Cointegrating Regressions with Messy Regressors: Missingness, Mixed Frequency, and Measurement Error

J. Isaac Miller[1]

Department of Economics

University of Missouri

Abstract

We consider a cointegrating regression in which the integrated regressors are messy in the sense that they contain data that may be mismeasured, missing, observed at mixed frequencies, or have other irregularities that cause the econometrician to observe them with mildly nonstationary noise. Least squares estimation of the cointegrating vector is consistent. Existing prototypical variance-based estimation techniques, such as canonical cointegrating regression (CCR), are both consistent and asymptotically mixed normal. This result is robust to weakly dependent but possibly nonstationary disturbances.

First draft: January 18, 2007
This draft: April 14, 2009

JEL Classification: C13, C14, C32

Key words and phrases: cointegration, canonical cointegrating regression, near-epoch dependence, messy data, missing data, mixed-frequency data, measurement error, interpolation

[1] I am grateful to participants of the North American Summer Meeting of the Econometric Society, the Midwest Econometrics Group, the Missouri Economics Conference, and invited seminars at OSU, TAMU, Rice, York, Exeter, Leeds, and Nottingham – particularly, Bill Brown, James Davidson, Robert de Jong, Mahmoud El-Gamal, Dave Mandy, Shawn Ni, Joon Park, Peter Phillips, Yongcheol Shin, Robin Sickles, and Rob Taylor – for helpful comments and suggestions. I am also grateful to Wensheng Kang for research assistance. The usual caveat applies.

Correspondence: millerjisaac@missouri.edu, 573-882-7282, 118 Professional Bldg., Columbia, MO 65203.

1. Introduction

The fully modified least squares (FM-OLS) and canonical cointegrating regression (CCR) techniques (Phillips and Hansen, 1990, and Park, 1992, respectively) use covariance estimators to estimate the cointegrating vector of a prototypical cointegrating regression. As covariance-based techniques, these estimators explicitly rely on covariance stationary errors. Very common data irregularities, such as measurement error or imputation error resulting from handling missing observations or mixed frequencies, may explicitly violate covariance stationarity. Intuitively, we may expect that mildly nonstationary errors will not substantially affect the superconsistency of the least squares estimator. However, the effect on the limiting (mixed) normality of multi-step covariance-based estimators (which rely on least squares in the first step) is not readily apparent. The asymptotic approximations that guide us to statistical inference about such models may be invalid. Our aim is to show that prototypical covariance-based techniques such as these are robust to mildly nonstationary error sequences. In particular, we rigorously show that least squares estimation of a canonical cointegrating regression is consistent and asymptotically mixed normal (CAMN) for a prototypical single-equation cointegrating regression. Our results are suggestive for other covariance-based techniques, such as FM-OLS.

In order to avoid the assumption of covariance stationary errors for covariance-based estimators, we make extensive use of near-epoch dependence, a concept that was defined and refined by *inter alia* Ibragimov (1962), Billingsley (1968), and McLeish (1975). Key results for near-epoch dependent processes developed by *inter alia* Davidson (1994), Davidson and de Jong (1997), and de Jong and Davidson (2000) provide a useful alternative paradigm to covariance stationarity. Allowing mild nonstationarity in the error term changes the asymptotic tools from those of Park (1992), but we show that the estimation technique, its consistency, and its limiting (mixed) normality are the same. The theoretical contribution of this paper adds to recent literature that seeks current and new econometric techniques that are robust to violations of classical covariance stationarity assumptions and generalizes the concept of I(0).[2]

The statistics and econometrics literatures abound with explicit techniques for handling certain types of messy time series, such as those with missing or mixed-frequency observations. For example, some early regression-based techniques were proposed by Friedman (1962) and Chow and Lin (1971, 1976). Recent approaches under the MIDAS moniker (Ghysels *et al.*, 2004) have proposed spectral methods to estimate mixed-frequency models with stationary series. A number of authors (Harvey and Pierse, 1984; Kohn and Ansley, 1986; Gomez and Maravall, 1994; Harvey *et al.*, 1998; Mariano and Murasawa, 2003; Seong *et al.,* 2007; *inter alia*) have used the Kalman filter with mixed-frequency data.

Although we motivate our concerns about nonstationarity with a particular imputation technique in a particular (mixed-frequency) setting, the theoretical contribution of this research is not associated with any particular imputation technique – rather, a wide variety

[2] For example, Müller and Watson (2008a) discuss tests for persistence alternatively to classical unit root tests. Müller and Watson (2008b) favor robust cointegration testing motivated by Wright's (2000) variance-ratio tests. Kurozumia and Hayakawa (2009) explore estimation of cointegrating regressions similar to those considered here, but with local to unity errors.

of imputation techniques may generate imputation error satisfying our sufficient conditions. Moreover, if the econometrician obtains data to which some such technique has already been applied, imputation error is instead measurement error. In fact, nonstationary error of the type allowed in our analysis could stem from sources well beyond either of these. GARCH(1,1) models, for example, contain mildly nonstationary error (Hansen, 1991). For the sake of generality, we refer to all noise beyond the usual stationary modeling error as *messy-data noise*. Letting (x_t^*) denote the messy analog of an integrated series (x_t) series, we let (z_t^*), defined by

$$z_t^* \equiv x_t^* - x_t, \tag{1}$$

denote the messy-data noise.

As a non-technical motivating example, we consider simple linear interpolation (lerp) in a mixed-frequency setting. Even this simplistic technique generates mild nonstationarity, and conventional wisdom suggests that linear interpolation is inferior to estimating a model at the lowest frequency (omitting data observed at higher frequencies). Simulations support the conventional wisdom for a *stationary* regression. In a *cointegrating* regression, however, the conventional wisdom does not hold, and simulations suggest that lerp may *outperform* omission. These results are suggestive for more sophisticated imputation techniques.

In the following section, we briefly discuss the example in order to motivate the general econometric concerns arising from messy data. In Section 3, we describe our econometric model in detail. We outline general sufficient conditions for the messy-data noise. In Sections 4 and 5, we present our main theoretical results. Specifically, we focus on consistent parameter estimation in Section 4, and we show that CCR is CAMN in Section 5. In Section 6, we briefly revisit the example of Section 2, verifying our theoretical sufficient conditions and presenting some small-sample results specific to the example. The final section concludes. Two technical appendices contain ancillary lemmas and their proofs, along with proofs of the main results.

Unless otherwise noted, summations are indexed by $t = 1, \ldots, n$ and integrals are evaluated over $s \in [0, 1]$. We use $||\cdot||_p$ to denote the L_p-norms $(\sum_i \sum_j \mathbf{E}|a_{ij}|^p)^{1/p}$, $(\sum_i \mathbf{E}|a_i|^p)^{1/p}$, and $(\mathbf{E}|a|^p)^{1/p}$ of matrices, vectors, and scalars, respectively.

2. Example: Lerping Mixed-Frequency Data

In order to further motivate the concept of messy data, we consider a very specific example: using linear interpolation (lerp) to impute a low-frequency series at a higher frequency. We consider lerp only as an illustrative example, but do not generally advocate lerp over other techniques in dealing with mixed-frequency data.

Consider two time-series of interest (y_t) and (x_t). We observe (y_t) at each $t = 1, \ldots, n$, but (x_t) is observed only over a subset of this index set. Let the observed subseries of (x_t) be denoted by (x_{τ_p}) with $p = 1, \ldots, l$, and let each unobserved subseries be denoted by $(x_{\tau_p + j})$ with $j = 1, \ldots, m$. There are thus l unobserved subseries of finite length m. For example, if interpolating 20 years of a quarterly data to a monthly frequency, then $n = 240$, $l = 80$, and $m = 2$.

