, d\xi_2 \right] \\
&= \sum_{t_1=1}^T \sum_{t_2=1}^T E[U_{1,t_1} U_{1,t_2}] \int_{\mathbb{R}^2} \Re \left[ g_{t_1}(\xi) \left( \overline{\omega_1(\xi)} - E[\omega_1(\xi)] \right) \right] f(\xi) \, d\xi \\
&\quad \times \int_{\mathbb{R}^2} \Re \left[ g_{t_2}(\xi) \left( \overline{\omega_1(\xi)} - E[\omega_1(\xi)] \right) \right] f(\xi) \, d\xi \\
&= \sum_{t_1=1}^T \sum_{t_2=1}^T E[U_{1,t_1} U_{1,t_2} \phi_{1,t_1} \phi'_{1,t_2}]
\end{align*}
\]

where \( \phi_{1,t} \) is defined by (32). Similarly

\[
\Var \left( \int_{\mathbb{R}^2} \Re \left[ \left( \sum_{t=1}^T g_t(\xi) \right) W^*(\xi) \right] f(\xi) \, d\xi \right) = \int v^2 dG(v) \sum_{t_1=1}^T \sum_{t_2=1}^T E[\phi_{1,t_1} \phi_{1,t_2}].
\]

Thus, in this case the variance matrix \( \Delta_1 \) in (33) can be written as

\[
\Delta_1 = \Var \left( \int_{\mathbb{R}^2} H(\xi) f(\xi) \, d\xi \right) + \Var \left( \int_{\mathbb{R}^2} \Re \left[ \left( \sum_{t=1}^T g_t(\xi) \right) W^*(\xi) f(\xi) \, d\xi \right] \right)
\]

\[
= \sum_{t_1=1}^T \sum_{t_2=1}^T E[U_{1,t_1} U_{1,t_2} \phi_{1,t_1} \phi'_{1,t_2}] + \int v^2 dG(v) \sum_{t_1=1}^T \sum_{t_2=1}^T E[\phi_{1,t_1} \phi_{1,t_2}]
\]

\[
= \sum_{t_1=1}^T \sum_{t_2=1}^T E[R_{1,t_1} R_{1,t_2} \phi_{1,t_1} \phi'_{1,t_2}],
\]

where \( R_{i,t} = U_{i,t} + V_i \), and the last equality follows from the condition that \( V_i \) is independent of the \( U_{i,t} \)'s and the \( X_{j,t} \)'s. The same applies for \( R_{i,t} = U_{i,t} + v_i \), where the fixed effects \( v_i \) satisfy Assumption 6.

Summarizing, the following results have been shown.

**Theorem 4.** In the fixed effects case, and under Assumption 6, the asymptotic normality result in Theorem 2 carries over, with random effects \( V_i \)
drawn independently from the distribution $G(v)$ in Assumption 6, provided that the function $w(u)$ in Assumption 4 is the real or complex exp() function. Moreover, the parameter estimate $\hat{\beta}_N$ is now weakly consistent, i.e., $p\lim_{N \to \infty} \hat{\beta}_N = \beta_0$, but no longer unbiased.

However, in the sequel only the random effects case will be considered.

8 A consistent ICM model misspecification test

8.1 The ICM test

In this section we will show that the statistic $N\hat{Q}_N(\hat{\beta}_N)$ yields a consistent ICM type test of the null hypothesis (8) against the general alternative hypothesis (14) that the null hypothesis is false.

Let us first consider the alternative hypothesis (14). Under this alternative hypothesis we no longer have $\hat{\beta}_N \xrightarrow{a.s.} \beta_0$, but it is not hard to verify that under the other conditions in Assumptions 1-5, $\hat{\beta}_N \xrightarrow{a.s.} \beta_\ast$ for some $\beta_\ast \in \mathbb{R}^p$. Then $R_{i,t} - \bar{R}_{N,t}$ becomes

$$R_{i,t}^* - \bar{R}_{N,t}^* = \bar{Y}_{i,t} - \beta_\ast \bar{X}_{i,t},$$

where the $R_{i,t}^*$'s are still i.i.d. with $E[R_{i,t}^*] = 0$, so that by the USLLN and $\bar{R}_{N,t}^* \xrightarrow{a.s.} 0$,

$$\frac{1}{N} \sum_{i=1}^{N} (R_{i,t}^* - \bar{R}_{N,t}) \omega_i(\xi) = \frac{1}{N} \sum_{i=1}^{N} R_{i,t}^* \omega_i(\xi) - \bar{R}_{N,t}^* \frac{1}{N} \sum_{i=1}^{N} \omega_i(\xi)
\xrightarrow{a.s.} E[R_{1,t}^* \omega_1(\xi)],$$

uniformly on $\Xi$, and similarly,

$$\frac{1}{N} \sum_{i=1}^{N} \bar{X}_{i,t} \omega_i(\xi) \xrightarrow{a.s.} E[(X_{1,t} - E[X_{1,t}]) \omega_1(\xi)]$$

uniformly on $\Xi$. Consequently,

$$\frac{1}{N} \sum_{i=1}^{N} \left( \bar{Y}_{i,t} - \beta_\ast \bar{X}_{i,t} \right) \omega_i(\xi) = \frac{1}{N} \sum_{i=1}^{N} (R_{i,t}^* - \bar{R}_{N,t}^*) \omega_i(\xi)$$
uniformly on $\Xi$, and thus,

$$
\hat{Q}_N(\beta_N) = \sum_{t=1}^{T} \int_{\Xi} \left| \frac{1}{N} \sum_{i=1}^{N} \left( \hat{Y}_{i,t} - \hat{\beta}_N^{'} \tilde{X}_{i,t} \right) \omega_i(\xi) \right|^2 f(\xi) d\xi
$$

$$
\xrightarrow{a.s.} \sum_{t=1}^{T} \int_{\Xi} \left| E[R_{i,t}^* \omega_1(\xi)] \right|^2 f(\xi) d\xi > 0 \text{ under } H_1,
$$

(50)

where the inequality follows from the fact that, instead of (8) we now have

$$
\Pr \left( E[R_{i,t}^* | G_{i,T}] = 0 \right) < 1,
$$

hence by Lemma 1, $E[R_{i,t}^* \omega_1(\xi)] \neq 0$ a.e.

Next, consider the null hypothesis (8). It follows trivially from (26), (27) and (36) that

$$
N\hat{Q}_N(\beta_N)
$$

$$
= N \sum_{t=1}^{T} \sum_{i=1}^{N} \tilde{R}_{i,t} \omega_1(\xi) \left( \sum_{\tau=1}^{T} \tilde{\Sigma}_{\tau,N} \right)^{-1} \sqrt{N} \sum_{t=1}^{T} \tilde{\eta}_{t,N}
$$

$$
= \sum_{t=1}^{T} \int_{\Xi} \left| \frac{1}{\sqrt{N}} \sum_{i=1}^{N} \tilde{R}_{i,t} \omega_1(\xi) \right|^2 f(\xi) d\xi
$$

$$
- \sum_{t=1}^{T} \int_{\Xi} \Re \left[ \left( \frac{1}{N} \sum_{i=1}^{N} \tilde{X}_{i,t} \omega_i(\xi) \right) \left( \frac{1}{\sqrt{N}} \sum_{j=1}^{N} \tilde{R}_{j,t} \omega_j(\xi) \right) \right] f(\xi) d\xi
$$

$$
\times \tilde{\Delta}_{2,N}^{-1} \sum_{t=1}^{T} \int_{\Xi} \Re \left[ \left( \frac{1}{N} \sum_{i=1}^{N} \tilde{X}_{i,t} \omega_i(\xi) \right) \left( \frac{1}{\sqrt{N}} \sum_{j=1}^{N} \tilde{R}_{j,t} \omega_j(\xi) \right) \right] f(\xi) d\xi.
$$

(51)

Since $\sum_{i=1}^{N} \tilde{R}_{i,t} = 0$ we can write

$$
\frac{1}{\sqrt{N}} \sum_{i=1}^{N} \tilde{R}_{i,t} \omega_i(\xi) = \frac{1}{\sqrt{N}} \sum_{i=1}^{N} \tilde{R}_{i,t} (\omega_i(\xi) - E[\omega_1(\xi)])
$$

$$
= \frac{1}{\sqrt{N}} \sum_{i=1}^{N} R_{i,t} (\omega_i(\xi) - E[\omega_1(\xi)])
$$

