Estimating a Cointegrating Regression with Missing Data

J. Isaac Miller
Department of Economics
University of Missouri

Abstract

We consider a cointegrating regression in which the integrated regressors, which may or may not be cointegrated with each other, contain missing observations. Under general conditions on the imputation error, we show that least squares estimation of the cointegrating vector is consistent, even though the estimator is neither asymptotically normal nor unbiased. In order to allow valid statistical inference, we construct a canonical cointegrating regression (CCR) and show that least squares estimation of the CCR provides a consistent and asymptotically normal estimator.

First draft: January 18, 2007
This draft: April 3, 2007

JEL Classification: C13, C32
Key words and phrases: cointegration, canonical cointegrating regression, missing data, mixed-frequency data, imputation, near-epoch dependence

1The author would like to thank participants of the Missouri Economics Conference (U. of Missouri) and of the U. of Missouri Economics brown bag series for helpful comments and suggestions.
2This preliminary draft was prepared for a seminar presentation at The Ohio State University.
1. Introduction

Missing data have plagued researchers across many fields of inquiry and many decades. Both patterns of missingness and missing data mechanisms may be quite complex, and techniques devised for handling such data sets are diverse. Simple techniques may include omission, linear interpolation, and using an ancillary regression to predict missing variables based on correlated covariates. More sophisticated techniques include the EM algorithm formalized by Dempster, Laird, and Rubin (1977), multiple imputation proposed by Rubin (1978), or resampling methods such as the bootstrap, applied to missing data by Efron (1994). Little and Rubin (2002) and Shafer (1997) provide recent texts that thoroughly survey techniques aimed especially at jointly independent and identically distributed (iid) sequences.

Missingness in time series may occur sporadically due to sampling issues, or missingness may be recurring in mixed-frequency data. Because of fundamental dependence of random variables in a time series, the likelihood structure of such sequences is more complicated. On the other hand, the inherent sortability of time series means that not only may covariates be used, but proximate observed data from the same series may also be used to forecast missing data. For example, Friedman (1962) proposed an imputation technique that neatly and simply maps the shape of a covariate onto the missing data, employing data at the endpoints of the missing period in order to anchor the imputed series at both ends to observed data. The chief difficulty faced by Friedman was in finding a good way to estimate the correlation between these series, precisely because of the missing data problem. In the mixed-frequency context, Chow and Lin (1971, 1976) used the observable data in each series to estimate this correlation. The least squares estimator based on only quarterly or yearly data provides an optimal and tractable way to impute the monthly data in these settings. As explored by Harvey and Pierse (1984), Kohn and Ansley (1986), Gomez and Maravall (1994), and others, the Kalman filter provides a powerful way to simultaneously estimate both unobserved data and model parameters. It does not require regular patterns of missingness, and it uses the underlying serial dependence of the data in order to estimate conditional expectations of the missing observations. More recent approaches, such as those taken by Mariano and Murasawa (2003) and Seong et al. (2007), use the Kalman filter to estimate cointegrated models with missing data.

For integrated time series, the issue of cointegration adds additional challenges (inter alia, the possibility of integrated error sequences, if improperly specified), but it opens doors to new opportunities – faster rates of convergence that may overcome biased estimators. The two most well-known of the non-likelihood based approaches are the instrumental variables approach of fully-modified least squares (FM-OLS) proposed by Phillips and Hansen (1990) and constructing a canonical cointegrating regression (CCR), such as that studied by Park (1992).

In this paper, we show how to construct a CCR such that least squares provides a consistent, asymptotically normal (CAN) estimator of the cointegrating vector when regressors contain missing data. Even though estimation requires several steps, each step may be done simply and quickly, and our procedure does not require strict distributional assumptions. Rather than focusing on a particular imputation technique, we place mild restrictions on
the imputation error that necessarily results from any such technique. This error sequence
need not be stationary, and the amount of missing data may be large relative to the sample
size. Moreover, the pattern of missingness in the data may be completely general. Either
sporadic or recurring missingness is permitted. An advantage of abstracting from the
technique is that the only theoretical burden on the researcher lies in verifying that a few
assumptions about the imputation error are satisfied for a given imputation technique. A
disadvantage is that our estimation procedure is sequential. Some imputation techniques,
such as the Kalman filter approaches used in some of the studies mentioned above, explicit-
ly involve maximizing likelihood over the whole system, in which case the missing data
are imputed and the parameters are estimated simultaneously. The statistical properties of
estimators based on such imputation techniques and involving integrated time series are not
well-known, and are beyond the scope of the present analysis. A second disadvantage of our
theoretical analysis is that we derive statistical properties of estimators assuming only a sin-
gle imputation. Simulation-based techniques such as multiple imputation or bootstrapping
could conceivably be used to increase small-sample performance of our estimators.

The rest of the paper is structured as follows. In the following section, we discuss the
theoretical model, and we address the possibilities of singularity and unobservable trends
in the regressors. Most importantly, we outline general assumptions about the imputation
error. If the imputation error from any imputation technique satisfies these conditions,
then our subsequent asymptotic results hold. In Sections 3 and 4, we present the main
theoretical results of the paper. Specifically, we show that all of the parameters may be
estimated consistently in Section 3. We show that the parameter vector of interest may
be estimated consistently and with an asymptotically normal distribution by running least
squares on an appropriately constructed CCR in Section 4. The final section concludes
by summarizing the steps involved in our estimation procedure. Two technical appendices
contain ancillary lemmas and their proofs, along with proofs of the main results of the
paper.

We use the following notational conventions throughout the paper. The vector \( \iota \) denotes
a vector of ones, with a subscript signifying the dimension. Similarly, a single subscript
on \( I \) denotes the dimension of an identity matrix, and single or double subscripts on \( 0 \)
denote a vector or matrix of zeros. Unless otherwise noted, summations are indexed across
\( t = 1, \ldots, n \) and integrals are evaluated over \( s \in [0, 1] \). We use \( \| z \|_p \) to denote the \( L_p \)-norm
\( (E|z|^p)^{1/p} \) of a random variable \( z \) and \( \| Z \|_p \) to denote the \( L_p \)-norm \( (\sum_i \sum_j E|Z_{ij}|^p)^{1/p} \) of a
random vector or matrix \( Z \).

2. Cointegrating Regression With Missing Data

Consider a cointegrating regression given by

\[
y_t = \alpha' w_t + \beta' x_t + v_t,
\]

where \( (x_t) \) is an \( r \)-dimensional I(1) series with missing data, \( (w_t) \) is a \( p \)-dimensional I(0)
series, \( (v_t) \) is a one-dimensional series of unobservable I(0) disturbances with mean zero,
and \( \alpha \) and \( \beta \) are conformable vectors of unknown parameters.
As is well-known, there are essentially eight very different possibilities when a regression contains more than one I(1) regressor. If the regressors are not cointegrated (with each other) or the regressors are cointegrated (with each other) but the cointegrating vector is not $\beta$, then there are three possibilities. If in these cases the regressand is either I(0) or I(1) but not cointegrated with the regressors, then the regression is spurious. On the other hand, if the regressand is I(1) but cointegrated with the regressands and $(1, -\beta)'$ is a cointegrating vector, then the regression is authentic. CCR and FM-OLS were originally designed to deal with the authentic case in which the regressors are not cointegrated with each other, and Phillips (1995) extended FM-OLS to deal with the authentic case in which the regressands are cointegrated with each other. There are two remaining cases. It is possible that the regressors are cointegrated with each other, but that $\beta$ happens to be a cointegrating vector, in which case $(\beta'x_t)$ is a stationary series. If this happens, then having an I(1) regressand clearly creates a spurious relationship, but having an I(0) regressand results in an authentic regression with $\sqrt{n}$-asymptotics, as if all of the series were I(0).

In this paper, we allow the regressors to be cointegrated with each other, but assume that $\beta$ is not a cointegrating relationship. The reasons for allowing cointegrated regressors is so that at least one stochastic trend common to all of the I(1) regressors exists. The importance of this will become clear below. However, we rule out the possibility that $\beta$ is a cointegrating vector, in order to obtain consistent parameter estimates from the faster rate of convergence of an estimator of $\beta$.

To specify the cointegrating relationship among the regressors, we consider $(x_t)$ given by

$$x_t = \mu + \Gamma q_t + u_t,$$

where $(q_t)$ is a $g$-dimensional I(1) series of linearly independent stochastic trends common to $(x_t)$ – and $(y_t)$ – with $1 \leq g \leq r$, $(u_t)$ is an $r$-dimensional I(0) series of unobservable disturbances, and $\mu$ and $\Gamma$ are an $r \times 1$ vector and an $r \times g$ matrix of unknown parameters. Specifically, $(x_t)$ has $r$ dimensions, $r - g$ cointegrating vectors, and $g$ common stochastic trends. Moreover, $(y_t, x_t)'$ has $r + 1$ dimensions and $r - g + 1$ cointegrating vectors (and still $g$ common stochastic trends).

Allowing for a constant term in (1) is possible, but the asymptotics would be somewhat different. Doing so would require first demeaning $(y_t), (x_t)$, and $(u_t)$, so that our asymptotic results would be defined in terms of demeaned Brownian motions, rather than simple Brownian motions. Otherwise, the results would not be fundamentally altered. For expositional simplicity, we instead assume in the theoretical analysis that $(y_t), (x_t)$, and $(u_t)$ have means of zero (or are already demeaned), so that a constant in (1) is unnecessary. We include the constant vector $\mu$ in (2) so that $(q_t)$ need not have a mean of zero. It is possible to consider more complicated deterministic trends in this model. However, the asymptotic theory and estimation technique would be substantially different if the deterministic trends dominated the stochastic trend. We focus explicitly on dominant I(1) stochastic trends in this analysis.

To summarize our assumptions about the I(0) and I(1) components of this model, we define

$$b_t \equiv (v_t, w_t, u_t, \triangle q_t)'$$
and assume throughout the paper that

[A1] \((b_t)\) is a \((1 + p + r + g)\)-dimensional mean-zero series that is stationary, ergodic, and \(\alpha\)-mixing of size \(-a\) with \(a > 1\).

Further, we require of the initial condition of the I(1) series \((q_t)\) that

[A2] Either \(q_0 = 0\) or \(q_0 = \varepsilon_0 O_p(1)\) and independent of \((u_t), (v_t),\) and \((w_t)\)

as is typical in this type of analysis. If the initial value is considered to be stochastic, then independence from the stationary series avoids additional nuisance parameters that add unnecessary complications to the model. Note that we only assume this independence with respect to the initial condition \(q_0\) — not the whole series. If \((q_t)\) is observable, common practice suggests subtracting the initial value from the entire series in order to approximate this condition.