We make the conventions that x_0 is observed, $\tau_0 = 0$, and $\tau_p + (m+1) = \tau_{p+1}$, so that $\tau_l = l\,(m+1) = n$. The messy-data noise (in this case, interpolation error) is

$$z_t^* = \frac{j}{m+1} \sum\nolimits_{i=1}^{m+1} \triangle x_{\tau_{p-1}+i} - \sum\nolimits_{i=1}^{j} \triangle x_{\tau_{p-1}+i} \tag{2}$$

for $t = \tau_{p-1} + j$, $p = 1, \ldots, l$, and $j = 1, \ldots, m+1$. By construction, there is no noise when $j = m+1$ (or, alternatively, when $j = 0$) – i.e., when (x_t) is observed. The noise (z_t^*) explicitly depends on the time index j within each interval of missing data. Even under the most optimistic assumptions about $(\triangle x_t)$, (z_t^*) is not covariance stationary. In spite of nonstationarity, as long as m is finite, (z_t^*) is only *mildly* nonstationary, in the sense of its asymptotic order. We will exploit the asymptotic dominance of (x_t) over (z_t^*).

3. Cointegrating Regression With Messy Regressors

Consider a cointegrating regression given by

$$y_t = \alpha' w_t + \beta' x_t + v_t, \tag{3}$$

where (x_t) is an r-dimensional I(1) series, (w_t) is a p-dimensional stationary series, (v_t) is a one-dimensional series of unobservable stationary disturbances with mean zero, and α and β are conformable vectors of unknown parameters such that β does not cointegrate (x_t). Under these assumptions, $(y_t, x_t')'$ is cointegrated with cointegrating vector $(1, -\beta')'$.

Allowing for the possibility of cointegrated regressors, we may define (x_t) in terms of its stochastic trends. Specifically, we let

$$x_t = \mu + \Gamma q_t + u_t, \tag{4}$$

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

We define $b_t \equiv (v_t, w_t', u_t', \triangle q_t')'$ such that (b_t) is an R-dimensional series, where $R \equiv 1 + p + r + g$. We assume throughout the paper that

[A1] (b_t) is a mean-zero series that is stationary and α-mixing of size $-a$ with $a > 1$ and finite moments up to $4a/\,(a-1)$, and
[A2] Either $q_0 = 0$ or $q_0 = O_p\,(1)$ and independent of (u_t), (v_t), and (w_t).

If the initial value is stochastic, independence from the stationary series avoids additional nuisance parameters that add unnecessary complications to the model.

In light of the integratedness of (q_t) and partial sums of the other stationary series in the model, we define a stochastic process $B_n\,(s) \equiv n^{-1/2} \sum_{t=1}^{[ns]} b_t$, where $[ns]$ denotes the greatest integer not exceeding ns. We assume an invariance principle (IP) for $B_n\,(s)$, so that

[A3] $B_n\,(s) \to_d B\,(s) \equiv (V(s), W(s)', U(s)', Q(s)')'$ and

[A4] $\Omega_{bb} > 0$,

for vector Brownian motion $B(s)$ with finite variance Ω_{bb}. Using Σ_{bb} to denote the contemporaneous variance of (b_t), the one-sided long-run variance Δ_{bb} is the sum of the covariances running from 0 to ∞, which is implicitly defined by $\Omega_{bb} = \Delta_{bb} + \Delta'_{bb} - \Sigma_{bb}$.

We also assume that

[A5] $\mathbf{E}v_t w'_{t-k} = 0$ for all k,

to allow $\sqrt{n}$-consistent estimation of α.

3.1 Messy Regressors

The heart of the analysis is messy integrated regressors. We define the messy-data noise by (1), where (x_t) contains r regressors observed with noise as (x_t^*). Using (1), feasible analogs of (3) and (4) are thus

$$y_t = \alpha' w_t + \beta' x_t^* + v_t^* \tag{5}$$

and

$$x_t^* = \mu + \Gamma q_t + u_t^*, \tag{6}$$

respectively, where $u_t^* \equiv u_t + z_t^*$ and $v_t^* \equiv v_t - \beta' z_t^*$. The feasible regressors (x_t^*) in (5) are explicitly correlated with the new error term (v_t^*). Herein lies potential for asymptotically biased estimation, if not properly addressed.

Definition. *A sequence* (z_t) *is* near-epoch dependent in L_p-norm (L_p-NED) *of size* $-\lambda$ *on a stochastic sequence* (η_t) *if* $||z_t - E(z_t|F_{t-K}^{t+K})||_p \leq d_t \nu_K$*, 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* F_{t-K}^{t+K} *is the* σ*-field generated by* $\eta_{t-K}, \ldots, \eta_{t+K}$.

We primarily deal with L_2-NED sequences in this paper, which are simply described as *near-epoch dependent* (NED). The reader is referred to Davidson (1994), e.g., for more details.

The NED framework allows the generality required to deal with messy-data noise generated by techniques as simple as lerp or much more complicated, as long as the rates at which the appropriate sample moments diverge are properly taken into account. Mildly nonstationary dependence and heterogeneity are allowed, as long as the nonstationarity of (z_t^*) does not dominate the nonstationarity of (x_t).

Letting (z_{it}^*), with $i = 1, \ldots, r$, be an element of (z_t^*), we assume that the following hold:

[NED1] For all i, (z_{it}^*) is L_2-NED of size -1 on (b_t) w.r.t. a bounded sequence (d_{it}^z) of constants, and
[NED2] $\mathbf{E}z_t^* = 0$ for all t.

We define (possibly time-dependent) covariance matrices $\Sigma_{**}^t \equiv \mathbf{E}z_t^* z_t^{*\prime}$ and $\Sigma_{*b}^t \equiv \mathbf{E}z_t^* b_t'$ that satisfy

[NED3.a] $\Sigma_{**}^t < \infty$ for all t, $\Sigma_{**} \equiv n^{-1} \sum \Sigma_{**}^t < \infty$, and
[NED3.b] $\Sigma_{*b}^t < \infty$ for all t, $\Sigma_{*b} \equiv n^{-1} \sum \Sigma_{*b}^t < \infty$,

so that even though the sample moments may converge to time-dependent spatial averages, the average of each across time is finite and independent of time. With the additional assumption that

[NED4] $\sup_t ||z^*_{it}||_{4a/(a-1)} < \infty$ for all i,

we may employ limit theory of Davidson (1994), Davidson and de Jong (1997), and de Jong and Davidson (2000) to partial sums of (z^*_t) and (b_t) and their products.

Without messy data, asymptotics involving integrated (q_t) are straightforward from an IP and other limiting distributions implied by [A1]-[A5]. Assuming that an IP holds for $Z^*_n(s) \equiv n^{-1/2}\sum_{t=1}^{[ns]} z^*_t$ 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 variances satisfy

[NED5.a] $\mathbf{E}Z^*_n Z^{*\prime}_n(s) = \Omega_{**}(s)$ with $\Omega_{**}(s) < \infty$, and
[NED5.b] $\mathbf{E}Z^*_n Q'_n(s) = \Omega_{*q}(s)$ with $\Omega_{*q}(s) < \infty$.