28
\[-\sqrt{N} R_{N,t} \frac{1}{N} \sum_{i=1}^{N} (\omega_i(\xi) - E[\omega_1(\xi)])\]
\[= \frac{1}{\sqrt{N}} \sum_{i=1}^{N} R_{i,t} (\omega_i(\xi) - E[\omega_1(\xi)]) + o_p(1) \quad (52)\]

where the \(o_p(1)\) term is uniform on \(\Xi\), due to \(\sqrt{N} R_{N,t} = O_p(1)\) and

\[\sup_{\xi \in \Xi} \left| \frac{1}{N} \sum_{i=1}^{N} \omega_i(\xi) - E[\omega_1(\xi)] \right| \xrightarrow{a.s.} 0.\]

Furthermore, it can be shown, using Lemma A.1 in Bierens and Ploberger (1997), that for each \(t\),

\[\hat{W}_{N,t}(\xi) = \frac{1}{\sqrt{N}} \sum_{i=1}^{N} R_{i,t} (\omega_i(\xi) - E[\omega_1(\xi)]) \Rightarrow W_t(\xi), \quad (53)\]

where \(W_t(\xi)\) is a zero-mean Gaussian process on \(\Xi\) with covariance function

\[\Gamma_t(\xi_1, \xi_2) = E[W_t(\xi_1)W_t(\xi_2)] = E \left[ R_{1,t}^2 (\omega_1(\xi) - E[\omega_1(\xi)]) (\omega_1(\xi) - E[\omega_1(\xi)]) \right].\]

Since we can now write

\[N \hat{Q}_N(\hat{\beta}_N) = \sum_{t=1}^{T} \int_{\Xi} \left| \hat{W}_{N,t}(\xi) \right|^2 f(\xi) d\xi\]
\[-\sum_{t=1}^{T} \int_{\Xi} \text{Re} \left[ \hat{g}_{N,t}(\xi) \overline{\hat{W}_{N,t}(\xi)} \right] f(\xi) d\xi\]
\[\times \Delta_{2,N}^{-1} \sum_{t=1}^{T} \int_{\Xi} \text{Re} \left[ \hat{g}_{N,t}(\xi) \overline{\hat{W}_{N,t}(\xi)} \right] f(\xi) d\xi + o_p(1).\]

where \(\hat{g}_{N,t}(\xi)\) is defined in (37), the latter result together with (53) suggest that under the null hypothesis (8), \(N \hat{Q}_N(\hat{\beta}_N)\) converges in distribution to

\[\sum_{t=1}^{T} \int_{\Xi} |W_t(\xi)|^2 f(\xi) d\xi - \sum_{t=1}^{T} \int_{\Xi} \text{Re} \left[ g_t(\xi) \overline{W_t(\xi)} \right] f(\xi) d\xi\]
\[\times \Delta_{2,N}^{-1} \sum_{t=1}^{T} \int_{\Xi} \text{Re} \left[ g_t(\xi) \overline{W_t(\xi)} \right] f(\xi) d\xi.\]
However, this is only correct if the $W_i$’s are independent across $t$, which requires that the $R_{i,t}$’s are independent across $t$. In other words, (53) does not imply that \( \left( \widehat{W}_{N,1}(\xi), \widehat{W}_{N,2}(\xi), \ldots, \widehat{W}_{N,T}(\xi) \right)' \) converges weakly to \( (W_1(\xi), W_2(\xi), \ldots, W_T(\xi))' \) because the possible serial dependence of the $R_{i,t}$’s has not been taken into account.

A somewhat similar problem occurred with the test in Bierens (1984), which involves a weighted sum of ICM test statistics. This problem was only recently solved, in Bierens (2017, Chapter 3: Addendum). Mimicking the latter solution in the present case, let

\[
\widehat{W}_N(\xi, u) = \sum_{t=1}^{T} \widehat{W}_{N,t}(\xi) \kappa_t(u)
\]

\[
= \frac{1}{\sqrt{N}} \sum_{i=1}^{N} \left( \sum_{t=1}^{T} R_{i,t} \kappa_t(u) \right) \left( \omega_i(\xi) - E[\omega_1(\xi)] \right), \quad u \in [0, 1],
\]

(54)

for example, where $\kappa_1(u), \kappa_2(u), \ldots, \kappa_T(u)$ are continuous orthonormal functions on $[0, 1]$. For example, let

\[
\kappa_t(u) = \sqrt{2} \cos(t \pi u),
\]

(55)

which satisfy \( \int_0^1 \kappa_{t_1}(u) \kappa_{t_2}(u) du = I(t_1 = t_2) \) for $t_1, t_2 \in \mathbb{N}$. Then it is easy to verify that

\[
\sum_{t=1}^{T} \int_{\Xi} \left| \widehat{W}_{N,t}(\xi) \right|^2 f(\xi)d\xi = \int_{\Xi \times [0,1]} \left| \widehat{W}_N(\xi, u) \right|^2 f(\xi)d\xi du
\]

(56)

and

\[
\sum_{t=1}^{T} \int_{\Xi} \text{Re} \left[ \widehat{g}_{N,t}(\xi) \overline{\widehat{W}_{N,t}(\xi)} \right] f(\xi)d\xi
\]

\[
= \int_{\Xi \times [0,1]} \text{Re} \left[ \widehat{g}_N(\xi, u) \overline{\widehat{W}_N(\xi, u)} \right] f(\xi)d\xi du,
\]

(57)

where

\[
\widehat{g}_N(\xi, u) = \sum_{t=1}^{T} \widehat{g}_{N,t}(\xi) \kappa_t(u)
\]

(58)
Hence, the ICM test statistic takes the form

\[
\hat{\Upsilon}_{0,N} \overset{\text{def.}}{=} N \hat{Q}_N(\hat{\beta}_N)
\]

\[
= \int_{\Xi \times [0,1]} \left| \hat{W}_N(\xi, u) \right|^2 f(\xi) d\xi du
\]

\[
- \int_{\Xi \times [0,1]} \text{Re} \left[ \hat{g}_N(\xi, u) \hat{W}_N(\xi, u) \right] f(\xi) d\xi du
\]

\[
\times \Delta_2^{-1} \int_{\Xi \times [0,1]} \text{Re} \left[ g_N(\xi, u) \hat{W}_N(\xi, u) \right] f(\xi) d\xi du. \quad (59)
\]

Again, it can be shown on the basis of Lemma A.1 in Bierens and Ploberger (1997) that under the null hypothesis (8),

\[
\hat{W}_N(\xi, u) \Rightarrow W(\xi, u) = \sum_{t=1}^{T} \kappa_t(u) W_t(\xi) \text{ on } \Xi \times [0,1], \quad (60)
\]

where \( W(\xi, u) \) is a zero-mean Gaussian process on \( \Xi \times [0,1] \) with covariance function

\[
\Gamma((\xi_1, u_1), (\xi_2, u_2)) = E[W(\xi_1, u_1)W(\xi_2, u_2)]
\]

\[
= \sum_{t_1=1}^{T} \sum_{t_2=1}^{T} E\left[R_{1,t_1}R_{1,t_2}(\omega_1(\xi_1) - E[\omega_1(\xi_1)])(\omega_1(\xi_2) - E[\omega_1(\xi_2)]) \right]
\]

\[
\times \kappa_{t_1}(u_1) \kappa_{t_2}(u_2). \quad (61)
\]

Then by the continuous mapping theorem, \( \hat{\Upsilon}_{0,N} \overset{d}{\to} \Upsilon_0 \) under \( H_0 \), where

\[
\Upsilon_0 \sim \int_{\Xi \times [0,1]} |W(\xi, u)|^2 f(\xi) d\xi du
\]

\[
- \int_{\Xi \times [0,1]} \text{Re} \left[ g(\xi, u) W(\xi, u) \right] f(\xi) d\xi du
\]

\[
\times \Delta_2^{-1} \int_{\Xi \times [0,1]} \text{Re} \left[ g(\xi, u) W(\xi, u) \right] f(\xi) d\xi du \quad (62)
\]

with

\[
g(\xi, u) = \sum_{t=1}^{T} q_t(\xi) \kappa_t(u). \quad (63)
\]
Note that the choice of the continuous orthonormal functions $\kappa_1(u), \kappa_2(u), \ldots, \kappa_T(u)$ in (54) is arbitrary. For example, the equalities (56) and (57) also hold if we choose $\kappa_t(u) = \sqrt{2} \sin(2t\pi u)$, because then $\int_0^1 \kappa_{t_1}(u)\kappa_{t_2}(u)du = I(t_1 = t_2)$ for $t_1, t_2 \in \mathbb{N}$ as well. Consequently, the null distribution (62) is invariant for the choice of $\kappa_t(u)$’s, as is easy to verify by integrating $u$ out in (62).