In light of the integratedness of \((q_t)\) and partial sums of the other I(0) series in the model, it is convenient to define a stochastic process

\[
B_n(s) \equiv \frac{1}{\sqrt{n}} \sum_{t=1}^{[ns]} b_t
\]

where \([ns]\) denotes the greatest integer not exceeding \(ns\). Since the series \((b_t)\) is stationary, we further assume that an invariance principle (IP) holds for \(B_n(s)\). Specifically,

[A3] \(B_n(s) \to_d B(s)\),

where

\[
B(s) \equiv (V(s), W(s)', U(s)', Q(s)')'
\]

is a vector Brownian motion with finite long-run variance \(\Omega_{bb}\), such that

[A4] \(\Omega_{bb} > 0\).

The latter assumption precludes degeneracy in any of the variances and covariance of the stationary components of the model. We use \(\Sigma_{bb}\) to denote the contemporaneous variance of \((b_t)\). The one-sided long-run variance \(\Delta_{bb}\) is the sum of the covariances running from 0 to \(\infty\), which is implicitly defined by \(\Omega_{bb} = \Delta_{bb} + \Delta_{bb}' - \Sigma_{bb}\). Double subscripts with these matrices denote submatrices, vectors, or scalars with specific elements of the these variance matrices.

2.2 Singularity

Aside from missing data, this model differs from the model assumed by the standard CCR and FM-OLS approaches, in that we allow for cointegrated regressors. This singularity is synonymous with (asymptotic) collinearity. Setting aside the issue of missing data for the moment, the limiting moment matrix to be inverted in the least squares estimator of \(\beta\) is not
invertible \((\text{unless } g = r)\), since it will be an \(r \times r\) matrix of rank \(g\). Asymptotically negligible terms may allow invertibility in finite samples, but this near collinearity results not only in bloated standard errors (as in the usual iid case), but also in bias and inconsistency. The bias stems from these negligible terms, which are no longer dominated as a result of the collinearity. They have different rates of divergence and may be correlated with many of the error sequences in the model, so the least squares estimator no longer has a sufficiently fast rate of convergence to attain consistency. This may be remedied by choosing any \(g \times r\) matrix \(C\) that does not contain a cointegrating vector of \((x_t)\). We may use \(Cx_t\) (a vector of \(g\) linearly independent regressors) in place of \(x_t\) (a vector of \(r\) linearly dependent regressors) and estimate a \(g \times 1\) vector \(\psi\) in place of \(\beta\). We may then make inferences about \(\beta\) using \(\psi\) and \(C\).

How might we choose \(C\)? If the regressors are not cointegrated, then \(g = r\) and we may choose \(C = I_g\). Otherwise, a natural choice for \(C\) would be a matrix such that \(C\Gamma = I_g\), so that \(\psi\) may be meaningfully interpreted as the vector of marginal effects of the trends themselves. Among this class, a logical choice reflects the relationship between \((q_t)\) and \((x_t)\) and therefore lies in the space spanned by the matrix \(\Gamma\). In particular,

\[
C = \left(\Gamma'\Gamma\right)^{-1}\Gamma',
\]

which is well-defined since \(\Gamma'\Gamma\) is full rank, is a natural choice. The marginal effects of the original regressors will reflect their respective contributions to the underlying trends. Note that this matrix contains no cointegrating vectors of \((x_t)\), since \(\Gamma\) does not cointegrate \((q_t)\).

By choosing \(C\) in this way, estimating \(\psi\) is similar to ignoring the series \((x_t)\) with problematic missing data and substituting \((q_t)\) directly into (1). Doing so yields

\[
y_t = \psi'C\mu + \alpha'w_t + \psi'q_t + \left(v_t + \psi'Cu_t\right),
\]

and applying least squares to this equation provides an alternative way to consistently estimate \(\psi\) that does not require any imputation. However, an estimator of \(\Gamma\) is still necessary to identify the parameter vector of interest \(\beta\) and estimate \(\psi\) efficiently. Using \((q_t)\) in place of \((Cx_t)\) simplifies estimation of \(\psi\), but does not circumvent the fundamental problem. We still need to impute the missing data in \((x_t)\).

If \(C\) is defined in terms of model parameters, as above, then it must be estimated. We assume that such an estimator \(\hat{C}\) satisfies

\[[A5]\quad (\hat{C} - C) = u_g\ell_n O_p\left(n^{-1}\right)\]

More specifically, if \(C\) is chosen as in (3), a natural choice of \(\hat{C}\) would be

\[
\hat{C} = \left(\hat{\Gamma}'\hat{\Gamma}\right)^{-1}\hat{\Gamma}'
\]

for some estimator \(\hat{\Gamma}\). As long as \((\hat{\Gamma} - \Gamma) = o_p\left(n^{-1}\right)\), which we show below for the least squares estimator of \(\Gamma\), \([A5]\) follows from the Slutsky theorem. In this specific case, \([A5]\) is redundant.
2.3 Observable vs. Unobservable Trends

The data generating process described above contains a $g$-dimensional series ($q_t$) of stochastic trends, and our theoretical results are based on increments of this vector of trends. In general, stochastic trends are not directly observable. In applications, however, observable series that are cointegrated with the true trend(s) may be used as proxies. For example, in the panel of city-level cost of living indices with missing data considered by Chang and Rhee (2005), an observable national cost of living index may be employed as a proxy for the common national trend. Similarly, a national stock market index may be used as a proxy for the trend common to individual stock price data with missing observations, such as the historical NYSE data examined by Goetzmann et al. (2000). Additionally, such an approach is natural when using mixed-frequency data, since the high frequency series may be used as proxies for the trends of the low frequency series.

Moreover, many simple and practical imputation techniques are tailored to situations in which such proxies are available. These are precisely the types of problems that the time series imputation techniques of Friedman (1962), Chow and Lin (1971, 1976), Harvey and Pierse (1984) and others are designed to handle – that is, imputation using a “related” time series. In fact, Chow and Lin (1976) roundly criticized a contemporary analysis for assuming that no related time series are observable.

What if Chow and Lin (1976) were wrong, and no proxies are available? It may be possible to use an approach such as those taken by Mariano and Murasawa (2003) or Seong, et al. (2007) to extract the common trends using the Kalman filter. In the presence of missing data, the properties of estimators based on such an approach are not well-known, but if the extracted trends are cointegrated with the true trends, it may be possible to use those trends in place of ($q_t$). In the absence of missing data, Chang et al. (2006) showed that when there is only one trend, the Kalman filter extracts a trend that is cointegrated with the true trend.\textsuperscript{4} They speculated that the Kalman filter might retain its optimal properties if used to extract more than one trend, if ($q_t$) is multidimensional. As the theory governing this particular situation is beyond the scope of the present analysis, we leave these challenges for future research.

2.3 Imputation Error

The heart of this analysis is missingness, as missing data in ($x_t$) is the way in which our analysis and other recent analyses in the literature most fundamentally depart from the bulk of the literature on cointegrated time series.\textsuperscript{5} We do not focus on a particular imputation

\textsuperscript{3}They did not require the related series to be cointegrated, since the notion of cointegration as it is known today did not yet exist. Had they been explicitly dealing with nonstationary data, an assumption that no cointegrated time series are observable would probably not have been as unrealistic as they claimed.

\textsuperscript{4}It is cointegrated up to a scale transformation, since those authors restricted the variance of ($\Delta q_t$). The scale transformation would mean that $\Gamma$ cannot be identified. However, if $C$ is defined in terms of $\Gamma$, as discussed above, $\psi$ would be identified.

\textsuperscript{5}Allowing ($y_t$) to have missing data is also possible. As with measurement error, imputation error causes the most concern when it involves regressors. Imputation error in the regressand may be considered to be part of the error term of the model, in which case we would simply have two imputation errors rather than one in (6) below.
technique, but simply on the imputation error generated by the technique. Let the hybrid series containing both imputed and observed data be denoted by \((x_t^*)\). We assume that this hybrid series may be written as

\[ x_t^* = x_t + z_t^* , \]

where \((z_t^*)\) is an \(r\)-dimensional sequence of imputation error, which is non-zero only when the corresponding datum is imputed. The imputation error may in general be a nonlinear function of other stationary components of the model, such as \((\Delta q_t)\), but interacts with \((x_t)\) linearly. In some well-known examples – see Friedman (1962), e.g. – it may be more appropriate to express \((x_t)\) in logs, in which case (4) could easily be modified. A feasible analog of (2) employs \((x_t^*)\) in place of \((x_t)\).

Specifically, substituting (2) into (4) yields

\[ x_t^* = \mu + \Gamma q_t + u_t^* , \]

where \(u_t^* = u_t + z_t^*\) for notational convenience. Thus defined, we may use \((x_t^*)\) in place of \((x_t)\) in the regression of interest (1), which becomes (also with the introduction of \(C\))

\[ y_t = \alpha' w_t + \psi' C x_t^* + v_t^* \]

where we define \(v_t^* = v_t - \psi' C z_t^*\). The model described by (5) and (6) is somewhat analogous to the classical case of measurement error, since the series of feasible regressors \((x_t^*)\) represents an improperly measured proxy for \((x_t)\) that is explicitly correlated with \((v_t^*)\).

If the imputation error is simply covariance stationary, then the results of Park (1992) or Phillips and Hansen (1992) would follow directly. Variance estimation of the stationary series \((u_t^*)\) and \((v_t^*)\) would replace variance estimation of the stationary series \((u_t)\) and \((v_t)\).

In practice, imputation techniques need not leave behind stationarity of imputation error. Even the simplest technique – linear interpolation – violates stationarity, as the variances and covariances depend explicitly on the length of the missing interval and the distance from the last or next observed \(x_t\).

Realistic imputation techniques generate complicated sequences of imputation error. We may allow for such generality, as long as we keep track of the rates at which the appropriate sample moments diverge. In order to do so, we work within the framework of near-epoch dependence, delineated by the following definition.

**Definition:** Near-epoch Dependence. A sequence \((z_t)\) is near epoch-dependent in \(L_p\)-norm \((L_p\text{-NED})\) of size \(-\lambda\) on a stochastic sequence \((\eta_t)\) if

\[ \left\| z_t - \mathbb{E} \left( z_t | \mathcal{F}_{t+K} \right) \right\|_2 \leq d_t \nu_K \]

where \(\nu_K \to 0\) as \(K \to \infty\) such that \(\nu_K = O \left( K^{-\lambda-\varepsilon} \right)\) for \(\varepsilon > 0\), \(d_t\) is a sequence of positive constants, and \(\mathcal{F}_{t+K}^{t-K}\) is the \(\sigma\)-field defined by \(\sigma (\eta_{t-K}, \ldots, \eta_{t+K})\).