Under these assumptions, $((Z^{*\prime}_n(s), Q'_n(s))' \to_d (Z^{*\prime}(s), Q'(s))'$, where $Z^*(s)$ is a vector Brownian motion with variance $\Omega_{**}(s)$.

Finally, we define a stochastic process (with an abuse of notation) $Z^*W_{n,j}(s) \equiv n^{-1/2}\sum_{t=1}^{[ns]} z^*_t w_{jt}$, for $j = 1, \ldots, p$, and assume

[NED6.a] $\mathbf{E}z^*_t w'_{t-k} = 0$ for all t, k,
[NED6.b] $\mathbf{E}Z^*W_{n,j}(Z^*W_{n,j})'(s) = \Omega_{*w_j*w_j}$ with $\Omega_{*w_j*w_j} < \infty$.

Analogously to [A5], the prohibition of correlation between the stationary regressors and the messy-data noise is necessary to obtain $\sqrt{n}$-consistent estimation of α.

The limit theory made accessible by assumptions [NED1]-[NED6] is collected in Lemma A.2 in an appendix. In some cases (see Section 6, for example), it may be more straightforward to verify the five results of Lemma A.2 directly, rather than verify sufficient conditions [NED2], [NED3.a], [NED3.b], [NED5.a], [NED5.b], [NED6.a], and [NED6.b] which are used only to prove Lemma A.2. If these results are verified directly, then [NED1] and

[NED4'] $\sup_t ||z^*_{it}||_{2a/(a-1)} < \infty$ for all i,

should also be verified. Note that [NED4'] relaxes [NED4] and is easier to verify. For the reader's convenience, alternative subsets of these assumptions are given for each result.

3.2 Cointegrated Regressors

Setting aside messy data, this model differs from the that of the standard CCR and FM-OLS approaches, in that we allow cointegrated regressors.[3] The limiting moment matrix to be inverted in the least squares estimator of β is not invertible (unless $g = r$), since it will be an $r \times r$ matrix of rank g. This collinearity may be remedied by choosing a $g \times r$ matrix C that does not contain any cointegrating vectors 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)

[3] Park (1992) considered cointegrated regressors in an early section in which he outlined the model to be estimated, but did not allow for this in estimation. Phillips (1995) expanded FM-OLS to estimate cointegrated VAR's.

and estimate a $g \times 1$ vector ψ in place of β. We may then make inferences about β' using $\psi' C$.

Since the only requirements for C are dimension and lack of cointegrating vectors of (x_t), a natural choice is the Moore-Penrose generalized inverse $C = (\Gamma'\Gamma)^{-1}\Gamma'$, so that $C\Gamma = I$ and ψ may be interpreted as the vector of marginal effects of the trends themselves. The marginal effects of the original regressors thus reflect their respective dependence on the underlying stochastic trends (q_t). Note that this matrix contains no cointegrating vectors of (x_t), since, by construction, Γ does not cointegrate (q_t).

In general, we must estimate C, and we assume the estimator $\hat{C}$ satisfies

[A6] $(\hat{C} - C) = O_p\left(n^{-1}\right)$.

Since C may be chosen arbitrarily, we just need to ensure that $\hat{C}$ converges to some matrix of appropriate rank that does not contain any cointegrating vectors.

Although not central to our main results, we briefly discuss alternative methods for choosing or estimating $\hat{C}$ to satisfy assumption [A6].

[1] If the regressors are *not* cointegrated, choose $\hat{C} = C = I$.

[2] If either common stochastic trend(s) or proxies are observed, choose $\hat{C} = (\hat{\Gamma}'\hat{\Gamma})^{-1}\hat{\Gamma}'$. As long as $(\hat{\Gamma} - \Gamma) = O_p\left(n^{-1}\right)$, which we show below for least squares, [A6] holds.[4]

[3] If there is only *one* common stochastic trend, but no observable proxy, choose $\hat{C} = C$ to be a binary vector with a single unit element and zeros elsewhere.

[4] If none of the above apply, we suggest using a standard trend estimation technique, such as the common factor approach of Gonzalo and Granger (1995) or the Kalman filter approach of Chang *et al.* (2009). If a "clean" subseries of the regressor series is available, such as lower frequency data in the example of mixed-frequency data, these techniques consistently estimate Γ at the appropriate rate, as long as the lower frequency sample size is a constant fraction of n. If no clean subseries is available, these techniques may still be consistent, just as least squares is still consistent with messy data (as we show below).

[5] An arbitrary $g \times r$ matrix $\hat{C} = C$ may be chosen, but this method is the least desirable. A near-cointegrating vector in C could change the rate of convergence of $\hat{\psi}$ to ψ. In the extreme case that a perfect cointegrating vector is chosen (which with probability 1 cannot happen if chosen randomly from a continuous distribution), the rate is slowed to $\sqrt{n}$ and the messiness may more fundamentally contaminate the estimator.

[4]Examples of observable proxies for single common stochastic trends are common in practice. They include, for example, the national cost of living index for a vector of metropolitan cost of living indices with missing data analyzed by Chang and Rhee (2005) and national stock market index for individual stock price data with missing data analyzed by Goetzmann *et al.* (2001).

3.3 Testing for Cointegration

In cases where cointegration between (y_t) and (x_t) is not obvious or expected,[5] testing is desirable. The variance ratio test and multivariate trace statistic proposed by Phillips and Ouliaris (1990) rely on the estimation of the long-run variance of a different residual series. Specifically, $(y_t, x_t')'$ is regressed on one lag of itself, and the long-run variance of the residual series from that regression is estimated. This series is I(0) under the null, so we can expect that adding (z_t^*) – i.e., using $(y_t, x_t^{*\prime})'$ – would have a similar effect to adding (z_t^*) to the residual series below. Specifically, it would inflate the variance of both the numerator and denominator of these statistics, so that the limits would be preserved. Approaches using more robust variance ratio tests (Wright, 2000; Müller and Watson, 2008) hold more promise still.

3.4 Further Generalizations

Messy Regressand. A messy regressand requires only a straightforward extension. If $e_t^* \equiv y_t^* - y_t$ for observed regressand (y_t^*), then $v_t^* \equiv v_t - \beta' z_t^* + e_t^*$ replaces the above definition, and we simply need the assumptions about (z_t^*) to hold jointly for $(e_t^*, z_t^{*\prime})'$.

Messy Stationary Regressors. Least squares no longer estimates α consistently, which is a key for variance estimation. Instead, (w_t) may be dropped from the regression, so that $v_t^* \equiv v_t - \beta' z_t^* + \alpha' w_t$. Our asymptotic results hold, but with a larger variance.

Deterministic Trends. If the deterministic trends dominate the stochastic trends in (3) and (4), the asymptotics would be fundamentally different. We focus on asymptotically dominant I(1) stochastic trends in this analysis.

4. Consistent Estimation

We turn to consistent estimation of the parameters described by the feasible system given by (5) and (6), with the substitution of $\psi' \hat{C}$ for β'.

4.1 Consistent Estimation of ψ

Least squares estimation of (5) provides a consistent (but neither asymptotically normal nor unbiased) estimator of ψ. The least squares estimator is $(\hat{\psi}_{LS} - \psi) = (\hat{C} N_n^* \hat{C}')^{-1} \hat{C} M_n^*$ with $M_n^* \equiv \sum x_t^* v_t^* - \sum x_t^* w_t' \left(\sum w_t w_t'\right)^{-1} \sum w_t v_t^*$ and $N_n^* \equiv \sum x_t^* x_t^{*\prime} - \sum x_t^* w_t' \left(\sum w_t w_t'\right)^{-1} \sum w_t x_t^{*\prime}$.