8.2 Bootstrap versions of the ICM test

The idea of the bootstrap is to create versions $\widehat{W}_{N,b}(\xi, u)$ of $\widehat{W}_N(\xi, u)$ such that

$$\widehat{W}_{N,b}(\xi, u) \Rightarrow W_b(\xi, u)$$

where for each bootstrap $b$, $W_b(\xi, u)$ is a zero-mean Gaussian process on $\Xi \times [0, 1]$ with the same covariance function (61) as $W(\xi, u)$, hence $W_b(\xi, u) \sim W(\xi, u)$, and where $\widehat{W}_{N,b}(\xi, u)$ is constructed such that the $W_b(\xi, u)$’s are independent of each other and of $W(\xi, u)$.

Replacing $\widehat{W}_N(\xi, u)$ in (59) by $\widehat{W}_{N,b}(\xi, u)$, the bootstrap version of $\Upsilon_N$ now takes the form

$$\widehat{\Upsilon}_{N,b} = \int_{\Xi \times [0, 1]} |\widehat{W}_{N,b}(\xi, u)|^2 f(\xi)d\xi du$$

$$- \int_{\Xi \times [0, 1]} \text{Re} \left[ \overline{\widehat{g}'_N(\xi, u)} \widehat{W}_{N,b}(\xi, u) \right] f(\xi)d\xi du$$

$$\times \overline{\Delta_{2,N}^{-1}} \int_{\Xi \times [0, 1]} \text{Re} \left[ \overline{\widehat{g}_N(\xi, u)} \overline{\widehat{W}_{N,b}(\xi, u)} \right] f(\xi)d\xi du,$$

(64)
satisfying $\widehat{\Upsilon}_{N,b} \overset{d}{\to} \Upsilon_b$, where

$$\Upsilon_b = \int_{\Xi \times [0, 1]} |W_b(\xi, u)|^2 f(\xi)d\xi du$$

$$- \int_{\Xi \times [0, 1]} \text{Re} \left[ g(\xi, u)\overline{W_b(\xi, u)} \right] f(\xi)d\xi du$$

$$\times \overline{\Delta_{2}^{-1}} \int_{\Xi \times [0, 1]} \text{Re} \left[ g(\xi, u)\overline{W_b(\xi, u)} \right] f(\xi)d\xi du$$

(65)

$$\sim \text{i.i.d. } \Upsilon_0,$$

with $g(\xi, u)$ defined by (63).
The proposed version of \( \hat{W}_{N,b}(\xi, u) \) is
\[
\hat{W}_{N,b}(\xi, u) = \frac{1}{\sqrt{N}} \sum_{i=1}^{N} e_{i,b} \sum_{t=1}^{T} \kappa_t(u)(\tilde{Y}_{i,t} - \tilde{\beta}_N^T \tilde{X}_{i,t})
\times \left( \omega_i(\xi) - \frac{1}{N} \sum_{j=1}^{N} \omega_j(\xi) \right)
\]
(66)

where the \( e_{i,b} \)'s are i.i.d. \( N(0, 1) \), motivated as follows. Note first that
\[
\hat{W}_{N,b}(\xi, u) = \frac{1}{\sqrt{N}} \sum_{i=1}^{N} e_{i,b} \sum_{t=1}^{T} \kappa_t(u)(\tilde{R}_{i,t} - (\tilde{\beta}_N - \beta_0)^T \tilde{X}_{i,t})
\times \left( \omega_i(\xi) - \frac{1}{N} \sum_{j=1}^{N} \omega_j(\xi) \right)
\]
\[
= \frac{1}{\sqrt{N}} \sum_{i=1}^{N} e_{i,b} \sum_{t=1}^{T} \kappa_t(u) \tilde{R}_{i,t} \left( \omega_i(\xi) - \frac{1}{N} \sum_{j=1}^{N} \omega_j(\xi) \right)
\]
\[
- \sum_{m=1}^{p} (\tilde{\beta}_{m,N} - \beta_{0,m}) \frac{1}{\sqrt{N}} \sum_{i=1}^{N} e_{i,b} \sum_{t=1}^{T} \kappa_t(u) \tilde{X}_{m,i,t}
\times \left( \omega_i(\xi) - \frac{1}{N} \sum_{j=1}^{N} \omega_j(\xi) \right)
\]
\[
= \hat{W}^*_N(\xi, u) - \sum_{m=1}^{p} (\tilde{\beta}_{m,N} - \beta_{0,m}) \hat{H}_{m,N}(\xi, u),
\]
say, where \( \tilde{\beta}_{m,N}, \beta_{0,m} \) and \( \tilde{X}_{m,i,t} \) are components \( m \) of \( \tilde{\beta}_N, \beta_0 \) and \( \tilde{X}_{i,t} \), respectively, and
\[
\hat{W}^*_N(\xi, u) = \frac{1}{\sqrt{N}} \sum_{i=1}^{N} e_{i,b} \sum_{t=1}^{T} \kappa_t(u) \tilde{R}_{i,t} \left( \omega_i(\xi) - \frac{1}{N} \sum_{j=1}^{N} \omega_j(\xi) \right),
\]
\[
\hat{H}_{m,N}(\xi, u) = \frac{1}{\sqrt{N}} \sum_{i=1}^{N} e_{i,b} \sum_{t=1}^{T} \kappa_t(u) \tilde{X}_{m,i,t} \left( \omega_i(\xi) - \frac{1}{N} \sum_{j=1}^{N} \omega_j(\xi) \right)
\]
Denoting
\[
W^*_N(\xi, u) = \frac{1}{\sqrt{N}} \sum_{i=1}^{N} e_{i,b} \left( \sum_{t=1}^{T} \kappa_t(u) R_{i,t} \right) \left( \omega_i(\xi) - E[\omega_1(\xi)] \right),
\]
(67)
it is easy to verify that
\[ \sup_{(\xi, u) \in \Xi \times [0, 1]} \left| \hat{W}_{N, b}^*(\xi, u) - W_{N, b}^*(\xi, u) \right| = o_p(1) \]
and
\[ W_{N, b}^*(\xi, u) \Rightarrow W_b(\xi, u), \]
where \( W_b(\xi, u) \) has the same covariance function (61) as \( W(\xi, u) \), i.e.,
\[ E \left[ W_{N, b}^*(\xi_1, u_1) W_{N, b}^*(\xi_2, u_2) \right] = \Gamma((\xi_1, u_1), (\xi_2, u_2)) \]

hence
\[ E \left[ W_b(\xi_1, u_1) W_b(\xi_2, u_2) \right] = \Gamma((\xi_1, u_1), (\xi_2, u_2)). \]

Similarly, denoting
\[ H_{m, N, b}(\xi, u) = \frac{1}{\sqrt{N}} \sum_{i=1}^{N} e_{i,b} \sum_{t=1}^{T} \kappa_t(u) X_{m,i,t} (\omega_i(\xi) - E[\omega_1(\xi)]) \]
we have \( \sup_{(\xi, u) \in \Xi \times [0, 1]} \left| \hat{H}_{m, N, b}(\xi, u) - H_{m, N, b}(\xi, u) \right| = o_p(1) \) and
\[ \sup_{(\xi, u) \in \Xi \times [0, 1]} |H_{m, N, b}(\xi, u)| = O_p(1). \]

Thus,
\[ \hat{W}_{N, b}(\xi, u) = W_{N, b}^*(\xi, u) + o_p(1) \Rightarrow W_b(\xi, u). \]

where the \( o_p(1) \) term is uniform on \( \Xi \times [0, 1] \).

Due to the \( e_{i,b} \)'s in (66), the bootstrap statistics \( \tilde{\Upsilon}_{N, b} \) are i.i.d., conditional on the current data. Moreover, for \( b_1 \neq b_2 \) it follows trivially from (67) that
\[ E \left[ W_{N, b_1}^*(\xi_1, u_1) W_{N, b_2}^*(\xi_2, u_2) \right] = E \left[ W_{b_1}(\xi_1, u_1) W_{b_2}(\xi_2, u_2) \right] = 0, \]
which by the Gaussianity of \( W_b \) implies that \( W_{b_1} \) and \( W_{b_2} \) are independent. More generally, the \( W_b \)'s are i.i.d. and so are the \( \Upsilon_b \)'s.