The reader is referred to Davidson (1994), e.g., for more details. The intuition underlying \(L_p\text{-NED}\) is that the random variable \(z_t\) is more dependent on proximate than distant elements of the sequence \((\eta_t)\). This definition allows for much more general patterns of dependence in \((z_t)\) than stationarity or mixing.
In order for our asymptotic results to hold, we must set limitations on the divergence of moments involving \((z_t^*)\) and other series in the model. In particular, we require limit theory to govern the sample moments defined by \(z_t^*, z_t^*z_t'^{-k}, \) and \(z_t^*b_t'^{-k}\). Let \((z_t^*)\) be an element of \((z_t^*)\), and assume that the following hold:

\[\text{[NED1]}\] For each \(i = 1, \ldots, r\), \((z_{it}^*)\) is \(L_2\)-NED of size \(-1\) on \((b_i)\), w.r.t. \((d_{it}^*)\) such that \(d_{it}^* \leq d_t^* \leq \infty\) for \((z_{it}^*)\) \(\neq 0\), and

\[\text{[NED2]}\] \(Ez_t^* = 0\) for all \(t\).

In addition, we define the (possibly time-dependent) covariance matrices

\[\Sigma_{**}^t (k) \equiv Ez_t^* z_t'^{-k} \quad \text{and} \quad \Sigma_{*b}^t (k) \equiv Ez_t^* b_t'^{-k}\]

where, for notational convenience, we simply omit the argument \(k\) when \(k = 0\) henceforth. We assume that for any integer \(k\), covariances \(\Sigma_{**}^t (k)\) and \(\Sigma_{*b}^t (k)\) satisfy

\[\text{[NED3.a]}\] \(\Sigma_{**}^t (k) < \infty\) for all \(t\), \(\Sigma_{**}^t (k) \equiv n^{-1} \sum_{t=k+1}^{n} \Sigma_{**}^t (k) < \infty\), and

\[\text{[NED3.b]}\] \(\Sigma_{*b}^t (k) < \infty\) for all \(t\), \(\Sigma_{*b}^t (k) \equiv n^{-1} \sum_{t=k+1}^{n} \Sigma_{*b}^t (k) < \infty\),

so that even though the covariances may converge to time-dependent spatial averages, the average of each across time is well-defined. With the additional assumption that

\[\text{[NED4]}\] \(\sup_t \|z_{it}^*\|_{4a/(a-1)} < \infty\) for each \(i = 1, \ldots, r\),

we may employ the types of laws of large numbers and central limit theorems for \(L_p\)-NED sequences analyzed by Davidson (1994), Davidson and de Jong (1997), de Jong and Davidson (1997), among others.

We also need to deal with asymptotics for sample moments involving integrated \((q_t)\). Without missing data, such asymptotics are straightforward from the IP and other limiting distributions implied by \([A1]-[A4]\). We need to make additional assumptions about \((z_t^*)\) in order to obtain analogous results for the sample moment defined by \(z_t^* q_t'^{-k}\). To do so, we define an additional stochastic process

\[Z_n^*(s) \equiv \frac{1}{\sqrt{n}} \sum_{t=1}^{[ns]} z_t^*\]

similarly to \(B_n (s)\). Assuming that an IP holds in this case is not as innocuous as in the case of stationary \((b_t)\), so we assume more primitive conditions to obtain the IP. Specifically, we assume that long-run variances satisfy

\[\text{[NED5.a]}\] \(EZ_n^* Z_n'^* (s) \to \Omega_{**} (s) > 0\) with \(\Omega_{**} (s) < \infty\), and

\[\text{[NED5.b]}\] \(EZ_n^* Q_n' (s) \to \Omega_{*q} (s) \) with \(\Omega_{*q} < \infty\).
Such assumptions strengthen [NED3], in the sense that they require finite long-run variances, as well as contemporaneous variances. In essence, these act like summability conditions on the autocovariances of the respective limits in [NED3]. They also allow us to employ some key results on IP’s presented by Davidson (1994).

Finally, we assume that

\[ \text{[NED6]} \quad E u_t^* w_{t-k} = \iota_r \iota'_p 0 \text{ for all } k. \]

The prohibition of contemporaneous and serial correlation between the stationary regressors and the imputation error is not strictly required to obtain consistent parameter estimates. However, it is necessary to obtain \( \sqrt{n} \)-convergence for the coefficients \( \alpha \) of these regressors, which we use to estimate the covariances consistently. Moreover, even if these are estimated consistently, relaxing this assumption would require a similar but different procedure to estimate \( \psi \) with asymptotic normality.

The limit theory made accessible by assumptions [NED1]-[NED6] is collected in Appendix A of this paper.

3. Consistent Estimation

We turn our attention to consistent estimation of the parameters described by the feasible system (5) and (6), assuming that \((q_t)\) is observable or that we have an observable series cointegrated with \((q_t)\). All of our theoretical results are built upon least squares estimation of (5) and (6).

3.1 Consistent Estimation of \( \psi \)

Under general conditions, least squares estimation of (6) still provides a consistent (but not asymptotically normal and unbiased) estimator of \( \psi \) – and therefore of \( \beta \). Using a partitioned regression, it is easy to see that the least squares estimator of \( \psi \) is

\[ \hat{\psi}_{LS} = (N^*_n)^{-1} M^*_n \]

with

\[ M^*_n = \hat{C} \sum x_t^* v_t^* - \hat{C} \sum x_t^* w_t' \left( \sum w_t w_t' \right)^{-1} \sum w_t v_t^* \]

and

\[ N^*_n = \hat{C} \sum x_t^* x_t^* \hat{C}' - \hat{C} \sum x_t^* w_t' \left( \sum w_t w_t' \right)^{-1} \sum w_t x_t^* \hat{C}', \]

since this estimator first projects orthogonally to the space of the stationary regressors \((w_t)\), and then projects \((y_t)\) onto the space of these residuals. The exact asymptotic distribution of the estimator is presented in the following theorem.
**Theorem 3.1** Assume that [A1]-[A5], and [NED1]-[NED6] hold. Define

\[ N(s) \equiv CT \left( \int QQ' \right) \Gamma' C' \]

and

\[ M^*(s) \equiv \Gamma \left( \int Qd(V(s) - \psi' CZ^*(s)) + \delta v - \Delta q s C' \psi \right) \]

\[ + (\sigma_{uv} - \Sigma u s C' \psi) + (\sigma_{v} - \Sigma v s C' \psi) . \]

The least squares estimator \( \hat{\psi}_{LS} \) has a distribution given by

\[ n \left( \hat{\psi}_{LS} - \psi \right) \rightarrow_d N(s)^{-1} M^*(s) \]

as \( n \rightarrow \infty \).

Due to the fast rate of convergence \( n \) of the estimator to its asymptotic distribution, \( \sqrt{n} \)-consistency is obtained, in spite of numerous nuisance parameters. Moreover, choosing \( \hat{C} \) (naively) to be an \( r \times 1 \) vector of ones – as long as this is not a cointegrating vector of \( (x_t) \) – would also yield a consistent estimator of \( \psi \) for that particular linear combination.

**3.2 Consistent Estimation of \( \Gamma, \alpha, \) and \( \mu \)**

The asymptotic distribution above critically depends on \( \Gamma \), so that constructing a feasible CCR requires a consistent estimator of \( \Gamma \). Moreover, if \( C \) is defined in terms of \( \Gamma \), we need a consistent estimator of \( \Gamma \) before estimating \( \psi \). Least squares estimation of (5) is accomplished by way of the least squares estimator defined by

\[ (\hat{\Gamma}_{LS} - \Gamma) = \sum u_t^* (q_t - \bar{q}_n) \left( \sum q_t (q_t - \bar{q}_n) \right)^{-1} \]

where \( \bar{q}_n \equiv n^{-1} \sum q_t \) is just the sample mean of \( (q_t) \). The following lemma establishes consistency of this estimator.

**Lemma 3.2** Assume that [A1]-[A4], [NED1]-[NED3.a], [NED4], and [NED5.a]-[NED5.b] hold. We have

\[ (\hat{\Gamma}_{LS} - \Gamma) = \omega_{r \times \omega} \text{Op}(n^{-1}) \]

as \( n \rightarrow \infty \).

In addition to \( \Gamma \), we also require consistent estimators of \( \alpha \) and \( \mu \) in order to estimate covariances consistently. We used a partitioned regression to analyze \( \psi \) and \( \Gamma \). With consistent estimators of these parameters in hand, we can use the next step of a partitioned regression to analyze \( \alpha \) and \( \mu \). The least squares estimators are given by

\[ \hat{\alpha}_{LS} = \left( \sum w_t w_t' \right)^{-1} \sum w_t \left( y_t - \hat{\psi}_{LS} C x_t^* \right) \]

\[ \hat{\mu}_{LS} = \left( \sum x_t x_t' \right)^{-1} \sum x_t \left( y_t - \hat{\psi}_{LS} C x_t^* \right) \]

\[ \hat{\Gamma}_{LS} = \left( \sum x_t x_t' \right)^{-1} \sum x_t \left( x_t - \hat{\psi}_{LS} C x_t^* \right) \]
for \( \alpha \) and
\[
\hat{\mu}_{LS} = \frac{1}{n} \sum (x_i^* - \hat{\Gamma}_{LS} q_t)
\]
for \( \mu \). Least squares consistently estimates these, too, as we show in the following lemma.

**Lemma 3.3**  Assume that [A1]-[A5], and [NED1]-[NED6]. We have

[a] \((\hat{\alpha}_{LS} - \alpha) = \nu_p O_p (n^{-1/2})\), and
[b] \((\hat{\mu}_{LS} - \mu) = \nu_r O_r (n^{-1/2})\)

as \( n \to \infty \).

Theorem 3.1 and Lemma 3.3[a] jointly establish consistency with appropriate rates of convergence for least squares estimation of (6), while Lemma 3.2 and Lemma 3.3[b] accomplish the same for (5).

In practice, a problem may arise for some imputation techniques. We need to impute missing data before running least squares, but we may need to estimate \( \Gamma \) and/or \( \mu \) in order to impute the missing data, if the technique employs the trends \( (q_t) \). In order to circumvent this difficulty, it may be necessary to use a preliminary (but not necessarily consistent) estimator of \( \Gamma \) and/or \( \mu \). If the resulting imputation error satisfies our conditions above, then these parameters may then be re-estimated consistently.