[5] An example of an *obvious* case is one in which the regressor(s) are lagged values of (y_t). Examples of cases in which cointegration may be *expected* include models with clear common stochastic trends, such as the cost of living index and the stock price index examples mentioned above.

Theorem 4.1 Let [A1]-[A6] hold. Further, assume that either [NED1]-[NED6] or the results of Lemma A.2[a]-[c], [e], and [f] hold. Define $N \equiv \Gamma \int QQ'\Gamma'$ and

$$M^* \equiv \Gamma \int Qd(V(s) - Z^*(s)' C'\psi) + \Gamma(\delta'_{vq} - \Delta'_{*q}C'\psi) \\ + (\sigma_{uv} - \Sigma_{u*}C'\psi) + (\sigma_{*v} - \Sigma_{**}C'\psi).$$

The least squares estimator $\hat{\psi}_{LS}$ has a distribution given by $n(\hat{\psi}_{LS} - \psi) \to_d (CNC')^{-1}CM^*$ as $n \to \infty$.

Due to the superconsistent rate of convergence of the estimator to its asymptotic distribution, $\hat{\psi}_{LS}$ is consistent in spite of numerous nuisance parameters.[6]

4.2 Consistent Estimation of Γ, α, and μ

The asymptotic distribution above critically depends on Γ, so that constructing a feasible CCR requires a consistent estimator of Γ. The least squares estimator, $(\hat{\Gamma}_{LS} - \Gamma) = \sum u_t^*(q_t - \bar{q}_n)'(\sum q_t(q_t - \bar{q}_n)')^{-1}$ with $\bar{q}_n \equiv (1/n)\sum q_t$, is superconsistent.

Lemma 4.2 Let [A1]-[A4] hold. Further, assume that either [NED1]-[NED2], [NED4'], and [NED5.a]-[NED5.b] or the results of Lemma A.2[d] and [f] hold. We have $(\hat{\Gamma}_{LS} - \Gamma) = O_p(1/n)$ as $n \to \infty$.

In addition to Γ, we also need consistent estimators of α and μ for consistent covariance estimation. In practice, these are estimated simultaneously with ψ and Γ. Since we have already shown that $\hat{\psi}_{LS}$ and $\hat{\Gamma}_{LS}$ are consistent, we may simply consider

$$\hat{\alpha}_{LS} = (\sum w_t w_t')^{-1} \sum w_t(y_t - \hat{\psi}'_{LS}\hat{C}x_t^*) \tag{7}$$

for α and

$$\hat{\mu}_{LS} = (1/n) \sum (x_t^* - \hat{\Gamma}_{LS} q_t) \tag{8}$$

for μ.

Lemma 4.3 Let [A1]-[A6] hold. Further, assume that either [NED1]-[NED6] or the results of Lemma A.2 hold. We have
[a] $(\hat{\alpha}_{LS} - \alpha) = O_p(n^{-1/2})$, and
[b] $(\hat{\mu}_{LS} - \mu) = O_p(n^{-1/2})$
as $n \to \infty$.

Theorem 4.1 and Lemma 4.3[a] jointly establish consistency with appropriate rates of convergence for least squares estimation of (5), while Lemma 4.2 and Lemma 4.3[b] accomplish the same for (6).[7] All further references to estimators of α, ψ, μ, and Γ pertain to the least squares estimators, unless otherwise specified, and we drop the LS subscript henceforth.

---

[6] Consistency holds under much more general assumptions, such as relaxing [A5] and [NED6], but with additional nuisance parameters.

[7] In practice, a problem may arise for some imputation techniques.We may need to impute data before

4.3 Consistent Covariance Estimation

To estimate long-run variances and covariances consistently, let $b_t^* \equiv b_t + Dz_t^*$, where D is an $R \times r$ matrix defined by $D \equiv (-C'\psi, 0, I, 0)'$. The first submatrix of zeros is $r \times p$, the second is $r \times g$, and the identity submatrix is $r \times r$. The long-run variance of (b_t) cannot be identified, but that of (b_t^*) is all that is required.

We first verify consistent covariance estimation when (b_t^*) is observable. In this case, the long-run variance estimator is

$$\tilde{\Omega}_{b^*b^*} = \frac{1}{n} \sum \sum_{s=1}^{n} b_t^* b_s^{*\prime} \pi((t-s)/\ell_n) \tag{9}$$

for some kernel function π with lag truncation parameter ℓ_n. In the absence of messy data, a vast literature on covariance estimation is available. Newey and West (1987), Andrews (1991), Hansen (1992), *inter alia*, have addressed this problem under stationarity or mixing assumptions.

Consider the class of kernel functions $\mathcal{K}$ defined by de Jong and Davidson (2000),

$$\mathcal{K} = \left\{ \pi(z) : \mathbf{R} \to [-1,1] \;\middle|\; \begin{array}{l} \pi(0) = 1,\ \pi(z) = \pi(-z) \quad \text{for all} \quad z \in \mathbf{R}, \\ \int_{-\infty}^{\infty} |\pi(z)|\, dz < \infty,\ \int_{-\infty}^{\infty} |\varpi(\xi)|\, d\xi < \infty, \text{ and} \\ \pi(z) \ \text{ is continuous at 0 and almost everywhere else,} \end{array} \right.$$

where $\varpi(\xi)$ is the Fourier transform of $\pi(z)$. For covariance estimation using (9) and throughout the paper, we assume that the kernel function and lag truncation parameter satisfy

[K1] $\lim_{n\to\infty} (1/\ell_n + \ell_n/n) = 0$,
[K2] $\pi \in \mathcal{K}$, and
[K3] $n^{-1/2} \sum_{k=0}^{n} \pi(k/\ell_n) = o(1)$.

The first condition imposes $\ell_n = o(n)$ on the lag truncation parameter. The second limits the class of admissible kernel functions, but still 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 ℓ_n, depending on the kernel function. For example, the Bartlett and Tukey-Hanning kernels require $\ell_n = o(n^{1/2})$ to satisfy [K3].

Under these additional assumptions, we have the following result.

Lemma 4.4 Let [A1], [NED1], [NED4'], [K1], and [K2] hold. We have
[a] $\tilde{\Sigma}_{b^*b^*} \to_p \Sigma_{b^*b^*}$,
[b] $\tilde{\Delta}_{b^*b^*} \to_p \Delta_{b^*b^*}$, and
[c] $\tilde{\Omega}_{b^*b^*} \to_p \Omega_{b^*b^*}$

running least squares on (6), but estimate Γ in order to conduct the imputation, if the technique employs the trends (q_t), such as Friedman's (1962) approach does. In order to circumvent this difficulty, it may be necessary to use a preliminary (but not necessarily consistent) estimator of Γ. As long as the resulting imputation error satisfies our conditions above, then Γ may then be re-estimated consistently.

as $n \to \infty$, where $\Sigma_{b^*b^*} \equiv \Sigma_{bb} + \Sigma_{b*}D + D\Sigma_{*b} + D\Sigma_{**}D$ with $\Delta_{b^*b^*}$ and $\Omega_{b^*b^*}$ defined accordingly.

Consequently, the long-run variance estimator is consistent if the I(0) series driving the model, the messy-data noise, and the model parameters are observed and known.