So far the bootstrap results have been derived under \( H_0 \), so the question arises: What happens under \( H_1 \)? First, recall that under \( H_1, \hat{\beta}_N \xrightarrow{a.s.} \beta_* \) for some \( \beta_* \in \mathbb{R}^p \). Then by (48), (67) now becomes
\[ W_{b,N}^*(\xi, u) = \frac{1}{\sqrt{N}} \sum_{i=1}^{N} e_{i,b} \left( \sum_{t=1}^{T} R_{i,t}^* \kappa_t(u) \right) (\omega_i(\xi) - E[\omega_1(\xi)]), \]
which converges weakly to a zero mean Gaussian process $W^*_b(\xi, u)$ with covariance function

$$
\Gamma^*((\xi_1, u_1), (\xi_2, u_2)) = \sum_{t_1=1}^{T} \sum_{t_2=1}^{T} E \left[ R^*_{1,t_1} R^*_{1,t_2} (\omega_1(\xi_1) - E[\omega_1(\xi_1)]) (\omega_1(\xi_2) - E[\omega_1(\xi_2)]) \right] \times \kappa_{t_1}(u_1) \kappa_{t_2}(u_2).
$$

Hence,

$$
\tilde{Y}_{b,N}^* = \int_{\Xi \times [0,1]} |W^*_b(\xi, u)|^2 f(\xi) d\xi du \xrightarrow{d} \int_{\Xi \times [0,1]} |W^*_b(\xi, u)|^2 f(\xi) d\xi du \sim \Upsilon^*_b,
$$

say, where similar as under $H_0$ the $\Upsilon^*_b$’s are i.i.d.

Summarizing, the following results hold.

**Theorem 5.**

(1) Let $\tilde{Y}_{0,N} = N.\hat{Q}_N(\hat{\beta}_N)$. Then under Assumptions 1-5, including the null hypothesis $H_0$ in (8),

$$
\tilde{Y}_{0,N} \xrightarrow{d} \Upsilon_0,
$$

where $\Upsilon_0$ represents a random variable with non-pivotal distribution, whereas if $H_0$ is false but the other conditions in Assumptions 1-5 hold,

$$
\tilde{Y}_{0,N}/N \xrightarrow{a.s.} \lambda > 0,
$$

where $\lambda$ is a constant.

(2) For some large natural number $B$ and for $b = 1, 2, ..., B$, let

$$
\tilde{A}_{N,b} = \tilde{C}_{N,b} = \tilde{D}_{2,N}^{-1} \tilde{C}_{N,b},
$$

where

$$
\tilde{A}_{N,b} = \frac{1}{N} \sum_{i=1}^{N} \sum_{j=1}^{N} e_{i,b} e_{j,b} \left( \sum_{t=1}^{T} (\tilde{Y}_{i,t} - \hat{\beta}_{N}^* \tilde{X}_{i,t}) (\tilde{Y}_{j,t} - \hat{\beta}_{N}^* \tilde{X}_{j,t}) \right) \times \int_{\Xi} \text{Re} \left[ \omega_i(\xi) \omega_j(\xi) \right] f(\xi) d\xi,
$$

$$
\tilde{C}_{N,b} = \frac{1}{\sqrt{N}} \sum_{i=1}^{N} e_{i,b} \sum_{t=1}^{T} (\tilde{Y}_{i,t} - \hat{\beta}_{N}^* \tilde{X}_{i,t}) \frac{1}{N} \sum_{j=1}^{N} \tilde{X}_{j,t}
$$

$$
\times \left( \int_{\Xi} \text{Re} \left[ \omega_j(\xi) \omega_i(\xi) \right] f(\xi) d\xi - \frac{1}{N} \sum_{k=1}^{N} \int_{\Xi} \text{Re} \left[ \omega_j(\xi) \omega_k(\xi) \right] f(\xi) d\xi \right),
$$

(70)
be the bootstrap versions of the ICM test \( \hat{\Upsilon}_{0,N} \), where the \( e_{i,b} \)'s are independent random drawing from the standard normal distribution and \( \hat{\Delta}_{2,N} \) is defined by (36). Then under \( H_0 \),

\[
\left( \hat{\Upsilon}_{0,N}, \hat{\Upsilon}_{1,N}, \hat{\Upsilon}_{2,N}, \ldots, \hat{\Upsilon}_{B,N} \right)' \xrightarrow{d} (\Upsilon_0, \Upsilon_1, \Upsilon_2, \ldots, \Upsilon_B)',
\]

where the random variables \( \Upsilon_b \) for \( b = 0, 1, 2, \ldots, B \) are i.i.d. \( \Upsilon_0 \), whereas under \( H_1 \),

\[
\left( \hat{\Upsilon}_{1,N}, \hat{\Upsilon}_{2,N}, \ldots, \hat{\Upsilon}_{B,N} \right)' \xrightarrow{d} (\Upsilon_1^*, \Upsilon_2^*, \ldots, \Upsilon_B^*)',
\]

where the random variables \( \Upsilon_b^* \) are also i.i.d. but no longer distributed as \( \Upsilon_0 \).

Note that the expressions (68), (69) and (70) are just straightforward further elaborations of the corresponding expressions in (64).

It remains to show how to use the results in part (2) of Theorem 5 to derive bootstrap critical values, using the empirical distribution function

\[
G_{B,N}(\tau) = \frac{1}{B} \sum_{b=1}^{B} I(\hat{\Upsilon}_{b,N} \leq \tau)
\]

as an approximation of the asymptotic null distribution function

\[
G(\tau) = \Pr[\Upsilon_0 \leq \tau].
\]

In particular, we are interested in approximating the true \( \alpha \times 100\% \) asymptotic critical value, \( c(\alpha) \), which is the \( 1 - \alpha \) quantile of \( G \), i.e.,

\[
c(\alpha) = \arg \min_{G(\tau) > 1 - \alpha} \tau,
\]

by the \( 1 - \alpha \) quantile of \( G_{B,N} \), i.e.,

\[
c_{B,N}(\alpha) = \arg \min_{G_{B,N}(\tau) > 1 - \alpha} \tau.
\]

First note that similar to Theorem 6.5 in Bierens (2017, p. 108) it follows that \( G(\tau) \) is continuous on \((0, \infty)\), so that exactly \( G(c(\alpha)) = 1 - \alpha \). On the other hand, for the bootstrap critical value \( c_{B,N}(\alpha) \) we only know that \( G_{B,N}(c_{B,N}(\alpha)) \geq 1 - \alpha \).
Moreover, similar to Theorem 6.3 in Bierens (2017, p. 105) and its proof it follows that pointwise in $\tau > 0$,

$$\lim_{N \to \infty} E \left[ (G_{B,N}(\tau) - G(\tau))^2 \right] = \frac{1}{B} (G(\tau) - G(\tau)^2) \leq \frac{1}{4B},$$

(71)

so that by Chebyshev’s inequality,

$$\lim_{N \to \infty} \Pr \left[ |G_{B,N}(\tau) - G(\tau)| \geq \delta \right] \leq \frac{G(\tau) - G(\tau)^2}{B\delta^2} \leq \frac{1}{4B\delta^2}$$

(72)

for arbitrary $\delta > 0$.

Next, observe that for $\varepsilon \in (0, c(\alpha))$,

$$\Pr [c_{B,N}(\alpha) > c(\alpha) + \varepsilon]$$

$$\leq \Pr [G_{B,N}(c_{B,N}(\alpha)) \geq G_{B,N}(c(\alpha) + \varepsilon)]$$

$$\leq \Pr [G_{B,N}(c(\alpha) + \varepsilon) \leq 1 - \alpha]$$

$$= \Pr [G_{B,N}(c(\alpha) + \varepsilon) - G(c(\alpha) + \varepsilon) \leq 1 - \alpha - G(c(\alpha) + \varepsilon)]$$

$$\leq \Pr [|G_{B,N}(c(\alpha) + \varepsilon) - G(c(\alpha) + \varepsilon)| \geq G(c(\alpha) + \varepsilon) - 1 + \alpha]$$

and similarly,

$$\Pr [c_{B,N}(\alpha) < c(\alpha) - \varepsilon]$$

$$\leq \Pr [|G_{B,N}(c(\alpha) - \varepsilon) - G(c(\alpha) - \varepsilon)| \geq 1 - \alpha - G(c(\alpha) - \varepsilon)] .$$

Hence by (72),

$$\lim \sup_{N \to \infty} \Pr [c_{B,N}(\alpha) > c(\alpha) + \varepsilon] \leq \frac{G(c(\alpha) + \varepsilon) - G(c(\alpha) + \varepsilon)^2}{B (G(c(\alpha) + \varepsilon) - 1 + \alpha)^2}$$

and

$$\lim \sup_{N \to \infty} \Pr [c_{B,N}(\alpha) > c(\alpha) - \varepsilon] \leq \frac{G(c(\alpha) - \varepsilon) - G(c(\alpha) - \varepsilon)^2}{B (1 - \alpha - G(c(\alpha) - \varepsilon))^2} .$$

Thus, the following result holds.