### 3.3 Consistent Covariance Estimation

In addition to the asymptotics given by the preceding lemmas, we need to estimate the appropriate long-run variances consistently. To this end, we define

\[
b_t^* \equiv b_t + D z_t^*
\]

where \( D \) is a \((1 + p + r + g) \times r\) matrix defined by

\[
D \equiv \left( -C' \psi, 0_{rp}, I_r, 0_{rg} \right)'.
\]

We need to estimate the long-run variance of the vector \((b_t^*)\) consistently. The other long-run variances and covariances that we must estimate are submatrices of this one. Moreover, the contemporaneous or one-sided long-run variances and covariances are special cases. We first verify consistent covariance estimation when \((b_t)\) and \((z_t^*)\) are observable. In this case, the long-run variance estimator is

\[
\tilde{\Omega}_{b^*b^*} = \frac{1}{n} \sum_{t=1}^{n} \sum_{s=1}^{n} b_t^* b_s^{*'} \pi \left( \frac{t - s}{h_n} \right)
\]

for some kernel function \( \pi \) with lag truncation parameter \( h_n \). In the absence of imputation error, a vast literature on covariance estimation is available. Papers by Newey and West (1987), Andrews (1991), and Hansen (1992), *inter alia*, address this problem under stationarity and/or mixing assumptions.
While earlier work on near-epoch dependence, such as that done by Davidson (1994), addresses covariance estimation, the theory of de Jong and Davidson (2000) provides a particularly nice framework, as it does not require sequences to be adaptable. Although adaptability is warranted in most time series contexts, it is too strict to deal with imputation error. Many imputation techniques use future as well as past information.

In addition to our previous assumptions, we also assume that the kernel function and lag truncation parameter satisfy

\[ K1 \lim_{n \to \infty} \left( h_n^{-1} + n^{-1}h_n \right) = 0, \]

\[ K2 \pi \text{ belongs to the function class defined by de Jong and Davidson (2000), and} \]

\[ K3 \sqrt[n]{\sum_{k=0}^{n} \pi\left( \frac{k}{h_n} \right)} = o\left(1\right). \]

The first condition imposes \( h_n = o(n) \) on the lag truncation parameter. The second limits the class of admissible kernel functions, which includes many well-known kernels, such as Bartlett, Parzen, quadratic spectral, and Tukey-Hanning kernels. The reader is referred to de Jong and Davidson (2000) for more details. We employ the third assumption for feasible estimators of the variances in the model. This may impose additional restrictions on \( h_n \), depending on the kernel function. For example, the Bartlett and Tukey-Hanning kernels would require \( h_n = o\left(n^{1/2}\right) \) to satisfy \( K3 \).

With these additional assumptions, we have the following result.

**Lemma 3.4** Assume that \( [A1], [NED1], [NED4], [K1], \) and \( [K2] \) hold. We have

\[ a \sum_{b} b^{*} \to_p \Sigma_{bb}^{*} \]

\[ b \Delta_{b}^{*} \to_p \Delta_{bb}^{*} \]

\[ c \Omega_{b}^{*} \to_p \Omega_{bb}^{*} \]

as \( n \to \infty \), where

\[ \Sigma_{bb}^{*} \equiv \Sigma_{bb} + \Sigma_{bs}D + D\Sigma_{sb} + D\Sigma_{ss}D, \]

with \( \Delta_{bb}^{*} \) and \( \Omega_{bb}^{*} \) defined accordingly.

The long-run variance estimator is consistent if the I(0) series driving the model and the imputation error are all known.

However, only \((w_t)\) and \((\triangle q_t)\) are known. The error sequences are obviously unknown, and the imputation error will in general be unknown. The unknown sequences \((v_t^*)\) and \((u_t^*)\) must be estimated. (Note that in general estimates of \((v_t)\), \((u_t)\), and \((z_t^*)\) cannot be identified.) Simple feasible estimators \((v_t^*)\) and \((u_t^*)\) are given by

\[ \hat{v}_t^* = y_t - \hat{\alpha}'w_t - \hat{\psi}'\hat{C}x_t^* \]

and

\[ \hat{u}_t^* = x_t^* - \hat{\mu} - \hat{\Gamma}q_t \]

for estimators \( \hat{\alpha}, \hat{\psi}, \hat{C}, \hat{\mu}, \) and \( \hat{\Gamma} \). Alternatively, these may be written as

\[ \hat{v}_t^* = \tilde{v}_t - \hat{\psi}'\hat{C}\tilde{z}_t^* \]

and

\[ \hat{u}_t^* = \tilde{u}_t + \tilde{z}_t^*, \]
where
\[ \hat{v}_t \equiv v_t + (\alpha - \hat{\alpha})'w_t + (\psi' \hat{C} - \hat{\psi}' \hat{C})x_t \quad \text{and} \quad \hat{u}_t \equiv u_t + (\mu - \hat{\mu}) + (\Gamma - \hat{\Gamma})q_t, \] (11)
The series \((\hat{b}_t^*)\) may thus be defined by
\[ \hat{b}_t^* \equiv \hat{b}_t + \hat{D}z_t^* \] (12)
where naturally \(\hat{w}_t \equiv w_t\), \(\triangle \hat{q}_t \equiv \triangle q_t\), and \(\hat{D}\) is defined by replacing \(C\) and \(\psi\) in \(D\) with suitable estimators. Define the feasible estimator
\[ \hat{\Omega}_{b^*b^*} \equiv \frac{1}{n} \sum_{t=1}^{n} \sum_{s=1}^{n} \hat{b}_t^* \hat{b}_s^* \pi \left( \frac{t-s}{h_n} \right) \] (13)
which may be used to estimate the long-run variance in practice. Using a change of variables \(k = t - s\), the symmetry of the kernel function, and \(\pi(0) = 1\), we may also write this as
\[ \hat{\Omega}_{b^*b^*} = \hat{\Delta}_{b^*b^*} + \hat{\Sigma}_{b^*b^*} \]
where
\[ \hat{\Delta}_{b^*b^*} \equiv \frac{1}{n} \sum_{k=0}^{n} \pi \left( \frac{k}{h_n} \right) \sum_{t=k+1}^{n} \hat{b}_t^* \hat{b}_{t-k}^* \quad \text{and} \quad \hat{\Sigma}_{b^*b^*} \equiv \frac{1}{n} \sum \hat{b}_t^* \hat{b}_s^* \] (14)
are feasible estimators used to estimate the one-sided long-run and contemporaneous variances, respectively.

**Lemma 3.5** Assume that \([A1]-[A5],[NED1]-[NED6],[K1]-[K3]\) hold, and consider estimators \(\hat{\alpha}, \hat{\psi}, \hat{\mu}, \) and \(\hat{\Gamma}\) defined by the least squares estimators. We have

\[ [a] \ \hat{\Sigma}_{b^*b^*} \rightarrow_p \Sigma_{b^*b^*}, \]
\[ [b] \ \hat{\Delta}_{b^*b^*} \rightarrow_p \Delta_{b^*b^*}, \] and
\[ [c] \ \hat{\Omega}_{b^*b^*} \rightarrow_p \Omega_{b^*b^*} \]
as \(n \rightarrow \infty\).
Consequently, the natural feasible estimators described by (13) and (14) are consistent.

With consistent estimators of all of the requisite parameters in hand, we now turn to asymptotically normal estimation of the parameter vector of interest.

### 4. CAN Estimation of \(\beta\)

For a prototypical cointegrated system, the most common choices for estimation are CCR, FM-OLS, or maximum likelihood. The first two are more attractive in this setting, because of the linearity of the model but otherwise complicated nature of the (imputed) error terms. It is well-known that both CCR and FM-OLS provide CAN estimators under standard assumptions. Aside from the missing data, our model has at least one fundamental difference
from that considered by Phillips and Hansen (1990) and Park (1992). We allow the regressors to be cointegrated and construct the model explicitly in terms of the stochastic trends \((q_t)\). The matrix \(C\) reduces this problem essentially to their model. The theoretical obstacle imposed by imputation is not trivial, since the asymptotics of these approaches were designed to accommodate stationary error terms.

Nevertheless, we show how to construct a CCR to account for this. In order to show CAN, we momentarily ignore the second problem (imputation), in order to find an appropriate transformation of the data under otherwise ideal conditions. Once we have an ideal estimator, we will deal with the imputation explicitly in order to find a feasible estimator.

4.1 Infeasible CAN Estimation

Let

\[
\Delta q_\bullet \equiv (\delta_{qv}, \Delta_{qw}, \Delta_{qu}, \Delta_{qq}),
\]

which is the \(g \times (1 + p + r + g)\) matrix formed by the rows corresponding to \((q_t)\) (the last \(g\) rows) in the one-sided long-run variance of \((b_t)\), and may be interpreted as the one-sided long-run covariance between \((q_t)\) and \((b_t)\). Similarly, let

\[
\Sigma u_\bullet \equiv (\sigma_{uv}, \Sigma_{uw}, \Sigma_{uu}, \Sigma_{uq})
\]

be the \(r \times (1 + p + r + g)\) matrix representing the contemporaneous covariance between \((u_t)\) and \((b_t)\). Define \(\kappa\) to be a \((1 + p + r + g)\times 1\) vector given by

\[
\kappa \equiv (1, 0_{1p}, 0_{1r}, -\omega_{vq} \Omega_{qq}^{-1})'.
\]

Now let

\[
x_t^{**} \equiv x_t - (\Gamma \Delta q_\bullet + \Sigma u_\bullet) \Sigma^{-1} b_t
\]

and

\[
y_t^{**} \equiv y_t - \psi' C (\Gamma \Delta q_\bullet + \Sigma u_\bullet) \Sigma^{-1} b_t - \omega_{vq} \Omega_{qq}^{-1} \Delta q_t,
\]

so that we estimate

\[
y_t^{**} = \alpha' w_t + \psi' C x_t^{**} + v_t^{**}
\]

in place of (1), where \((v_t^{**})\) is defined as

\[
v_t^{**} = b_t' \kappa
\]

implicitly.

In order to analyze the least squares estimator of the CCR above, we cannot use the first step of a partitioned regression, as we did to show consistency of \(\hat{\psi}_{LS}\). This is because in creating \((v_t^{**})\) we add a stationary series to the error sequence. This series may be correlated with \((w_t)\), even if we maintain \(\text{NED6}\), the assumption that \((w_t)\) is uncorrelated with the original error sequence. Estimating the above equation directly would bias our estimator in a fundamental way. Since we already have a consistent estimator of \(\alpha\), we may instead
regress the fitted residuals using $\hat{\alpha}$ onto the space of $(x_t^{**})$. This amounts to simply running least squares on

$$y_t^{**} - \hat{\alpha}' w_t = \psi' \hat{C} x_t^{**} + v_t^{**}$$

instead. It is not difficult to show that when $(x_t)$ contains no missing data and the true values of all nuisance parameters are known, running least squares on (15) provides a CAN estimator of $\psi$.