Of course, all of the parameters must be estimated, and the unknown sequences (v_t^*) and (u_t^*) that comprise (Dz_t^*) must be estimated. Simple feasible estimators of (v_t^*) and (u_t^*) are given by

$$\hat{v}_t^* = y_t - \hat{\alpha}' w_t - \hat{\psi}' \hat{C} x_t^* \quad \text{and} \quad \hat{u}_t^* = x_t^* - \hat{\mu} - \hat{\Gamma} q_t, \tag{10}$$

or simply by $\hat{v}_t^* = \hat{v}_t - \hat{\psi}' \hat{C} z_t^*$ and $\hat{u}_t^* = \hat{u}_t + z_t^*$, where

$$\hat{v}_t \equiv v_t + (\alpha - \hat{\alpha})' w_t + (\psi' C - \hat{\psi}' \hat{C}) x_t \quad \text{and} \quad \hat{u}_t \equiv u_t + (\mu - \hat{\mu}) + (\Gamma - \hat{\Gamma}) q_t. \tag{11}$$

The series $(\hat{b}_t^*)$ may thus be defined by

$$\hat{b}_t^* \equiv \hat{b}_t + \hat{D} z_t^* \tag{12}$$

where naturally $\hat{w}_t \equiv w_t$, $\triangle \hat{q}_t \equiv \triangle q_t$, and $\hat{D}$ is defined by replacing ψ and C in D with $\hat{\psi}$ and $\hat{C}$. The feasible long-run variance estimator

$$\hat{\Omega}_{b^*b^*} \equiv \frac{1}{n} \sum \sum_{s=1}^{n} \hat{b}_t^* \hat{b}_s^{*\prime} \pi \left((t-s)/\ell_n \right) \tag{13}$$

may replace the infeasible estimator in (9). Using a change of indices, symmetry of the kernel function, and $\pi(0) = 1$, we may also write this as $\hat{\Omega}_{b^*b^*} = \hat{\Delta}_{b^*b^*} + \hat{\Delta}'_{b^*b^*} - \hat{\Sigma}_{b^*b^*}$, where

$$\hat{\Delta}_{b^*b^*} \equiv \frac{1}{n} \sum_{k=0}^{n} \pi\left(k/\ell_n\right) \sum_{t=k+1}^{n} \hat{b}_t^* \hat{b}_{t-k}^{*\prime} \quad \text{and} \quad \hat{\Sigma}_{b^*b^*} \equiv \frac{1}{n} \sum \hat{b}_t^* \hat{b}_t^{*\prime} \tag{14}$$

are feasible estimators used to estimate the one-sided long-run and contemporaneous variances, respectively.

Lemma 4.5 Let [A1]-[A6], [NED1], [NED4'], and [K1]-[K3] hold. Further, assume that either [NED2]-[NED6] or the results of Lemma A.2 hold, and consider the least squares estimators $\hat{\alpha}$, $\hat{\psi}$, $\hat{\mu}$, and $\hat{\Gamma}$. We have
[a] $\hat{\Sigma}_{b^*b^*} \to_p \Sigma_{b^*b^*}$,
[b] $\hat{\Delta}_{b^*b^*} \to_p \Delta_{b^*b^*}$, and
[c] $\hat{\Omega}_{b^*b^*} \to_p \Omega_{b^*b^*}$
as $n \to \infty$.

The natural feasible estimators described by (13) and (14) are consequently consistent.

5. CAMN Estimation of β Using a CCR

A properly constructed CCR may achieve CAMN estimation with messy regressors – even with mildly nonstationary messy-data noise. We first show how to construct the CCR under ideal (infeasible) conditions, since this model differs somewhat from that considered by Park (1992). Once we have an ideal estimator, the path to feasible estimation is clear.

5.1 An Infeasible CCR

Let $\Delta_{bq} \equiv (\delta'_{vq}, \Delta'_{wq}, \Delta'_{uq}, \Delta'_{qq})'$ be the $R \times g$ matrix formed by the columns corresponding to $(\triangle q_t)$ (the last g columns) in the one-sided long-run variance of (b_t). This matrix may be interpreted as the one-sided long-run covariance between (b_t) and $(\triangle q_t)$. Similarly, let $\Sigma_{ub} \equiv (\sigma_{uv}, \Sigma_{uw}, \Sigma_{uu}, \Sigma_{uq})$ be the $r \times R$ matrix representing the contemporaneous covariance between (u_t) and (b_t). Define κ to be an $R \times 1$ vector given by $\kappa \equiv (1, 0, 0, -\omega_{vq}\Omega_{qq}^{-1})'$, where the first and second zeros are $1 \times p$ and $1 \times r$ vectors of zeros, respectively. Now let

$$x_t^{**} \equiv x_t - (\Gamma\Delta'_{bq} + \Sigma_{ub})\Sigma_{bb}^{-1} b_t$$

and

$$y_t^{**} \equiv y_t - \psi' C(\Gamma\Delta'_{bq} + \Sigma_{ub})\Sigma_{bb}^{-1} b_t - \omega_{vq}\Omega_{qq}^{-1}\triangle q_t,$$

so that we estimate $y_t^{**} = \alpha' w_t + \psi' \hat{C} x_t^{**} + v_t^{**}$ in place of (3), where $v_t^{**} \equiv b_t'\kappa$.

Since we already have a consistent estimator of α, we may simply run least squares on

$$y_t^{**} - \hat{\alpha}' w_t = \psi' \hat{C} x_t^{**} + v_t^{**} \tag{15}$$

for CAMN estimation of ψ.

5.2 A Feasible CCR

A feasible estimation procedure must overcome not only the usual obstacles of unknown nuisance parameters and unobserved error sequences. We face the additional obstacle of messy data, and the variance estimators described above are contaminated by this messiness, as is clear from the limits of Lemmas 4.4 and 4.5. All of the variance estimators below are defined as submatrices, vectors, or individual elements of $\hat{\Sigma}_{b^*b^*}$, $\hat{\Delta}_{b^*b^*}$, and $\hat{\Omega}_{b^*b^*}$. For notational simplicity, we drop the $*$ superscripts in the subscripts of the variance estimators throughout the rest of the paper. For example, $\hat{\Sigma}_{bb}$ denotes $\hat{\Sigma}_{b^*b^*}$ and $\hat{\omega}_{vq}$ denotes $\hat{\omega}_{v^*q}$. (Note that the probability limit of these feasible estimators are *not* Σ_{bb} and ω_{vq}. Rather the limiting variances are $\Sigma_{b^*b^*}$ and $\omega_{vq} - \psi' C\Omega_{*q}$.)

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

$$x_t^{**} \equiv x_t^* - (\hat{\Gamma}\hat{\Delta}'_{bq} + \hat{\Sigma}_{ub})\hat{\Sigma}_{bb}^{-1}\hat{b}_t^* \tag{16}$$

and

$$y_t^{**} \equiv y_t - \hat{\psi}'\hat{C}(\hat{\Gamma}\hat{\Delta}'_{bq} + \hat{\Sigma}_{ub})\hat{\Sigma}_{bb}^{-1}\hat{b}_t^* - \hat{\omega}_{vq}\hat{\Omega}_{qq}^{-1}\triangle q_t, \tag{17}$$

where all variances and covariances are estimated simultaneously using a single nonparametric procedure as discussed above. The error term in (15) is now

$$v_t^{**} = b_t^{*\prime}\hat{\kappa}, \tag{18}$$

and $(\hat{\psi}_{CCR} - \psi) = (\hat{C}\sum x_t^{**} x_t^{**\prime}\hat{C}')^{-1}\hat{C}\sum x_t^{**}(v_t^{**} + (\alpha - \hat{\alpha})' w_t)$ is the least squares estimator of this feasible CCR.