**Theorem 6.** Under the conditions of Theorem 5, including the null hypothesis $H_0$ in (8), and for $\varepsilon \in (0, c(\alpha))$,

$$\lim \sup_{N \to \infty} \Pr [|c_{B,N}(\alpha) - c(\alpha)| > \varepsilon]$$

$$\leq \frac{G(c(\alpha) + \varepsilon) - G(c(\alpha) + \varepsilon)^2}{B (G(c(\alpha) + \varepsilon) - 1 + \alpha)^2} + \frac{G(c(\alpha) - \varepsilon) - G(c(\alpha) - \varepsilon)^2}{B (1 - \alpha - G(c(\alpha) - \varepsilon))^2}$$

37
so that
\[
\lim_{B \to \infty} \lim_{N \to \infty} \Pr \left[ |c_{B,N}(\alpha) - c(\alpha)| > \varepsilon \right] = 0.
\]

**Remark 5.** This result carries over under the alternative hypothesis (14) as well, except that then \(G(\tau) = \Pr[T^*_I \leq \tau]\), so that \(c(\alpha)\) is no longer the correct asymptotic \(\alpha \times 100\%\) critical value. However, due to the consistency of the ICM test involved that does not matter, because then \(\hat{\Upsilon}_{0,N} > c(\alpha)\) as \(N \to \infty\).

**Remark 6.** Note that the result (71) implies
\[
\lim_{N \to \infty} E \left[ (G_{B,N}(c(\alpha)) - G(c(\alpha)))^2 \right] = \lim_{N \to \infty} E \left[ (G_{B,N}(c(\alpha)) - (1 - \alpha))^2 \right] = (\alpha - \alpha^2) / B.
\]
This result can be used, as a rule of thumb, to select the value of \(B\) for given \(\alpha\). In particular, choose \(B\) so large that for a small \(\varepsilon > 0\), 
\((\alpha - \alpha^2) / B \leq \varepsilon\). For example, for \(\alpha = 0.05\) and \(\varepsilon = 0.0001\) this rule of thumb indicates to choose \(B \geq 475\).

**Remark 7.** Rather than computing bootstrap critical values, it is easier and almost equivalent to use the bootstrap p-value, i.e.,
\[
\hat{p}_{B,N} = \frac{1}{B} \sum_{b=1}^{B} I \left( \hat{\Upsilon}_{b,N} > \hat{\Upsilon}_{0,N} \right),
\]
(73)
because the event \(\hat{\Upsilon}_{0,N} > c_{B,N}(\alpha)\) implies that \(\hat{p}_{B,N} < \alpha\), and the other way around, \(\hat{p}_{B,N} < \alpha\) is equivalent to \(G_{B,N}(\hat{\Upsilon}_{0,N}) > 1 - \alpha\), hence \(\hat{\Upsilon}_{0,N} \geq \arg \min_{G_{B,N}(\tau > 1 - \alpha)} \tau = c_{B,N}(\alpha)\).

### 9 Testing for time varying \(\beta\)’s

If the parameter vector \(\beta_0\) is no longer constant but time varying, then the original model (7) is misspecified, which by the consistency of the ICM test \(\hat{\Upsilon}_{1,N}\) defined in (59) will be detected for large sample sizes \(N\). Thus, if the ICM test \(\hat{\Upsilon}_{0,N}\) does not reject the null hypothesis (8) then there is no need to test for time varying \(\beta\)’s. On the other hand, if the ICM test rejects the null
hypothesis, then either the $\beta$ parameters are time varying, or the functional form of the model is nonlinear, or both. In the former case model (7) reads

$$Y_{i,t} = \alpha_{0,t} + \beta_0 t X_{i,t} + R_{i,t}, \quad i = 1, 2, \ldots, N, \quad t = 1, 2, \ldots, T.$$  \hspace{1cm} (74)

As we have seen before, given the null hypothesis (8), for each $t$ the parameter vector $\beta_t$ can be estimated consistently by $\hat{\beta}_{t,N} = \arg\min_{\beta \in \mathbb{R}^p} \hat{Q}_{t,N}(\beta) = \beta_t + \hat{\Sigma}_{t,N}^{-1}\hat{\eta}_{t,N}$, so that

$$\hat{Q}_{t,N}(\hat{\beta}_{t,N}) = (\hat{\beta}_{t,N} - \beta_t)^T \hat{\Sigma}_{t,N}(\beta_t - \beta_t) - 2(\hat{\beta}_{t,N} - \beta_t)^T \hat{\eta}_{t,N} + \hat{\rho}_{t,N}$$

This suggest to use, in the time varying $\beta$'s case, the ICM test statistic

$$\tilde{\Upsilon}_{1,N} \overset{\text{def.}}{=} N \sum_{t=1}^{T} \hat{Q}_{t,N}(\hat{\beta}_{t,N}) = N \sum_{t=1}^{T} \hat{\rho}_{t,N} - \sum_{t=1}^{T} \sqrt{N} \hat{\eta}_{t,N} \hat{\Sigma}_{t,N}^{-1} \sqrt{N} \hat{\eta}_{t,N}$$

Note that similar to (51), we can write,

$$\tilde{\Upsilon}_{1,N} = \sum_{t=1}^{T} \int_{\Xi} \left| \frac{1}{\sqrt{N}} \sum_{i=1}^{N} \tilde{R}_{t,i} \omega_i(\xi) \right|^2 f(\xi) d\xi$$

$$- \sum_{t=1}^{T} \int_{\Xi} \text{Re} \left[ \left( \frac{1}{N} \sum_{i=1}^{N} \tilde{X}_{t,i} \omega_i(\xi) \right) \left( \frac{1}{\sqrt{N}} \sum_{j=1}^{N} \tilde{R}_{t,j} \omega_j(\xi) \right) \right] f(\xi) d\xi$$

$$\times \hat{\Sigma}_{t,N}^{-1} \int_{\Xi} \text{Re} \left[ \left( \frac{1}{N} \sum_{i=1}^{N} \tilde{X}_{t,i} \omega_i(\xi) \right) \left( \frac{1}{\sqrt{N}} \sum_{j=1}^{N} \tilde{R}_{t,j} \omega_j(\xi) \right) \right] f(\xi) d\xi$$

where $\tilde{W}_N(\xi, u)$ is the same as in (54), and $\tilde{g}_{N,t}(\xi)$ is defined in (37). Thus, under the null hypothesis (8), $\tilde{\Upsilon}_{1,N} \overset{d}{\rightarrow} \Upsilon_0$, where now

$$\Upsilon_0 = \int_{\Xi \times [0,1]} \left| W(\xi, u) \right|^2 f(\xi) d\xi du$$
whereas under the alternative hypothesis (14),

\[
\hat{\Upsilon}_{1,N}/\sqrt{N} \xrightarrow{a.s.} \sum_{t=1}^{T} \int_{\Xi} \left( E[R_{1,t}(\xi)] \right)^{2} f(\xi) d\xi > 0.
\]

Moreover, the corresponding bootstrap test statistics take the form

\[
\tilde{\Upsilon}_{b,N} = A_{N,b} - \sum_{t=1}^{T} \tilde{C}_{t,N,b} \tilde{\Sigma}_{t,N} \tilde{C}_{t,N,b}, \quad \text{with}
\]

\[
A_{N,b} = \frac{1}{N} \sum_{i=1}^{N} \sum_{j=1}^{N} \varepsilon_{i,b} \varepsilon_{j,b} \left( \sum_{t=1}^{T} \left( \tilde{Y}_{i,t} - \tilde{\beta}_{t,N} \tilde{X}_{i,t} \right) \left( \tilde{Y}_{j,t} - \tilde{\beta}_{t,N} \tilde{X}_{j,t} \right) \right)
\]

\[
\times \int_{\Xi} \text{Re} \left[ \omega_{i}(\xi) \omega_{j}(\xi) \right] f(\xi) d\xi,
\]

\[
\tilde{C}_{t,N,b} = \frac{1}{\sqrt{N}} \sum_{i=1}^{N} e_{i,b}(\tilde{Y}_{i,t} - \tilde{\beta}_{t,N} \tilde{X}_{i,t}) \frac{1}{N} \sum_{j=1}^{N} \tilde{X}_{j,t}
\]

\[
\times \left( \int_{\Xi} \text{Re} \left[ \omega_{j}(\xi) \omega_{i}(\xi) \right] f(\xi) d\xi - \frac{1}{N} \sum_{k=1}^{N} \int_{\Xi} \text{Re} \left[ \omega_{j}(\xi) \omega_{k}(\xi) \right] f(\xi) d\xi \right).
\]

It follows now straightforwardly that:

**Theorem 7.** With \( \hat{\Upsilon}_{0,N} \) replaced by \( \hat{\Upsilon}_{1,N} = N \sum_{t=1}^{T} \tilde{Q}_{t,N}(\tilde{\beta}_{t,N}) \) and \( \hat{\Upsilon}_{b,N} \) replaced by (75) the results in Theorems 5 and 6 carry over.