### 4.2 Feasible CAN Estimation

It should be clear that we are dealing with variance estimators that capture both idiosyncratic error and imputation error. All of the variances estimators are defined as submatrices, vectors, or individual elements of $\hat{\Sigma}_{b^*b^*}$, $\hat{\Delta}_{b^*b^*}$, and $\hat{\Omega}_{b^*b^*}$. For notational convenience, we drop the $*$ superscripts in the subscripts of the variance estimators, and the $b^*$ subscripts entirely. For example, $\hat{\Sigma}$ denotes $\hat{\Sigma}_{b^*b^*}$ and $\hat{\omega}_{vq}$ denotes $\hat{\omega}_{vq}$. Note that the probability limit of these feasible estimators are not $\Sigma_{bb}$ and $\omega_{vq}$. Rather the limiting variances are $\Sigma_{b^*b^*}$ and $\omega_{vq} - \psi' \Omega_{vq}$.

Replacing all of the parameters in $(x_t^{**})$ and $(y_t^{**})$ with feasible consistent estimators and using the imputed series $(x_t^*)$, necessitates redefining $(x_t^{**})$ and $(y_t^{**})$ as

$$x_t^{**} = x_t^* - (\hat{\Gamma} \hat{\Delta}_{q^*} + \hat{\Sigma}_{u^*}) \hat{\Sigma}^{-1} \hat{b}_t^*$$

and

$$y_t^{**} = y_t^* - \psi' \hat{C}(\hat{\Gamma} \hat{\Delta}_{q^*} + \hat{\Sigma}_{u^*}) \hat{\Sigma}^{-1} \hat{b}_t^* - \hat{\omega}_{vq} \Omega_{qq}^{-1} \Delta q_t,$$

which is not difficult to estimate, since all variances and covariances can be estimated simultaneously. Now, the error term is simply

$$v_t^{**} = b_t^* \kappa$$

in (15), and

$$\hat{\psi}_{CCR} = \left( \hat{C} \sum x_t^{**} x_t^{**'} \hat{C}' \right)^{-1} \hat{C} \sum x_t^{**} (y_t^{**} - \hat{\alpha}' w_t)$$

is a feasible CCR estimator.

Define Brownian motion

$$V_{\bot Q} (s) \equiv (V (s) - \psi' C Z^* (s)) - (\omega_{vq} - \psi' C \Omega_{vq}) \Omega_{qq}^{-1} Q (s),$$

which may be interpreted as the projection of $V (s) - \psi' C Z^* (s)$ orthogonally to $Q (s)$. This has a long-run variance of

$$\text{lrvar} (V_{\bot Q} (s)) = (\omega_{vv} - \omega_{vq} \Omega_{qq}^{-1} \omega_{qv}) + \psi' C (\Omega_{ss} - \Omega_{sv} \Omega_{qq}^{-1} \Omega_{qs}) C' \psi$$

$$- \psi' C (\omega_{sv} - \Omega_{sv} \Omega_{qq}^{-1} \omega_{qv}) - (\omega_{vs} - \omega_{vq} \Omega_{qq}^{-1} \Omega_{qs}) C' \psi$$

where the first term is the long-run variance in the standard CCR model. Using this notation, we show in the following theorem that $(\hat{\psi}_{CCR} - \psi)$ has ideal properties and rate-$n$ convergence.
**Theorem 4.1** Assume that \([A1]-[A5], [NED1]-[NED6], [K1]-[K3]\) hold, and consider estimators \(\hat{\alpha}, \hat{\psi}, \hat{\mu}, \text{ and } \hat{\Gamma}\) defined by the least squares estimators. We have

\[
n(\hat{\psi}_{CCR} - \psi) \to_d N(s)^{-1} C \int QdV^\ast_{\perp Q}(s)
\]
as \(n \to \infty\).

Moreover, since

\[
\text{var} \left( C \int QdV^\ast_{\perp Q}(s) \right) = \text{lrvar} \left( V^\ast_{\perp Q}(s) \right) N(s),
\]
the asymptotic distribution may be rewritten as

\[
n(\hat{\psi}_{CCR} - \psi) \to_d \sqrt{\text{lrvar} \left( V^\ast_{\perp Q}(s) \right)} N(s)^{-1/2} N(0, I_r),
\]

where \(N(s)^{-1/2}\) is the Cholesky decomposition of the inverse of \(N(s)\). This distribution is simply a \(\ast \times 1\) vector of mixed normal variates. Note that this variance has the standard least squares form, so that standard errors and test statistics from any software package are asymptotically valid. Since \(\hat{C} \to_p C\), we may recover the asymptotic distribution for \(\hat{\beta}_{CCR} = \hat{C}'\hat{\psi}_{CCR}\), which may be written as

\[
n(\hat{\beta}_{CCR} - \beta) \to_d \sqrt{\text{lrvar} \left( V^\ast_{\perp Q}(s) \right)} \left( C' N(s)^{-1} C \right)^{1/2} N(0, I_r),
\]
an \(\ast \times 1\) vector of mixed normal variates. Estimates and standard errors from software packages must be transformed accordingly. If \(C\) is chosen to be \(C = (I'\Gamma)^{-1} I'\), this distribution is

\[
\sqrt{\text{lrvar} \left( V^\ast_{\perp Q}(s) \right)} \left( C' \left( \int QQ' \right)^{-1} C \right)^{1/2} N(0, I_r),
\]
so that the standard errors are proportional to the contributions of each element of \((x_t)\) to their common stochastic trends.

### 5. Summary and Outlook

The procedure we detail in this paper for estimating \(\beta\) in the cointegrated system given by (1) and (2) can be briefly summarized in six straightforward steps.

1. **Create** \((x_t^\ast)\). Impute the missing data in the regressors \((x_t)\) using an imputation technique satisfying assumptions \([NED1]-[NED6]\). If this is accomplished using the stochastic trends \((q_t)\), a preliminary imputation using inconsistent estimates of \(\Gamma\) and \(\mu\) may be necessary. As long as the preliminary imputation satisfies \([NED1]-[NED6]\), these may be re-estimated. If the trends \((q_t)\) are unobservable, it may be possible to estimate them consistently. See discussion above.
Estimate $\alpha$, $\psi$, $\mu$, and $\Gamma$. Least squares on (5) provides consistent estimates of $\mu$ and $\Gamma$. If the regressors $(x_t)$ are not cointegrated with each other, let $\hat{C} = I_g$. Otherwise, let $\hat{C} = (\hat{\Gamma}'\hat{\Gamma})^{-1}\hat{\Gamma}'$. Least squares on (6) provides consistent estimates of $\alpha$ and $\psi$.

Estimate $\Omega$, $\Sigma$, and $\Delta$. Use observable $(w_t)$ and $(\Delta q_t)$, and sample analogs $(\hat{v}_t^*)$ and $(\hat{u}_t^*)$ of unobservable $(v_t)$ and $(u_t)$ to get consistent estimates. Long-run variance estimation must use a kernel function and lag truncation parameter satisfying [K1]-[K3].

Create $(x_t^{**})$ and $(y_t^{**})$. Use (16) and (17) and consistent parameter estimates.

Re-Estimate $\psi$. Least squares on (15) provides consistent and asymptotically normal estimates.

Estimate $\beta$. Use $\hat{\beta}_{CCR} = \hat{C}'\hat{\psi}_{CCR}$ and $\text{Var}(\hat{\beta}_{CCR}) = \hat{C}'\text{Var}(\hat{\psi}_{CCR})\hat{C}$ to get standard errors.

The missing data problem adds only one step, the imputation step. All of the other steps are required for estimating a standard cointegrated model – with cointegrated regressors and well-identified parameters – using a CCR.

Unanswered questions present multiples avenues for future research. Unobservable trends and dominant deterministic trends, which we discussed briefly in Section 2, present challenges that remain. We expect that these would require different approaches and different theory, and we leave these for future research.
References


Appendix A: Useful Lemmas and Their Proofs

Lemma A.1  For sequences \((b_t)\) satisfying [A1] and \((z_t^*)\) satisfying [NED1] and [NED4],

[a] \((z_{t,k}^*, z_{t,k}^{**})\) is a matrix of \(L_1\)-NED sequences of size \(-1\) on \((b_t)\) w.r.t. bounded sequences of constants, and

\[ \{ z_{t,k}^* b_{t-k} \} \text{ is a matrix of } L_1-\text{NED sequences of size } -1 \text{ on } (b_t) \text{ w.r.t. bounded sequences of constants.} \]

[c] \((b_t', z_t^{**})'\) is a vector of \(L_2\)-NED sequences of size \(-1\) on \((b_t)\) w.r.t. bounded sequences of constants.

Proof of Lemma A.1  Under [NED1], each element \((z_{t,k}^*)\) of \((z_t)\) is \(L_2\)-NED on \((b_t)\), which means that \((z_{t,k}^*, z_{t,k}^{**})\) is also \(L_2\)-NED (with the same size) on \((b_t)\) w.r.t. constants \(2d_{t,k}^2\), from the proof of Theorem 17.10 of Davidson (1994). Since both of these are \(L_2\)-NED, an arbitrary element \((z_{t,k}^*, z_{t,k}^{**})\) of \(L_1\)-NED (with the same size) w.r.t. constants defined by

\[ \max(\|z_{t,k}^*, z_{t,k}^{**}\|_2, \|z_{t,k}^*, z_{t,k}^{**}\|_2, \|z_{t,k}^*, z_{t,k}^{**}\|_2, \|z_{t,k}^*, z_{t,k}^{**}\|_2, \|z_{t,k}^*, z_{t,k}^{**}\|_2, \|z_{t,k}^*, z_{t,k}^{**}\|_2, \|z_{t,k}^*, z_{t,k}^{**}\|_2) \]

from the proof of Theorem 17.9 of Davidson (1994). Since this sequence is bounded by [NED1] and [NED4], the proof of part [a] is complete.

Part [b] of the lemma follows in the same way by noting that \((b_t)\) is \(L_2\)-NED on itself w.r.t. constants that are bounded by the covariance stationarity of \((b_t)\). In particular, a Wold representation may be used to show that each element of \((b_t)\) can be decomposed into a stationary linear filter (with summable coefficients) and a deterministic sequence. The deterministic sequence is eliminated when taking the difference in the definition of \(L_2\)-NED, and the rest follows along the lines of Example 17.3 in Davidson (1994), with the difference that we define the filtration in the definition to be the natural filtration. The rest of the proof of part [b] is similar to that of part [a].