Define Brownian motion $V_{\perp Q}^{*}(s) \equiv (V(s) - \psi' C Z^{*}(s)) - (\omega_{vq} - \psi' C \Omega_{*q}) \Omega_{qq}^{-1} Q(s)$, which may be interpreted as the projection of $V(s) - \psi' C Z^{*}(s)$ onto the space orthogonal to $Q(s)$. Such Brownian motion has variance

$$\begin{aligned}\mathbf{var}(V_{\perp Q}^{*}(1)) &= (\omega_{vv} - \omega_{vq}\Omega_{qq}^{-1}\omega_{qv}) + \psi' C(\Omega_{**} - \Omega_{*q}\Omega_{qq}^{-1}\Omega_{q*})C'\psi \\ &\quad - \psi' C(\omega_{*v} - \Omega_{*q}\Omega_{qq}^{-1}\omega_{qv}) - (\omega_{v*} - \omega_{vq}\Omega_{qq}^{-1}\Omega_{q*})C'\psi,\end{aligned}$$

where the first term is the long-run variance in the standard CCR model.

Theorem 5.1 Let [A1]-[A6], [NED1], [NED4'], and [K1]-[K3] hold. Further, assume that either [NED2]-[NED6] or the results of Lemma A.2 hold, and consider estimators $\hat{\alpha}$, $\hat{\psi}$, $\hat{\mu}$, and $\hat{\Gamma}$ defined by the least squares estimators above. We have $n(\hat{\psi}_{CCR} - \psi) \rightarrow_d (CNC')^{-1} C \Gamma \int Q dV_{\perp Q}^{*}(s)$ as $n \rightarrow \infty$.

Since the variance of this distribution conditional on Q is $\mathbf{var}(V_{\perp Q}^{*}(1)) CNC'$, the asymptotic distribution may be rewritten as $\mathbf{var}(V_{\perp Q}^{*}(1))^{1/2} (CNC')^{-1/2\prime} \mathbf{N}(0, I_g)$, where $(CNC')^{-1/2}$ is the inverse of the Cholesky decomposition of CNC'. This distribution is simply a $g \times 1$ vector of mixed normal variates. Note that this variance has the standard least squares form, so that standard errors and test statistics from standard software packages are asymptotically valid.

Since $\hat{C} \rightarrow_p C$, we may recover the asymptotic distribution of $\hat{\beta}_{CCR} = \hat{C}' \hat{\psi}_{CCR}$, which is $\mathbf{var}(V_{\perp Q}^{*}(1))^{1/2} (C'(CNC')^{-1} C)^{1/2} \mathbf{N}(0, I_r)$, an $r \times 1$ vector of mixed normal variates. Estimates and standard errors from software packages must be transformed accordingly. If C is chosen to be $C = (\Gamma'\Gamma)^{-1}\Gamma'$, this distribution is $\mathbf{var}(V_{\perp Q}^{*}(1))^{1/2} (C'(\int QQ')^{-1} C)^{1/2} \mathbf{N}(0, I_r)$, so that the standard errors are proportional to the contributions of each element of (x_t) to the stochastic trend(s) of (x_t).

6. Lerp Revisited: Large- and Small-Sample Results

At least two concerns remain for the econometrician. Of primary concern, the sufficient conditions of either [NED2]-[NED6] or the results of Lemma A.2, [NED1], and [NED4'] should be verified for a particular messy-data generating process, in order for the results to hold in large samples. Secondarily, although the bias from messiness (indeed, even the bias from serial correlation without the added complication of messiness) is overcome in large samples, we should question to what extent the asymptotic approximation is valid in small samples. We briefly address these two concerns for the specific example of linear interpolation in a mixed-frequency setting.

Using the feasible system given by (5) and (6), we may write the univariate messy-data noise from linear interpolation in (2) as

$$\begin{aligned} z_t^{*} &= (u_{\tau_{p-1}} - u_{\tau_{p-1}+j}) + \frac{j}{m+1}(u_{\tau_p} - u_{\tau_{p-1}}) \\ &\quad - \Gamma\left(\sum_{i=1}^{j} \triangle q_{\tau_{p-1}+i} - \frac{j}{m+1}\sum_{i=1}^{m+1} \triangle q_{\tau_{p-1}+i}\right) \end{aligned} \tag{19}$$

Sufficient conditions are verified in the proof of the following proposition.

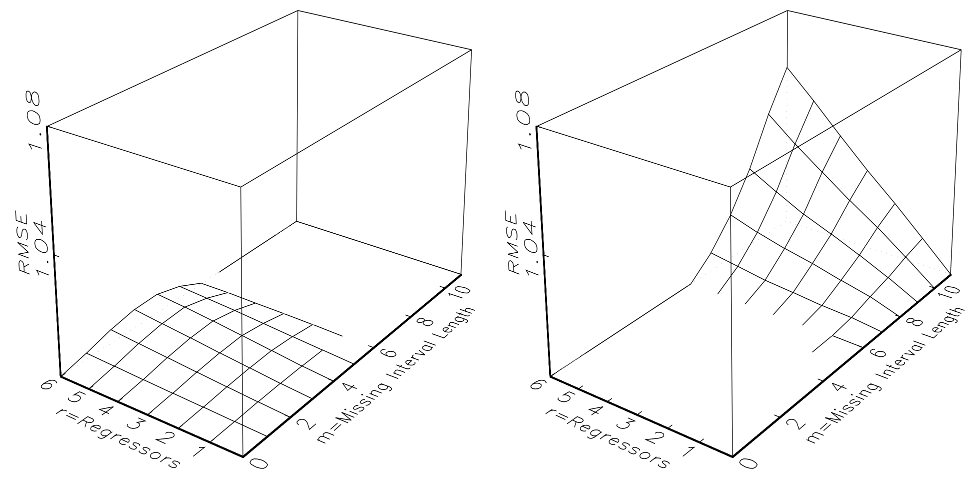

Figure 1: Relative average RMSE using cointegrated series: $\max(RMSE_{omit}/RMSE_{lerp}, 1)$ (left panel) and $\max(RMSE_{lerp}/RMSE_{omit}, 1)$ (right panel).

Proposition 6.1 Let [A1]-[A4] hold. Define (z_t^*) by (19) and assume that the sequence (w_t) is contemporaneously and serially uncorrelated with (u_t) and (q_t). Then the results of Lemma A.2, [NED1], and [NED4'] hold.

This proposition (in conjunction with assumptions [A5] and [A6]) validates CCR with linearly interpolated series, and is suggestive for other covariance-based techniques and for more sophisticated imputation techniques.

In addressing small-sample concerns in a mixed-frequency setting, we compare a messy-data generating imputation technique (lerp, in particular) with estimation at the lower frequency (data omission). Adding small amounts of bias and inefficiency through imputation may be preferable to the larger inefficiency resulting from data omission. Using root mean-squared error (RMSE), simulations illustrate this bias/variance trade-off.

For a sample size of 480, we simulate $6 \times 11 = 66$ models $5,000$ times each, differentiated by $r = 1, \dots, 6$ and $m = 1, \dots, 11$ (holding the number of regressors observed at the lower frequency to 1).[8] Variances and covariances are estimated using a Bartlett kernel.