As said before, there is no need to test for time varying \( \beta \)'s if ICM test \( \hat{\Upsilon}_{0,N} \) does not reject. If the latter test rejects the null hypothesis (8) but ICM test \( \hat{\Upsilon}_{1,N} \) does not, then the previous test result was due to time varying \( \beta \)'s. Note that if we conduct ICM test \( \hat{\Upsilon}_{1,N} \) only if ICM test \( \hat{\Upsilon}_{0,N} \) rejects, then due to the consistency of the latter there is, asymptotically, no pretesting problem.
However, if ICM test $\hat{\Upsilon}_{1,N}$ rejects the null hypothesis (8) as well, then the functional form of models (7) or (74) is incorrect, which is much more difficult to repair because our approach hinges crucially on the linearity in the parameters hypothesis. If so, we can still interpret these models as linear projections, as in Arrellano and Honore (2001), but then the ICM estimation procedure is no longer valid, in the sense that the consistency and asymptotic normality results in Theorems 1-3 no longer hold.

10 Implementation

10.1 Standardization of the conditional variables

Bierens (1982) proposed to standardize the conditioning variables, i.e., the components $Z_{i,m}$ of $Z_i \in \mathbb{R}^{q+(p-q)T}$ defined by (9) in our case, before applying a bounded one-to-one transformation, by first taking them in deviation of their sample means and then dividing them by their sample standard errors, i.e., replace $Z_{i,m}$ by $(Z_{i,m} - \bar{Z}_m)/S_m$, where $\bar{Z}_m = (1/N) \sum_{j=1}^{N} Z_{j,m}$ and $S_m = \sqrt{(N - 1)^{-1} \sum_{j=1}^{N} (Z_{j,m} - \bar{Z}_m)^2}$, and then use the arctan(.) transformation, i.e., use

$$
\tilde{Z}_{i,m} = \arctan((Z_{i,m} - \bar{Z}_m)/S_m) \in (-\pi/2, \pi/2), \quad (78)
$$

instead of the original $Z_{i,m}$'s. As shown in Theorem 7.1 in Bierens (2017, p.110), and similar for the present ICM estimation and testing results, asymptotically the effect is the same as in the case that the transformed standardized conditional variables $Z^*_{i,m} = \arctan((Z_{i,m} - \mu_m)/\sigma_m)$ were used, where $\mu_m = E[Z_{i,m}]$ and $\sigma_m = \sqrt{E[(Z_{i,m} - \mu_m)^2]}$.

Strictly speaking, the arctan(.) transformation in (78) is only necessary if the corresponding $Z_{i,m}$ is unbounded, but we advocate the transformation (78) in the latter case as well because then the $\tilde{Z}_{i,m}$'s have approximately the same (zero) location and scale, which proves to be convenient in choosing $f(\xi)$ and $\Xi$. 

41
10.2 The choice of the density $f(\xi)$ and its support $\Xi$

If we choose $w(u) = \exp(u)$ in Assumption 4, the weight function (13) reads

$$\omega_i(\xi) = \prod_{m=1}^{\ell} \exp(\xi_m \tilde{Z}_{i,m}), \text{ with } \ell = q + (p - q)T,$$

where $\xi_m$ is components $m$ of $\xi \in \mathbb{R}^\ell$ and $\tilde{Z}_{i,m}$ is defined by (78), whereas in the case $w(u) = \exp(iu)$,

$$\omega_i(\xi) = \prod_{m=1}^{\ell} \exp(i \xi_m \tilde{Z}_{i,m}) = \prod_{m=1}^{\ell} \left( \cos(\xi_m \tilde{Z}_{i,m}) + i \sin(\xi_m \tilde{Z}_{i,m}) \right).$$

From a practical point of view it is recommended to choose $f(\xi)$ and $\Xi$ such that the integrals $\int_\Xi \omega_i(\xi) \omega_j(\xi) f(\xi) d\xi$ or $\int_\Xi \text{Re}[\omega_i(\xi) \omega_j(\xi)] f(\xi) d\xi$ have a closed form expression, which is, for example, the case if we let $\Xi = \mathcal{X}_{m=1}^{\ell} [-c_m, c_m]$ for some $c_m > 0$ and choose for $f(\xi)$ the uniform density on $\Xi$.

However, since the $\tilde{Z}_{i,m}$’s have approximately the same (zero) location and scale, there is not much loss of generality to choose $\Xi = [-c, c]^\ell$ for some $c > 0$, keeping $f(\xi)$ the uniform density on $\Xi$. Then in the case $w(u) = \exp(u)$,

$$\int_\Xi \omega_i(\xi) \omega_j(\xi) f(\xi) d\xi = \prod_{m=1}^{\ell} \frac{1}{2c} \int_{-c}^{c} \exp \left( \xi_m \tilde{Z}_{i,m} + \tilde{Z}_{j,m} \right) d\xi_m$$

$$= \prod_{m=1}^{\ell} \exp \left( c(\tilde{Z}_{i,m} + \tilde{Z}_{j,m}) \right) - \exp \left( -c(\tilde{Z}_{i,m} + \tilde{Z}_{j,m}) \right)$$

$$= \vartheta_{i,j}(c), \text{ say.} \quad (79)$$

In the case $w(u) = \exp(iu)$ we have

$$\int_\Xi \omega_i(\xi) \omega_j(\xi) f(\xi) d\xi$$

$$= \prod_{m=1}^{\ell} \frac{1}{2c} \int_{-c}^{c} \exp \left( i \xi_m(\tilde{Z}_{i,m} - \tilde{Z}_{j,m}) \right) d\xi_m$$

$$= \prod_{m=1}^{\ell} \frac{1}{2c} \int_{-c}^{c} \left( \cos (\xi_m (\tilde{Z}_{i,m} - \tilde{Z}_{j,m})) + i \sin (\xi_m (\tilde{Z}_{i,m} - \tilde{Z}_{j,m})) \right) d\xi_m$$

42
\[
\prod_{m=1}^{\ell} \frac{\sin (c(\bar{Z}_{i,m} - \bar{Z}_{j,m}))}{c(\bar{Z}_{i,m} - \bar{Z}_{j,m})}
\]

where the last equality follows from \(\frac{1}{2c} \int_{-c}^{c} \sin \left(\xi_m (\bar{Z}_{i,m} - \bar{Z}_{j,m})\right) d\xi_m = 0\), by symmetry. Hence,

\[
\int_{\mathbb{Z}} \text{Re} \left[ \omega_i(\xi) \overline{\omega_j(\xi)} \right] f(\xi) d\xi = \prod_{m=1}^{\ell} \frac{\sin (c(\bar{Z}_{i,m} - \bar{Z}_{j,m}))}{c(\bar{Z}_{i,m} - \bar{Z}_{j,m})} = \vartheta_{i,j}(c), \text{ say.} \quad (80)
\]

Now in either case

\[
\hat{\Sigma}_{t,N} = \frac{1}{N^2} \sum_{i=1}^{N} \sum_{j=1}^{N} \tilde{X}_{i,t} \tilde{X}_{j,t} \vartheta_{i,j}(c),
\]

\[
\hat{\mu}_{t,N} = \frac{1}{N^2} \sum_{i=1}^{N} \sum_{j=1}^{N} \tilde{X}_{i,t} \tilde{Y}_{j,t} \vartheta_{i,j}(c),
\]

and (40) now reads

\[
\hat{\Delta}_{1,N} = \frac{1}{N} \sum_{i=1}^{N} \sum_{t_1=1}^{T} \sum_{t_2=2}^{T} \left( \tilde{Y}_{i,t_1} - \hat{\beta}_N \tilde{X}_{i,t_1} \right) \left( \tilde{Y}_{i,t_2} - \hat{\beta}_N \tilde{X}_{i,t_2} \right)
\]

\[
\times \left( \frac{1}{N} \sum_{j=1}^{N} \tilde{X}_{j,t_1} \left( \vartheta_{j,i}(c) - \frac{1}{N} \sum_{k=1}^{N} \vartheta_{j,k}(c) \right) \right)
\]

\[
\times \left( \frac{1}{N} \sum_{j=1}^{N} \tilde{X}_{j,t_2} \left( \vartheta_{j,i}(c) - \frac{1}{N} \sum_{k=1}^{N} \vartheta_{j,k}(c) \right) \right)
\]