The proof of part [c] is trivial and therefore omitted.

\[\square\]

Lemma A.2  Let [NED1] and [NED4] hold. Then

[a] \(n^{-1} \sum_{t=1}^n z_t^* = \iota_r o_p (1)\), under [NED2],

[b] \(n^{-1} \sum_{t=1}^n z_t^* z_{t-k}^{**} - \Sigma_{ss} (k) \equiv \iota_r o_p (1)\), under [NED3.a],

[c] \(n^{-1} \sum_{t=1}^n z_t^* b_{t-k}^{**} - \Sigma_{sb} (k) \equiv \iota_r o_p (1)\), under [A1] and [NED3.b],

[d] \(n^{-1/2} \sum_{t=1}^n z_t^* = \iota_r O_p (1)\), under [NED3.a] and [NED5.a],

[e] \(n^{-1/2} \sum_{t=1}^n z_t^* w_t' = \iota_r O_p (1)\), under [A1], [NED3.a], [NED5.a], and [NED6],

[f] \(n^{-1} \sum_{t=1}^n z_t^* q_{t-k} \rightarrow_d \Xi_{sq} (k)\), where \(\Xi_{sq} (k) = O_p (1)\) and \(\Xi_{qs} \equiv \int QdZ^* (s) + \Delta_{qs}\)

for \(k = 0\), under [A1]-[A4], [NED2], and [NED5.a]-[NED5.b],

as \(n \rightarrow \infty\).
Proof of Lemma A.2 To prove parts [a]-[c], we verify the conditions for a law of large numbers proven by Davidson and de Jong (1997). For parts [d] and [e], we take a similar approach using a central limit theorem of de Jong (1997). Subsequently, we use a theorem from Davidson (1994) based on an IP to prove part [f]. For parts [d]-[f], all stochastic arrays are created by dividing the underlying stochastic sequences by \( \sqrt{n} \).

[a] Consider scalar-valued imputation error. Clearly, \( d_2 = O(\|z^*_t\|_2) \), since both sided are bounded under [NED1] and [NED4]. In order to obtain the rate of convergence in our stated result, a sufficient condition for Theorem 3.3 of Davidson and de Jong (1997) is

\[
t^{-1}\|z^*_t\|_2 = O(t^{-5/6-\varepsilon})
\]

for some \( \varepsilon > 0 \), since \( (z^*_t) \) has mean zero by [NED2]. Under [NED4], this condition holds for any \( \varepsilon \leq 1/6 \). The result trivially extends to encompass vector-valued \( (z^*_t) \).

[b] Consider an arbitrary element \( (z^*_t z^*_j t-k) \) of \( (z^*_t z^*_j t-k) \), which is \( L_1 \)-NED of size \( -1 \) on \( (b_t) \) w.r.t. a bounded sequence of constants by Lemma A1[a]. Note that

\[
\|z^*_t z^*_j t-k - [\Sigma_{xx}(k)]_{ij}\|_2 \leq \|z^*_t z^*_j t-k\|_4 \|z^*_j t-k\|_4 + \|[\Sigma_{xx}(k)]_{ij}\|
\]

by the Minkowski and Cauchy-Schwarz inequalities, where \( [\Sigma_{xx}(k)]_{ij} \) is simply the \( ij^{th} \) element of \( \Sigma_{xx}(k) \). Since \( \|z^*_t z^*_j t-k\|_4 \) is bounded for any \( t \) by [NED4], so is \( \|z^*_j t-k\|_4 \). Moreover, \( \|[\Sigma_{xx}(k)]_{ij}\| \) is bounded by [NED3.a], which is sufficient to employ Theorem 3.3 of Davidson and de Jong (1997). Again, the scalar result trivially extends to the whole matrix.

[c] The proof for part [c] is identical to that for part [b], with the replacement of Lemma A1[a] with A1[b] and of [NED3.a] with [NED3.b].

[d] Consider a random vector \( \Omega_{xx}^{-1/2} z^*_t / \sqrt{n} \) constructed from \( (z^*_t) \) and the Cholesky decomposition of the inverse of \( \Omega_{xx} \), which is invertible by [NED5.a]. By defining an \( r \times 1 \) vector \( e_i \) that has a one in the \( i^{th} \) row and zeros elsewhere, the \( i^{th} \) element of this random vector is \( e_i^t \Omega_{xx}^{-1/2} z^*_t / \sqrt{n} \). If the \( L_2 \)-norm of this element is unity, the conditions for Theorem 2 of de Jong (1997) are satisfied for constants \( (c_{nt}) = n^{-1/2} \). For verification, note that under [NED3.a] and [NED5.a] (with \( s = 1 \))

\[
E \left| \sum e_i^t \Omega_{xx}^{-1/2} z^*_t / \sqrt{n} \right|^2 = e_i^t \Omega_{xx}^{-1/2} \Omega_{xx} \Omega_{xx}^{-1/2} e_i,
\]

which is in fact unity for each \( i \).

[e] Part [e] is a special case and the only part that relies on [NED6], under which \( \Sigma_{sw} \) is a matrix of zeros. The proof follows is similar to that of part [d], and is therefore omitted.

[f] We look explicitly at the case in which both \( (z^*_t) \) and \( (q_t) \) are scalars, which may be generalized to the matrix case. We show that the conditions for Theorem 30.13 of Davidson (1994) are satisfied under our assumptions. This theorem requires sufficient conditions from Theorem 29.6 and Corollary 29.14 of Davidson (1994) to hold. Conditions [NED1] and [NED2] directly satisfy conditions [a], [c], and [e] of Theorem 29.6. Condition [d] of this theorem is also

\footnote{The number \(-5/6\) comes from applying the formula in Davidson and de Jong (1997) with \( q = 2 \), \( b = 1 \), and \( a > 1 \), which is appropriate for either \( L^2 \) or \( L^1 \)-NED sequences of size \(-1\) defined on mixing sequences with size \(-a \) where \( a > 1 \).}
satisfied by the bound of $d_t^2$ in [NED1]. Conditions [A1]-[A4], [NED5.a] and [NED5.b] jointly satisfy condition [f'] of Corollary 29.14, because in order for $E(Q(s), Z^*(s))'(Q(s), Z^*(s))$ to have a finite limit, $EQ^2(s), E Z^* Q(s)$, and $E Z^*^2(s)$ must have finite limits. All that remains to show is that condition [b] of Theorem 29.6,

$$\sup_t \|z_t^*/d_t^2\|_{2n/(a-1)} < \infty,$$

is also satisfied. Since $d_t^2 > 0$ for all $t$, it may be pulled out of the norm, and since [NED4] bounds $\|z_t^*\|_{2n/(a-1)}$, condition [b] is satisfied. Theorem 30.13 of Davidson (1994) yields the stochastically bounded result for any $k$, as well as the specific distributional result for $k = 0$. These results may be generalized to the case in which $(z_t^*)$ and $(q_t)$ are vectors, along the lines of Theorem 29.18 of Davidson (1994). □

Lemma A.3 Assume that [A1]-[A4] and [NED1]-[NED5] hold. We have

[a] $\frac{1}{n} \sum x_t^* v_t^* \rightarrow_d \Gamma \left( \int Qd(V - \psi' CZ^*(s)) + \delta_{qv} - \Delta_{qV} C' \psi \right) + (\sigma_{uv} - \Sigma_{ux} C' \psi) + (\sigma_{sv} - \Sigma_{sx} C' \psi)$,

[b] $\frac{1}{n} \sum x_t^* w_t^* \rightarrow_d \Gamma \left( \int Qdw + \Delta_{qw} \right) + \Sigma_{uw} + \Sigma_{sw}$,

[c] $\frac{1}{n} \sum w_t v_t^* \rightarrow_p \sigma_{uv} - \Sigma_{uw} C' \psi$,

[d] $\frac{1}{n^2} \sum x_t^* x_t'' \rightarrow_d \Gamma \left( \int QQ' \right) \Gamma'$

as $n \rightarrow \infty$.

Proof of Lemma A.3
[a] We may rewrite the summation in terms of $(x_t), (v_t)$, and $(z_t^*)$ using (4) and our definition of $(v_t^*)$. Expanding the product yields

$$\sum x_t^* v_t^* = \sum x_t v_t - \sum x_t z_t'' C' \psi + \sum z_t^* v_t - \sum z_t^* z_t'' C' \psi,$$

and to find the limiting distribution of the first term of (19), we may further expand this term using the data generating process of $(x_t)$ given by (2). We obtain

$$\frac{1}{n} \sum (\mu + \Gamma q_t + u_t) v_t \rightarrow_d \Gamma \left( \int QdV (r) + \delta_{qv} \right) + \sigma_{uv}$$

as $n \rightarrow \infty$. (Note that if $q_0 = O_p(1)$ but not independent of $(v_t)$, we would have to contend with an additional nuisance parameter.) Similarly, the second term of (19) has a distribution given by

$$-\frac{1}{n} \sum (\mu + \Gamma q_t + u_t) z_t'' C' \psi \rightarrow_d -\Gamma \Xi_q C' \psi - \Sigma_{ux} C' \psi$$

using Lemma A.2[a], [f], and [c], respectively. The third and fourth terms of (19) are similarly governed by Lemma A.2[c] and [b], respectively, so that the stated result is obtained.
As in the proof of part [a], we use (4) to write
\[ \sum x_i^* w_i' = \sum x_i w_i' + \sum z_i^* w_i', \]
and the stated result follows along similar lines.

Expanding the summation in part [c] yields
\[ \sum w_i v_i^* = \sum w_i v_i - \sum w_i z_i^* C' \psi, \]
and, again, the stated result immediately follows.

Finally, expanding the summation in part [d] reveals a structure similar to part [a]. Specifically,
\[ \sum x_i^* x_i^* = \sum x_i x_i' + \sum z_i^* z_i' + \sum x_i z_i' + \sum z_i^* x_i', \]
where the first term has an asymptotic distribution of
\[ \frac{1}{n^2} \sum x_i x_i' \to_d \Gamma' \left( \int QQ' dr \right) \Gamma' \]
which dominates under our conditions.

Lemma A.4 Assume that [A1]-[A5], [NED1]-[NED6], [K1]-[K3] hold, and consider estimators \( \hat{\alpha}, \hat{\psi}, \hat{\mu}, \) and \( \hat{\Gamma} \) defined by the least squares estimators. We have
\[ \frac{1}{n} \sum x_i^* v_i^* \to_d \Gamma \left( \int V_{\perp} Q \right) \]
\[ \frac{1}{n} \sum x_i^* w_i^* \to_d \Gamma \left( \int W \right) \]
\[ \frac{1}{n^2} \sum x_i^* x_i' \to_d \Gamma \left( \int QQ' \right) \Gamma' \]
as \( n \to \infty. \)

Proof of Lemma A.4 [a] The summation may be expanded using (16) and (18) as
\[ \sum x_i^* v_i^* - \sum x_i^* \hat{\omega}_{vq} \hat{\Omega}_{qq}^{-1} \hat{q}_t - (\hat{\Gamma} \hat{\Delta}_q + \hat{\Sigma}_w) \hat{\Sigma}^{-1} \sum b_i^* b_i' \hat{k} \]
using feasible estimators of all parameters. The distribution of the first term of (20) follows from Lemma A.3[a]. The second term of (20) may be written as
\[ - \left( \sum x_i \Delta q_i^* + \sum z_i \Delta q_i^* \right) \hat{\Omega}_{qq}^{-1} \hat{\omega}_{qv}, \]
where the variance estimators have a limiting distribution of
\[ \hat{\Omega}_{qq}^{-1} \hat{\omega}_{qv} \to_p \Omega_{qq}^{-1} (\omega_{qv} - \Omega_{qv} C' \psi) \]
by Lemma 3.5. When normalized by \( 1/n, \) the first summation in (21) has an asymptotic distribution given by
\[ \Gamma \int QdQ \left( \int \right) + \Gamma \Delta_{qq} + \Sigma_{uq}, \]
and the probability limit of the second summation in (21) is $\Sigma_{uv}$ when similarly normalized.