The left panels of Figures 1 and 2 show $\max(RMSE_{omit}/RMSE_{lerp}, 1)$, while the right panels show $\max(RMSE_{lerp}/RMSE_{omit}, 1)$. Regions over which lerp is preferable to omission (lerp has a lower RMSE) appear on the left, while regions over which lerp is not preferable to omission (lerp has a higher RMSE) appear on the right.

Figure 1 suggests that linear interpolation is preferable to high-frequency omission when m is relatively small. At around $m = 6$ (every 7^{th} observation is non-missing), we are indifferent between these two simple methods. For $m > 6$ omission is preferable. If the higher frequency is monthly, these results suggest that linearly interpolating quarterly data

[8]We generate regressors with serially dependent but contemporaneously independent increments. Specifically, $(v_t, \triangle x_t')'$ is a VAR(1) with an $(r+1) \times (r+1)$ diagonal autoregressive matrix with $1/2$ along the diagonals and $\Sigma = I$. We set β to be an $r \times 1$ vector of ones.

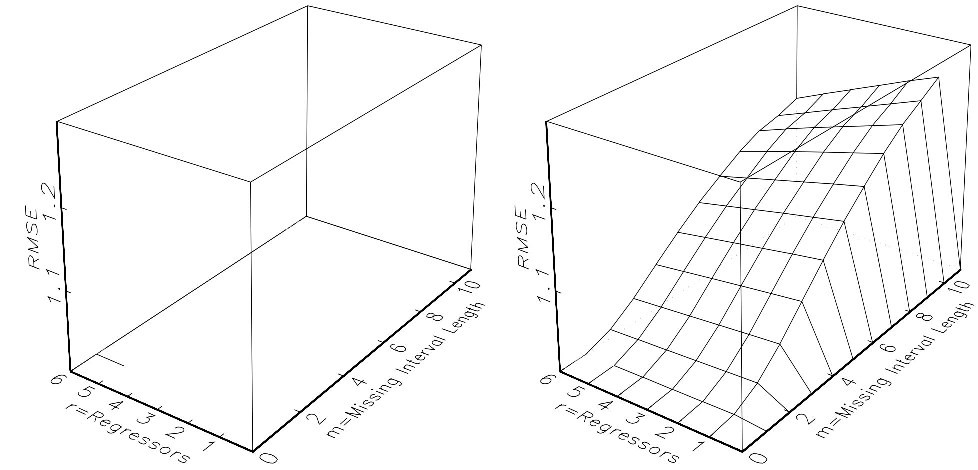

Figure 2: Relative average RMSE using stationary series: $\max\left(RMSE_{omit}/RMSE_{lerp}, 1\right)$ (left panel) and $\max\left(RMSE_{lerp}/RMSE_{omit}, 1\right)$ (right panel).

($m = 2$) is more appropriate than estimating the whole model at the quarterly frequency. However, linearly interpolating annual data ($m = 11$) is inferior to estimation at the annual frequency. Again, these results are suggestive for more sophisticated imputation techniques.

In order to emphasize the specificity of this result to cointegrated series, we repeat a similar exercise with stationary series using least squares.[9] In stark contrast, Figure 2 confirms conventional wisdom: interpolation is *almost never* preferable to data omission for stationary models, although the RMSE's are close when m is only 1 or 2.

7. Concluding Remarks

We have shown that covariance-based methods for estimating cointegrating regressions (CCR, in particular) may be valid even when the error term is not covariance stationary. Although consistency of least squares may be intuitive in this context, asymptotic normality and unbiasedness of multi-step covariance-based estimators is not obvious. We have rigorously shown that in fact these properties hold under general conditions about the rates of convergence of the error terms, even in cases where these errors have time-varying variances. The well-known covariance-based techniques are robust.

The importance of these results is underscored by the prevalence of data irregularities, such as missingness, mixed frequency, and measurement error, which create additional noise that may violate covariance stationarity assumptions on the model error. Our theoretical results may hold for a wide variety of techniques for handling data irregularities.

[9] We let $(v_t) \sim \mathbf{iidN}\,(0, I)$ with (x_t) a VAR(1) with 1/2 along the diagonals, independent of (v_t).

Appendix A: Useful Lemmas and Their Proofs

Throughout the proofs, the notation e_i is employed for a conformable vector that has a one in the i^{th} row and zeros elsewhere – i.e., the i^{th} column of an identity matrix of appropriate dimension. Also, we let $\zeta_t \equiv (b_t', z_t^{*\prime})'$ be an $(R+r)\times 1$ vector.

Lemma A.1 For sequences (b_t) satisfying [A1] and (z_t^*) satisfying [NED1] and [NED4'],
[a] $(z_t^* z_t^{*\prime})$ is a matrix of sequences that are L_1-NED on (b_t),
[b] $(z_t^* b_t')$ is a matrix of sequences that are L_1-NED on (b_t), and
[c] (ζ_t) is a vector of sequences that are L_2-NED on (b_t),
all of which have a size of -1 and are defined w.r.t. bounded sequences of constants.

Proof of Lemma A.1 Under [NED1], each element (z_{it}^*) of (z_t^*) is L_2-NED on (b_t), which implies that an arbitrary element $(z_{it}^* z_{jt}^*)$ of $(z_t^* z_t^{*\prime})$ is L^1-NED (with the same size) w.r.t. constants defined by $\max(||z_{jt}^*||_2 d_{jt}^z, \|z_{it}^*\|_2\, d_{it}^z, d_{it}^z d_{jt}^z)$ from the proof of Theorem 17.9 of Davidson (1994), where (d_{it}^z) and (d_{jt}^z) are the sequences of constants from the definition of near-epoch dependence for (z_{it}^*) and (z_{jt}^*), respectively. 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).[10] The filtration in the definition is simply defined to be the natural filtration, and the rest of the proof follows that of part [a]. The proof of part [c] is trivial and therefore omitted. □

Lemma A.2 Let [NED1] hold. Then
[a] $n^{-1}\sum z_t^* = o_p(1)$ under [NED2], [NED4'],
[b] $n^{-1}\sum (z_t^* z_t^{*\prime} - \Sigma_{**}) = o_p(1)$ under [NED3.a], [NED4],
[c] $n^{-1}\sum (z_t^* b_t' - \Sigma_{*b}) = o_p(1)$ under [A1], [NED3.b], [NED4],
[d] $n^{-1/2}\sum z_t^* = O_p(1)$ under [NED2], [NED4'], [NED5.a],
[e] $n^{-1/2}\sum z_t^* w_t' = O_p(1)$ under [NED4], [NED6.a], [NED6.b], and
[f] $n^{-1}\sum q_t z_t^{*\prime} \to_d \int Q dZ^*(s)' + \Delta_{*q}'$ under [A1]-[A4], [NED2], [NED4'], [NED5.a]-[NED5.b]
as $n\to\infty$.

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 a functional central limit theorem to prove part [f]. For parts [d]-[f], all stochastic arrays are created by dividing the underlying stochastic sequences by $\sqrt{n}$.
[a] Consider an arbitrary element (z_{it}^*) of (z_t^*). Letting (d_{it}^z) denote the sequence of constants in the definition, clearly we have $d_{it}^z = O\left(\|z_{it}^*\|_2\right)$ as $n\to\infty$, since both sides are bounded under [NED1] and [NED4']. A sufficient condition for Theorem 3.3 of Davidson and de Jong (1997), in order to obtain the rate of convergence in our stated result, is that $t^{-1}\|z_{it}^*\|_2 =$

[10] The reader is referred to Example 17.3 in Davidson (1994).