Finally, the bootstrap ICM statistics (68) and (75) now become

\[
\tilde{\Upsilon}_{b,N} = \tilde{A}_{N,b} - \tilde{C}_{N,b} \tilde{\Delta}_{2,N}^{-1} \tilde{C}_{N,b}, \text{ with }
\]

\[
\tilde{A}_{N,b} = \frac{1}{N} \sum_{i=1}^{N} \sum_{j=1}^{N} e_{i,b} e_{j,b} \left( \sum_{t=1}^{T} \left( \tilde{Y}_{i,t} - \hat{\beta}_N \tilde{X}_{i,t} \right) \left( \tilde{Y}_{j,t} - \hat{\beta}_N \tilde{X}_{j,t} \right) \right) \vartheta_{i,j}(c),
\]

\[
\tilde{C}_{N,b} = \frac{1}{\sqrt{N}} \sum_{i=1}^{N} \sum_{j=1}^{T} \left( \tilde{Y}_{i,t} - \hat{\beta}_N \tilde{X}_{i,t} \right) \frac{1}{N} \sum_{j=1}^{N} \tilde{X}_{j,t} \left( \vartheta_{j,i}(c) - \frac{1}{N} \sum_{k=1}^{N} \vartheta_{j,k}(c) \right),
\]

43
and
\[ \tilde{Y}_{b,N} = \tilde{A}_{N,b} = \sum_{t=1}^{T} \tilde{C}_{t,N,b}^r \tilde{\Sigma}_{t,N}^{-1} \tilde{C}_{t,N,b}, \] with
\[ \tilde{A}_{N,b} = \frac{1}{N} \sum_{i=1}^{N} \sum_{j=1}^{N} e_{i,b} e_{j,b} \left( \sum_{t=1}^{T} \left( \tilde{Y}_{i,t} - \tilde{\beta}_{i,N}^t \tilde{X}_{i,t} \right) \left( \tilde{Y}_{j,t} - \tilde{\beta}_{i,N}^t \tilde{X}_{j,t} \right) \right) \varrho_{i,j}(c), \]
\[ \tilde{C}_{t,N,b} = \frac{1}{\sqrt{N}} \sum_{i=1}^{N} e_{i,b} (\tilde{Y}_{i,t} - \tilde{\beta}_{i,N}^t \tilde{X}_{i,t}) \frac{1}{N} \sum_{j=1}^{N} \tilde{X}_{j,t} \left( \varrho_{j,i}(c) - \frac{1}{N} \sum_{k=1}^{N} \varrho_{j,k}(c) \right), \]
respectively.

We now have all the ingredients to implement our ICM estimation and testing approaches in practice.

Admittedly, in either case (79) or (80) the estimators in Lemma 1 and Theorems 1-3 will depend on the choice of \( c \), and so does the finite sample power of the ICM test in Theorem 5 and its bootstrap versions (68), and therefore also the bootstrap critical values in Theorem 6 will depend on \( c \).

As to the choice of \( c \), note that in the case (79),
\[ \lim_{c \to 0} \varrho_{i,j}(c) = 1, \lim_{c \to \infty} \varrho_{i,j}(c) = \infty, \]
whereas in the case (80),
\[ \lim_{c \to 0} \varrho_{i,j}(c) = 1, \lim_{c \to \infty} \varrho_{i,j}(c) = I(i = j). \]

Therefore, in either case one should not choose \( c \) too small or too large. But how small is too small and how large is too large?

The answer depends the criterion for choosing \( c \). One such a criterion is to choose \( c \) such the parameter estimates have the smallest asymptotic variances, for example by choosing \( c \) such that the maximum eigenvalue or the trace of the estimated asymptotic variance matrix \( \hat{\Sigma}_{1,2}^{-1} \hat{\Sigma}_{1,2}^{-1} \hat{\Sigma}_{2,2}^{-1} \) in Theorem 3 is minimal. However, this criterion is only relevant under the null hypothesis (8).

Another criterion is to choose \( c \) such that the finite sample power of the ICM test is maximal. The problem with this criterion is that we do not know the \( R_{i,t}^* \)'s in (49) and (50). On the other hand, as long as we choose \( \beta_* \) in
(48) away from its possible true value $\beta_0$ the resulting residuals $R_{i,t}^*$ satisfy
$\Pr[E[R_{i,t}^*|Z_i] = 0] < 1$ for $t = 1, 2, \ldots, T$, hence
$$\hat{Q}_N(\beta_*) = \sum_{t=1}^{T} \int_{\mathbb{X}} \left( \frac{1}{N} \sum_{i=1}^{N} (\tilde{Y}_{i,t} - \beta_* \tilde{X}_{i,t}) \omega_i(\xi) \right)^2 f(\xi) d\xi$$
\[\xrightarrow{a.s.} \sum_{t=1}^{T} \int_{\mathbb{X}} |E[R_{i,t}^* \omega_1(\xi)]|^2 f(\xi) d\xi > 0.\]

It is easy to verify from (79) or (80) that $\hat{Q}_N(\beta_*)$ can be written as
$$\hat{Q}_N(\beta_*) = \theta'_* \tilde{\Omega}_N(c) \theta_*$$
where $\theta_* = (1, -\beta'_*)'$ and
$$\tilde{\Omega}_N(c) = \sum_{t=1}^{T} \frac{1}{N^2} \sum_{i=1}^{N} \sum_{j=1}^{N} \left( \frac{\tilde{Y}_{i,t}}{\tilde{X}_{i,t}} \right) \left( \frac{\tilde{Y}_{j,t}}{\tilde{X}_{j,t}} \right) \int_{\mathbb{X}} \omega_i(\xi) \omega_j(\xi) f(\xi) d\xi$$
$$= \sum_{t=1}^{T} \frac{1}{N^2} \sum_{i=1}^{N} \sum_{j=1}^{N} \left( \frac{\tilde{Y}_{i,t}}{\tilde{X}_{i,t}} \right) \left( \frac{\tilde{Y}_{j,t}}{\tilde{X}_{j,t}} \right) \vartheta_{i,j}(c).$$

To guarantee that $\beta_*$ is away from a possible true value $\beta_0$, we could use,
for example,
$$\sup_{\theta_* \in \mathbb{R}^{p+1}} \frac{\theta'_* \tilde{\Omega}_N(c) \theta_*}{\theta'_* \theta_*} = \lambda_{\text{max}} \left( \tilde{\Omega}_N(c) \right)$$
as the criterion for choosing $c$, where $\lambda_{\text{max}}$ is the largest eigenvalue of $\tilde{\Omega}_N(c)$. However, instead of the maximum eigenvalue we propose to use the trace of $\tilde{\Omega}_N(c)$, because the latter is much easier and faster to calculate, as
$$\text{trace} \left( \tilde{\Omega}_N(c) \right) = \frac{1}{N^2} \sum_{i=1}^{N} \sum_{j=1}^{N} \sum_{t=1}^{T} \left( \tilde{Y}_{i,t} \tilde{Y}_{j,t} + \tilde{X}_{i,t} \tilde{X}_{j,t} \right) \vartheta_{i,j}(c)$$
$$= \frac{1}{N^2} \sum_{i=1}^{N} \sum_{t=1}^{T} \sum_{t=1}^{T} \left( \tilde{Y}_{i,t}^2 + \tilde{X}_{i,t} \tilde{X}_{i,t} \right) \vartheta_{i,i}(c)$$
$$+ 2 \frac{1}{N^2} \sum_{i=1}^{N} \sum_{j=1}^{N} \sum_{t=1}^{T} \left( \tilde{Y}_{i,t} \tilde{Y}_{j,t} + \tilde{X}_{i,t} \tilde{X}_{j,t} \right) \vartheta_{i,j}(c).$$
and dominates $\lambda_{\text{max}}(\tilde{\Omega}(c))$. Thus, choose $c > 0$ such that $\text{trace}(\tilde{\Omega}(c))$ is maximal. This approach is the one we will use, by maximizing $\text{trace}(\tilde{\Omega}(c))$ over a finite grid of $c$'s. In particular, we will use the grid $c = k/10$ for $k = 1, 2, ..., 50$.

As to the choice of $w(u)$ in Assumption 4, we favor the complex $\exp(.)$ function $w(u) = \exp(i.u)$ because then $\varphi_{i,j}(c)$ is bounded for all $c > 0$, hence $\text{trace}(\tilde{\Omega}(c))$ remains finite on $(0, \infty)$, whereas in the case $w(u) = \exp(u)$, $\lim_{c \to \infty} \text{trace}(\tilde{\Omega}(c)) = \infty$.

11 A limited Monte Carlo analysis

11.1 The model

The true model is an artificial Mincer type log-wage equation:

$$Y_{i,t} = \alpha_0 + \theta_1 E_{i,t} - \theta_2 E_{i,t}^2 + \gamma S_{i,t} + \delta W_i + V_{i,t} + U_{i,t}, t = 1, 2$$

where $t = 1, 2$ represents two consecutive years, and

- $Y_{i,t}$ represents the log-wage,
- $E_{i,t}$ represents years of experience,
- $S_{i,t}$ represents years of schooling,
- $W_i$ represents a race dummy variable: $W_i = 1$ if white, $W_i = 0$ if not,
- $V_i$ is the random effect,
- $U_{i,t}$ is an error term.