To determine the limit of the final term of (20), we need to deal with the limit of $\sum \hat{b}_i^* b_i'^*$. Expanding this as

$$
\sum \hat{b}_i^* b_i' + \hat{D} \sum z_i^* b_i' + \hat{D} \sum z_i^* z_i'^* D',
$$

it is clear using (11) that this consistently estimates $\Sigma_{b^* b'^*}$, as does $\hat{\Sigma}$. We may thus write this term as

$$
-(\hat{\Gamma} \hat{\delta}_{qv} + \hat{\Sigma}_{uv}) + (\hat{\Gamma} \hat{\Delta}_{qq} + \hat{\Sigma}_{uq}) \hat{\Omega}^{-1} \hat{\omega}_{qv} + o_p(1).
$$

Note that

$$
(\hat{\Gamma} \hat{\delta}_{qv} + \hat{\Sigma}_{uv}) \to_p \Gamma (\delta_{qv} - \Delta q C' \psi) + (\sigma_{uv} - \Sigma_{us} C' \psi) + (\sigma_{sv} - \Sigma_{ss} C' \psi)
$$

and that

$$
(\hat{\Gamma} \hat{\Delta}_{qq} + \hat{\Sigma}_{uq}) \to_p \Gamma \Delta_{qq} + \Sigma_{uq} + \Sigma_{sq}
$$

as $n \to \infty$. Combining all of these terms (after appropriate cancellations) yields the stated result.

[b] Define a $(1 + p + r + g) \times p$ nonstochastic matrix

$$
E_w \equiv (0_p, I_p, 0_p, (r + g))
$$

such that $E_w$ selects the columns corresponding to $w$. In particular, $\hat{b}_i'^* E_w = w_i'$, so that using (16), the summation is equal to

$$
\sum x_i^* w_i' - (\hat{\Gamma} \hat{\delta}_{qv} + \hat{\Sigma}_{uv}) \hat{\Omega}^{-1} \sum \hat{b}_i^* \hat{b}_i'^* E_w
$$

where the distribution of the first term comes from Lemma A.3[b]. The second term is also $O_p (n)$ by Lemma 3.5, with a probability limit given by $-(\Gamma \Delta_{qw} + \Sigma_{uw} + \Sigma_{sw})$.

[c] Expanding the summation yields

$$
\sum x_i^* x_i'^* + (\hat{\Gamma} \hat{\Delta}_{qs} + \hat{\Sigma}_{us}) \hat{\Sigma}^{-1} \sum \hat{b}_i^* \hat{b}_i'^* \hat{\Sigma}^{-1} (\hat{\Gamma} \hat{\Delta}_{qs} + \hat{\Sigma}_{us}) - (\hat{\Gamma} \hat{\Delta}_{qs} + \hat{\Sigma}_{us}) \hat{\Sigma}^{-1} \sum \hat{b}_i^* x_i'^*
$$

where the asymptotics of the first term are derived in Lemma A.3[d]. It remains to show that the other terms are $o_p (n^2)$. This is clearly true for the second term, which is $O_p (n)$ as a direct result of Lemma 3.5. The summation $\sum x_i^* \hat{b}_i'^*$ in the third and fourth terms of (22) may be partitioned as

$$
\sum (x_i^* \hat{\psi}_i^*, x_i^* w_i', x_i^* \hat{\psi}_i'^*, x_i^* \Delta q_i^*),
$$

and we examine each partition separately. The first partition may be expanded as

$$
\sum x_i^* v_i^* + \sum x_i^* w_i' (\alpha - \hat{\alpha}) + \sum x_i^* x_i'^* (\psi' C - \hat{\psi}' \hat{C})
$$
These are clearly no more than $O_p(n)$ from Lemma A.3[a], [b], [d], and the superconsistency of $\hat{\psi}'\hat{C}$. The second partition is obviously $O_p(n)$ from Lemma A.3[a], [b], [d], and the superconsistency of $\hat{\psi}'\hat{C}$. The second partition is obviously $O_p(n)$ as a special case of the first. The third partition in (23) admits the expansion

$$
\sum x_t u_t' + \sum z_t^* u_t' + \sum x_t (\mu - \hat{\mu})' + \sum z_t^* (\mu - \hat{\mu})' + \sum x_t q_t' (\Gamma - \hat{\Gamma})' + \sum x_t z_t^{**} + \sum z_t^* z_t''$

using (4) and (11). All of these are $o_p(n^2)$ under our assumptions. The last partition in (23) is $O_p(n)$ for the same reasons as the second partition. Finally, returning to the third and fourth terms of (22), since $\hat{\Sigma}^{-1}(\hat{\Gamma} \hat{\Delta} q + \hat{\Sigma} u C') = O_p(1)$, the proof is complete. \qed

**Appendix B: Proofs of the Main Results**

**Proof of Theorem 3.1** Since we assume [NED6], we have $\sigma_{wv} - \Sigma_{wv} C' \psi = 0$ in Lemma A.3[b]. Consequently, using Lemma A.3[a]-[d] and the continuous mapping theorem,

$$
\frac{1}{n} M_n^* \rightarrow d C T \left( \int Q d (V - \psi' C Z^*(s)) + \delta_{qv} - \Delta_{qv} C' \psi \right) + C (\sigma_{uv} - \Sigma_{uv} C' \psi) + C (\sigma_{wv} - \Sigma_{wv} C' \psi)
$$

as $n \to \infty$. Similarly,

$$
\frac{1}{n^2} N_n^* = \frac{1}{n^2} \hat{C} \sum x_t^* x_t^{**'} + o_p(1) \rightarrow d C T \left( \int QQ' \right) \Gamma' C',
$$

so that the stated result is obtained. \qed

**Proof of Lemma 3.2** Using the definition of $u_t^*$, we may rewrite the first summation as

$$
\sum u_t q_t' - \sum u_t q_t^{**} + \sum z_t^* q_t' - \sum z_t^* q_t^{**}
$$

The first term is $O_p(n)$ under [A1]-[A4], using standard nonstationary asymptotics, and the second is also $O_p(n)$ since

$$
n^{-1/2} q_t \rightarrow_d Q_t,
$$

and since our assumption that $E u_t = 0$ allows a central limit theorem for $\sum u_t$. The third term is $O_p(n)$ by Lemma A.2[f]. And, by Lemma A.2[d], the fourth term is $O_p(n)$. It remains to show that the second summation in the estimator is $O_p(n^2)$. This is straightforward, since

$$
\frac{1}{n^2} \sum q_t q_t' - \frac{1}{n^{3/2}} \sum q_t q_t^{**} - \frac{1}{n^{3/2}} \sum q_t q_t' \rightarrow_d \int QQ' - \int Q \int Q'
$$

as $n \to \infty$. This is invertible since the trends $(q_t)$ are distinct. \qed
The proof of Lemma 3.3 is as follows.

**Proof of Lemma 3.3** The estimator \((\hat{\alpha}_{LS} - \alpha)\) may be rewritten as

\[
\left(\sum w_t w_t'\right)^{-1} \sum w_t \left((\psi'C - \psi'_t \hat{C})x_t - \psi'_t \hat{C} z_t^* + \nu_t\right)
\]

by substituting (4) and our definition of \((\psi'_t)\) into (7). Under our assumptions, \(\sum w_t w_t' = O_p(n)\) and invertible. The second summation may be rearranged as

\[
\sum w_t (v_t - z_t^* \hat{C}' \hat{\psi}_{LS}) + \sum w_t x_t' (C' \psi - \hat{C}' \hat{\psi}_{LS}).
\]

Under [NED6], \(E w_t (v_t - z_t^* \hat{C}' \hat{\psi}_{LS}) = 0\), so we may apply a central limit theorem in addition to Lemma A.2\([e]\) to see that the first term is \(O_p\left(n^{1/2}\right)\). In the second term, note that

\[
(\psi'C - \hat{\psi}_{LS} \hat{C}) = (\psi - \hat{\psi}_{LS})'C + (\hat{\psi}_{LS} - \psi)'(C - \hat{C}) + \psi'(C - \hat{C}),
\]

each term of which is \(O_p\left(n^{-1}\right)\) under our assumptions. Consequently, the whole estimator is \(O_p\left(n^{-1/2}\right)\).

Similarly, the estimator \((\hat{\mu}_{LS} - \mu)\) is equal to

\[
(\Gamma - \hat{\Gamma}_{LS}) \frac{1}{n} \sum q_t + \frac{1}{n} \sum u_t + \frac{1}{n} \sum z_t^*
\]

using (8) and (5). The first term is \((\hat{\Gamma}_{LS} - \Gamma)O_p\left(n^{1/2}\right)\), which is \(O_p\left(n^{-1/2}\right)\) by Lemma 3.2. The second term is \(O_p\left(n^{-1/2}\right)\) using a CLT, as is the third term by Lemma A.2\([d]\). \(\square\)

**Proof of Lemma 3.4** First, expand (9) as

\[
\hat{\Omega}_{b'r'} = \frac{1}{n} \sum_{t=1}^{n} \sum_{s=1}^{n} \left\{ b_t b_s' + D z_t^* z_s^{*'} D' + D z_t^* b_s' + b_t z_s^{*'} D' \right\} \pi \left( \frac{t - s}{h_n} \right) \tag{24}
\]

Define the vector \(\zeta_t \equiv (b_t', z_t^{*'})'\), the elements of which are \(L_2\)-NED by Lemma A.1\([c]\). Note that the matrix \(\zeta_t^*\zeta_t^*\) may be partitioned so that the partitions are the four terms enclosed by curly brackets in (24) (up to linear transformations involving \(D\)). In order to prove the lemma, we show consistency of the analogous estimator based on \(\zeta_t^*\) rather than \(b_t'\), using the limit theory of de Jong and Davidson (2000). Consistency of the estimator of interest then follows by the Slutsky theorem.