Throughout will be assumed that $\theta_1 E_{i,t} - \theta_2 E_{i,t}^2$ takes a maximum at 25 years, hence by the first-order condition, $\theta_2 = \theta_1/50$. Thus, the true model becomes

$$Y_{i,t} = \alpha_0 + \theta_0 E_{i,t} - 0.02\theta_0 E_{i,t}^2 + \gamma_0 S_{i,t} + \delta_0 W_i + R_{i,t}, \quad (81)$$

$$R_{i,t} = V_i + U_{i,t}, \text{ with}$$

$$\Pr [E[R_{i,t}|W_i, E_{i,1}, E_{i,2}, S_{i,1}, S_{i,2}] = 0] = 1, \quad (82)$$
which will be used to analyze the size of the ICM test with bootstrap critical values. Moreover, the power of the ICM tests will be analyzed on the basis of the false null hypothesis

$$Y_{i,t} = \alpha + \theta E_{i,t} + \gamma S_{i,t} + \delta W_i + R^*_{i,t} \quad (83)$$

where for all parameter values

$$\Pr [E[R^*_{i,t}|W_i, E_{i,1}, E_{i,2}, S_{i,1}, S_{i,2}] = 0] < 1. \quad (84)$$

### 11.2 Generation of the variables

The covariates $E_{i,t}$, $S_{i,t}$ and $W_i$ have been generated as follows.

- The race dummy variable $W_i$ has been generated with probability $\Pr[W_i = 1] = 0.5$.

- Conditional on $W_i$, the years of schooling in year $t = 1$, $S_{i,1}$, is generated by a random drawing from the Bin($p, n$) distribution, with $n = 18 + 4W_i$ and $p = 0.5$, so that $E[S_{i,1}|W_i] = 9 + 2W_i$, hence $E[S_{i,1}] = 10$.

- Moreover, $S_{i,2} = S_{i,1} + D_i$, where $D_i$ is a dummy variable with conditional probability $\Pr[D_i = 1|S_{i,1}] = 1/(1 + S_{i,1})$. Thus, the more schooling individual $i$ has in year $t = 1$, the smaller the probability of an extra year of schooling in year $t = 2$.

- Conditional on $S_{i,1}$, the years of experience in year $t = 1$, $E_{i,t}$, is generated as a random drawing from the Bin($p, n$) distribution, with $n = 60 - S_{i,1}$ and $p = 0.75$, so that $E[E_{i,1}|S_{i,1}] = 45 - 0.75S_{i,1}$ and thus $E[E_{i,1}] = 37.5$.

- Moreover, $E_{i,2} = E_{i,1} + D^*_i$, where $D^*_i$ is a dummy variable with probability $\Pr[D^*_i = 0] = 0.25$. The latter represents the probability that individual $i$ is unemployed during year 2.

Furthermore, $V_i$ and the $U_{i,t}$'s have been generated as follows.

- The random effects $V_i$ have been drawn randomly from the uniform $[-1, 1]$ distribution.
The error $U_{i,1}$ has been drawn independently from the standard normal distribution, and $U_{i,2} = 0.5U_{i,1} + \sqrt{0.75}e_i$, where $e_i$ is drawn independently from the standard normal distribution, so that $E[U_{i,1}^2] = E[U_{i,2}^2] = 1$ and $E[U_{i,1}U_{i,2}] = 0.5$.

The $Y_{i,t}$’s are now generated according to data generating process (81), for

$$\alpha_0 = \theta_0 = \gamma_0 = \delta_0 = 1.$$  

Thus, the size of the ICM test will be analyzed on the basis of the model

$$Y_{i,t} = \alpha + X'_{i,t}\beta + R_{i,t}, \text{ where } X'_{i,t} = (E_{i,t}, E_{i,t}^2, S_{i,t}, W_{i}),$$

whereas the power will be analyzed on the basis of the model

$$Y_{i,t} = \alpha_* + X'_t\beta_* + R^*_{i,t}, \text{ where } X'_t = (E_{i,t}, S_{i,t}, W_{i}).$$

In both cases the initial conditioning variables in the weight function are

$$Z_i = (W_i, E_{i,1}, E_{i,2}, S_{i,1}, S_{i,2})'.$$

C.f. (9). The actual conditioning variables are the standardized and transformed ones according to (78).

Finally, as motivated in Remark 6, the bootstrap sample size will be chosen as $B = 500$.

### 11.3 Estimation and testing results

The performance of our ICM estimation approach will be demonstrated by an example for $N = 500$, with $c$ determined by maximizing trace $\left(\hat{\Omega}_N(c)\right)$ for $c = k/10$, with $k = 1, 2, ..., 50$, on the basis of (80). The optimal $c$ appears to be $c = 1.1$, under $H_0$ as well as under $H_1$.

In the following table the estimation results are presented, for $N = 500$. The entries in columns ”s.d.” are the usual standard deviations, i.e., the square roots of the diagonal elements of the matrix $\hat{\Delta}_{1,N}^{-1}\hat{\Delta}_{2,N}\hat{\Delta}_{1,N}^{-1}$ in Theorem 3, divided by $\sqrt{N}$. Moreover, ICM is the value of the ICM test statistic, BPV stands for bootstrap p-value, and BCV stands for bootstrap critical value. As said before, the bootstrap sample size is $B = 500$. 

48
Table 1. Estimation and ICM testing results for $N = 500$, $T = 2$, $c = 1.1$

<table>
<thead>
<tr>
<th>$X_{i,t}$</th>
<th>True $\beta_0$</th>
<th>$\beta_N$</th>
<th>s.d.</th>
<th>$\beta_N$</th>
<th>s.d.</th>
</tr>
</thead>
<tbody>
<tr>
<td>$E_{i,t}^2$</td>
<td>$-0.02$</td>
<td>$-0.021216$</td>
<td>$0.00570$</td>
<td>$-0.574780$</td>
<td>$0.01520$</td>
</tr>
<tr>
<td>$E_{i,t}$</td>
<td>$1$</td>
<td>$1.078081$</td>
<td>$0.44228$</td>
<td>$-0.574780$</td>
<td>$0.01520$</td>
</tr>
<tr>
<td>$S_{i,t}$</td>
<td>$1$</td>
<td>$0.970521$</td>
<td>$0.02436$</td>
<td>$0.995543$</td>
<td>$0.02450$</td>
</tr>
<tr>
<td>$W_{i}$</td>
<td>$1$</td>
<td>$0.945638$</td>
<td>$0.09800$</td>
<td>$0.937017$</td>
<td>$0.09308$</td>
</tr>
<tr>
<td>ICM</td>
<td></td>
<td>$0.574$</td>
<td></td>
<td>$3.054$</td>
<td></td>
</tr>
<tr>
<td>BPV</td>
<td></td>
<td>$0.392$</td>
<td></td>
<td>$0.000$</td>
<td></td>
</tr>
<tr>
<td>BCV</td>
<td></td>
<td>$1.4110$</td>
<td>$1.0480$</td>
<td>$0.9097$</td>
<td>$1.7949$</td>
</tr>
</tbody>
</table>

The large standard deviation of the coefficient of $E_{i,t}$ under $H_0$ may be due to the high correlation between $E_{i,t}$ and $E_{i,t}^2$, namely 0.9977, so that the null model suffers from near multicollinearity. Nevertheless, under $H_0$ the parameter estimates involved are close to their true values, and are not significantly different from their true values at the 5% significance level.

The performance of the ICM test is what is expected. In view of the bootstrap p-values, the true model is not rejected at any conventional significance level, and the misspecified model is strongly rejected at any conventional significance level.

However, the results in Table 1 are based on only two independently generated data sets, so that the ICM test results might be outliers. To check the latter, we have repeated the estimation and testing procedures in Table 1 independently 500 times, using the same value of $c$, i.e., $c = 1.1$, and bootstrap sample size $B = 500$ as in the cases in Table 1. The resulting percentage rejection rates (PRR) at the 1%, 5% and 10% significance levels are presented in Table 2.

Table 2. Bootstrap p-values rejection rates after 500 replications

<table>
<thead>
<tr>
<th>$H_0$</th>
<th>$H_1$</th>
</tr>
</thead>
<tbody>
<tr>
<td>$1%$</td>
<td>$5%$</td>
</tr>
<tr>
<td>PRR</td>
<td>$1.2$</td>
</tr>
</tbody>
</table>

The results in Table 2 indicate that the bootstrap procedure works quite well. In particular, the actual sizes under $H_0$ (the PRR’s) are close to their
nominal sizes,\(^9\) whereas in view of the sample size \(N = 500\) the finite sample power is excellent.

**References**


\(^9\)Of course, it would be better to use more replications than only 500, but the results in Table 2 took about 45 hours to compute on a Lenovo Thinkpad.