First, note that for an \(m\)-dimensional random vector \(z\),

\[
\sum_i E |z|^2 \leq m \sup_i E |z_i|^2
\]

for \(i = 1, \ldots, m\) so that

\[
\|z\|_2 \leq m^{1/2} \sup_i \|z_i\|_2
\]

where the norm on the LHS is an \(L_2\)-norm for vectors, whereas that on the RHS is an \(L_2\)-norm for scalars. Using this inequality, we may write

\[
\left\| \zeta_t - E \left( \zeta_t | F_{t+K}^{L-K} \right) \right\|_2 \leq (1 + p + 2r + g)^{1/2} \sup_i \left\| \zeta_t - E \left( \zeta_t | F_{t+K}^{L-K} \right) \right\|_2 \leq d_t^{h_z} \psi_{K}^{b_z}
\]

The second summation may be rearranged as

\[
\sum w_t (v_t - z_t^* \hat{C}' \hat{\psi}_{LS}) + \sum w_t x_t' (C' \psi - \hat{C}' \hat{\psi}_{LS}).
\]
for the vector sequence \((\zeta_t)\). Sequences \((d_{t}^{b,z})\) and \((\nu_{K}^{b,z})\) are defined in terms of the respective sequences \((d_{t})\) and \((\nu_{K})\) that are implicitly defined in Lemma A.1[c]. Specifically, \((d_{t}^{b,z})\) is \((1 + p + 2r + g)^{1/2}\) times the maximum the \(d_{t}\)'s and \((\nu_{K}^{b,z})\) is defined as the maximum of the \(\nu_{K}'\)s.

In order for Theorem 2.1 of de Jong and Davidson (2000) to hold, a sufficient condition is that

\[
\sup_{t} (\|\zeta_{t}\|_{2a/(a-1)} + d_{t}^{b,z}) c_{t}^{-1} < \infty
\]

(25)

with

\[
\frac{1}{n} \sum_{t} c_{t}^{2} < \infty,
\]

for some constant sequence \((c_{t})\). We may write

\[
\sup_{t} (\|\zeta_{t}\|_{2a/(a-1)} + d_{t}^{b,z}) c_{t}^{-1} \leq \sup_{t} \|\zeta_{t}\|_{2a/(a-1)} c_{t}^{-1} + \sup_{t} d_{t}^{b,z} c_{t}^{-1}
\]

and note that the first term is

\[
\sup_{t} \|\zeta_{t}\|_{2a/(a-1)} c_{t}^{-1} \leq (1 + p + 2r + g)^{1/2} \sup_{t} (\|\zeta_{t}\|_{2a/(a-1)} c_{t}^{-1}).
\]

Of course, \(1 + p + 2r + g\) is assumed to be finite. Let \(c_{t} = 1\) for all \(t\), which clearly satisfies (26) and bounds \((c_{t}^{-1})\). The moment restriction on \((z_{t}^{*})\) given by [NED4] and the covariance stationarity of \((b_{t})\) assumed in [A1] imply that \(\sup_{t} \|\zeta_{t}\|_{2a/(a-1)}\) is bounded. Clearly, \(\sup_{t} d_{t}^{b,z}\) is also bounded by [NED1]. As a result, the sufficient condition (25) holds. Moreover, for this choice of \((c_{t})\), the remaining sufficient conditions for Theorem 2.1 of de Jong and Davidson (2000) are directly satisfied by [K1] and [K2]. We may thus apply the results of their theorem to the covariance matrix of \((\zeta_{t})\), so that our result holds.

\[\Box\]

**Proof of Lemma 3.5** The proof is inspired by the proof of Lemma 4.3 of Park (1992), with the main complication being the sequence \((z_{t}^{*})\) of nonstationary imputation error. Because of this complication, we cannot restrict attention to the scalar case, as in that proof. To that end, let the \(ij^{th}\) element of a matrix \(X\) be denoted by \([X]_{ij}\).

Consider \(\Delta_{b^{*}b^{*}}\). If we can show that \(\hat{\Delta}_{b^{*}b^{*}} = \Delta_{b^{*}b^{*}} + o_{p}(1)\), then we may apply Lemma 3.4[b] to obtain the stated result for part [b]. (Part [a] is a special case, and part [c] is a trivial extension.) The absolute value of difference between the two estimators of \(\Delta_{b^{*}b^{*}}\) is

\[
|\hat{\Delta}_{b^{*}b^{*}} - \Delta_{b^{*}b^{*}}| \leq \frac{1}{\sqrt{n}} \sum_{k=0}^{n} \left| \frac{k}{h_{n}} \right| \left( \frac{1}{\sqrt{n}} \sum_{k=1}^{n} \left| \hat{b}_{t}^{*} (\hat{b}_{t-k}^{*} - b_{t-k}^{*}) \right| \right)
\]

(27)

by the triangle inequality. We will show that each element of this matrix is \(o_{p}(1)\). Since the sum over \(k\) of the kernel function evaluated at \(k/h_{n}\) is \(o\left(n^{1/2}\right)\) by [K3], the result holds if we can show that the two remaining sums in (27) are both \(O_{p}\left(n^{1/2}\right)\).

Examining the \(ij^{th}\) element of the matrix in the second summation in (27), we may write

\[
\frac{1}{\sqrt{n}} \sum_{t=k+1}^{n} \left| \hat{b}_{t}^{*} (\hat{b}_{t-k}^{*} - b_{t-k}^{*}) \right|_{ij} \leq \left( \frac{1}{n} \sum_{t} \left| \hat{b}_{t}^{*} b_{t}^{*} \right|_{ii} \sum_{t} \left| \hat{b}_{t}^{*} - b_{t}^{*} \right|_{jj} \right)^{1/2}
\]

(28)
using the Cauchy-Schwarz inequality and the fact that \( k \geq 0 \). We may expand
\[
\sum [\hat{b}_t \hat{b}_t']_{ii} = \sum [\hat{b}_t \hat{b}_t']_{ii} + \sum [\hat{D} \hat{z}_t^* \hat{z}_t^*']_{ii} + 2 \sum [\hat{b}_t \hat{z}_t^* \hat{D}']_{ii}
\]  
(29)
using (12). The stochastic boundedness of \( \frac{1}{n} \sum [\hat{b}_t \hat{b}_t']_{ii} \) follows by the same reasoning as that employed by Park (1992, Lemma 4.3), since some of the subseries – \((w_t)\) and \((\Delta q_t)\) – in \((\hat{b}_t)\) are stationary, and the remaining subseries consistently estimate stationary series \((v_t)\) and \((u_t)\) by way of (11) and the explicit assumptions about the rates of convergence of \( \hat{\psi} - \psi \), etc. The second term of (29) is \( O_p(1/n) \) as a direct result of Lemma A.2[b] and the consistency of the estimator \( \hat{D} \). The third term of (29) is slightly more complicated. It involves sums of element of both outer products of \((w_t)\) and \((\Delta q_t)\) with \((\hat{z}_t^*)\), and of \((\hat{v}_t)\) and \((\hat{u}_t)\) with \((z_t^*)\). The former sums are \( O_p(n) \) by Lemma A.2[c], posing no problem. The latter sums are
\[
\sum v_t z_t^* + (\alpha - \hat{\alpha})' \sum w_t \hat{z}_t^* + (\psi' C - \hat{\psi}' \hat{C}) \sum x_t z_t^*
\]
and
\[
\sum u_t z_t^* + (\mu - \hat{\mu}) \sum \hat{z}_t^* + (\Gamma - \hat{\Gamma}) \sum q_t z_t^*
\]
which are also also \( O_p(n) \) by [A5], Lemma A.2[a], [c], and [f], Theorem 3.1, and Lemmas 3.2 and 3.3. An expansion of \( \sum [(\hat{b}_t - b_t)(\hat{b}_t - b_t')]'_{jj} \) in (28) yields
\[
\sum [(\hat{b}_t - b_t)(\hat{b}_t - b_t')]'_{jj} + \sum [(\hat{D} - D) \hat{z}_t^* \hat{z}_t^*'(\hat{D} - D)']_{jj} + \sum [(\hat{b}_t - b_t) z_t^* \hat{z}_t^*(\hat{D} - D)']_{jj},
\]  
(30)
the first term of which is \( O_p(1) \) – again, as a straightforward extension of the proof in Park (1992, Lemma 4.3). Again, the summation \( \sum \hat{z}_t^* \hat{z}_t^* \) is \( O_p(n) \), by Lemma A.2[b]. Elements of the matrix \((\hat{D} - D)\) are either identically 0 or \( O_p(1/n) \) by construction and superconsistency of \( \hat{\psi} \) and \( \hat{C} \). Consequently, the second term of (30) is also \( O_p(1) \). Turning to the third term of (30), the vector \((\hat{b}_t - b_t)\) contains zeros for observable series \((w_t)\) and \((\Delta q_t)\). For the subseries \((\hat{v}_t)\) and \((\hat{u}_t)\) of estimates, we have
\[
(\alpha - \hat{\alpha})' \sum w_t \hat{z}_t^* (\hat{D} - D)' + (\psi' C - \hat{\psi}' \hat{C}) \sum x_t z_t^* (\hat{D} - D)'
\]
and
\[
(\mu - \hat{\mu}) \sum z_t^* (\hat{D} - D)' + (\Gamma - \hat{\Gamma}) \sum q_t \hat{z}_t^* (\hat{D} - D)'
\]
which are \( O_p(1) \) under our assumptions.

Finally, returning to (27), we must show that the third summation is stochastically bounded. This follows in exactly the same way as the second term, except that the first summation on the RHS of (28) is replaced with \( \frac{1}{n} \sum [\hat{b}_t \hat{b}_t']_{ii} \), which is stochastically bounded as a direct result of our assumptions. This completes the proof for \( \hat{\Delta} b_t b_t' \).  \( \square \)
Proof of Theorem 4.1  The estimator $\hat{\psi}_{CCR}$ may be rewritten as

$$
\hat{\psi}_{CCR} = \left( \hat{C} \sum x_i^{**} x_i^{**'} \hat{C}' \right)^{-1} \left( \hat{C} \sum x_i^{**} w_i' (\alpha - \hat{\alpha}) + \hat{C} \sum x_i^{**} x_i^{**'} \hat{C}' \psi + \hat{C} \sum x_i^{**} v_i^{**} \right)
$$

which (since $\hat{C} \to C$) may be written as

$$
\left( \hat{\psi}_{CCR} - \psi \right) = \left( C \sum x_i^{**} x_i^{**'} C' \right)^{-1} \left( C \sum x_i^{**} w_i' (\alpha - \hat{\alpha}) + C \sum x_i^{**} v_i^{**} \right) + o_p(1)
$$

using the Slutsky theorem. Now, since $(\alpha - \hat{\alpha}) = o_p(1)$, Lemma A.4[b] implies that $\sum x_i^{**} w_i' (\alpha - \hat{\alpha}) = o_p(n)$. The resulting distribution of $(\hat{\psi}_{CCR} - \psi)$ follows directly from Lemma A.4[a] and [c]. □