$O(t^{-5/6-\varepsilon})$ for some $\varepsilon > 0$,[11] since (z_t^*) has mean zero by [NED2]. Under [NED4'], this condition holds for any $\varepsilon \leq 1/6$. The result trivially extends to the entire vector (z_t^*).
[b] Consider an arbitrary element $(z_{it}^* z_{jt}^*)$ of $(z_t^* z_t^{*\prime})$, which is L_1-NED of size -1 on (b_t) w.r.t. a bounded sequence of constants by Lemma A.1[a]. Note that $||z_{i,t}^* z_{jt}^* - e_i' \Sigma_{**} e_j||_2 \leq ||z_{it}^*||_4 ||z_{jt}^*||_4 + e_i' \Sigma_{**} e_j$ by the Minkowski and Cauchy-Schwarz inequalities. Since $\|z_t^*\|_4$ is bounded by [NED4], and $e_i' \Sigma_{**} e_j$ is bounded by [NED3.a], we may employ Theorem 3.3 of Davidson and de Jong (1997). Again, the 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 A.1[a] with A.1[b], [NED3.a] with [NED3.b], and the addition of [A1] to ensure that A.1[b] holds and that the fourth moment of (b_t) is finite.
[d] If $\Omega_{**}(s) = 0$, then (z_t^*) must have a degenerate variance and degenerate autocovariances, so the result trivially holds. More generally, let $\Omega_{**}(s) > 0$. Consider a random vector $n^{-1/2} \Omega_{**}^{-1/2} z_t^*$ constructed from (z_t^*) and the Cholesky decomposition of the inverse of Ω_{**}. The i^{th} element of this random vector is $n^{-1/2} e_i' \Omega_{**}^{-1/2} z_t^*$. If the L_2-norm of this element is unity, the conditions for Corollary 1 of de Jong (1997) are satisfied for constants $c_{nt} \equiv n^{-1/2}$ and under [NED1], [NED2], and [NED4'] . For verification, note that

$$\mathbf{E} \left| \sum n^{-1/2} e_i' \Omega_{**}^{-1/2} z_t^* \right|^2 = e_i' \Omega_{**}^{-1/2} \left(\mathbf{E} Z^* Z^* (1)' \right) \Omega_{**}^{-1/2\prime} e_i$$

which under [NED5.a] is in fact unity for each i.
[e] The proof proceeds as in part [d], by looking the random vector $n^{-1/2} \Omega_{*w_j*w_j}^{-1/2} z_t^* w_t' e_j$, which corresponds to the j^{th} column of $z_t^* w_t'$. To consider the i^{th} element of this vector, we look at $n^{-1/2} e_i' \Omega_{*w_j*w_j}^{-1/2} z_t^* w_t' e_j$, which has an L_2-norm of

$$\mathbf{E} \left| \sum n^{-1/2} e_i' \Omega_{*w_j*w_j}^{-1/2} z_t^* w_t' e_j \right|^2 = e_i' \Omega_{*w_j*w_j}^{-1/2} \left(\mathbf{E} Z^* W_{n,j} \left(Z^* W_{n,j} \right)' (1) \right) \Omega_{*w_j*w_j}^{-1/2\prime} e_i$$

which again is unity under [NED6.b]. We only need to show that $\sup_t \|z_t^* w_t' / d_t^{zw}\|_{2a/(a-1)} < \infty$, where (d_t^{zw}) is the bounded sequence of constants defined implicitly in Lemma A.2[b]. A sufficient condition is that element of both (z_t^*) and (w_t) have finite moments up to $4a/(a-1)$, which we assume in [NED4] and [A1].
[f] We employ Theorem 30.14 of Davidson (1994). We may write

$$\frac{1}{n} \sum q_t z_t^{*\prime} = (0,0,0,I,0) \frac{1}{n} \sum \sum_{i=1}^{t} \zeta_i \zeta_t' (0,0,0,0,I)',$$

where the $(0,0,0,I,0)$ is a $g \times (R+r)$ matrix with $g \times g$ identity submatrix after column $R-g$ and $(0,0,0,0,I)$ is a $g \times (R+r)$ matrix with $r \times r$ identity submatrix after column R, and show that the sufficient conditions for Theorem 29.6 and Corollary 29.14 of Davidson (1994) hold for $e_j' \zeta_t$ with $j = 1, \ldots, R+r$. Condition [a] of Theorem 29.6 is jointly satisfied by [A1] and [NED2]. Condition [b] requires that $\sup_t \|z_t^* / d_t^z\|_{2a/(a-1)} < \infty$, which is satisfied by [NED4'], since constants (d_t^z) may be chosen to be nonzero w.l.o.g. Conditions [c] and [e]

[11] 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 $a > 1$.

of Theorem 29.6 follow directly from Lemma A.1[c]. Condition [d] of this theorem is also satisfied by the boundedness of (d_t^z) in [NED1]. Conditions [A1]-[A3], [NED5.a] and [NED5.b] jointly satisfy condition [f'] of Corollary 29.14 (and condition [b] of Theorem 29.18), because in order for $\mathbf{E}(Q_n(s)', Z_n^*(s)')'(Q_n(s)', Z_n^*(s)')$ to have a well-defined limit, $\mathbf{E}Q_nQ_n'(s)$, $\mathbf{E}Z_n^*Q_n'(s)$, and $\mathbf{E}Z_n^*Z_n^{*\prime}(s)$ must have finite limits. □

Lemma A.3 Let [A1]-[A3] hold. Further, assume that either [NED1]-[NED5] or the results of Lemma A.2[a]-[c] and [f] hold. We have
[a] $n^{-1}\sum x_t^* v_t^* \to_d M^*$,
[b] $n^{-1}\sum x_t^* w_t' \to_d \Gamma\left(\int QdW(s)' + \Delta_{wq}'\right) + \Sigma_{uw} + \Sigma_{*w}$,
[c] $n^{-1}\sum w_t v_t^* \to_p \sigma_{wv} - \Sigma_{w*}C'\psi$, and
[d] $n^{-2}\sum x_t^* x_t^{*\prime} \to_d N$
as $n \to \infty$.

Proof of Lemma A.3 Since [NED1]-[NED5] are sufficient for Lemma A.2[a]-[c] and [f], we prove the lemma using the latter.
[a] We may rewrite the summation in terms of (x_t), (v_t), and (z_t^*) using (1) and our definition of (v_t^*). Expanding the product yields

$$\sum x_t^* v_t^* = \sum x_t v_t - \sum x_t z_t^{*\prime} C'\psi + \sum z_t^* v_t - \sum z_t^* z_t^{*\prime} C'\psi, \tag{20}$$

and to find the limiting distribution of the first term of (20), we may further expand this term using the data generating process of (x_t) given by (4). We obtain

$$\frac{1}{n}\sum (\mu + \Gamma q_t + u_t) v_t \to_d \Gamma\left(\int QdV(s) + \delta_{vq}'\right) + \sigma_{uv}$$

as $n \to \infty$. (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 (20) has a distribution given by

$$-\frac{1}{n}\sum (\mu + \Gamma q_t + u_t) z_t^{*\prime} C'\psi \to_d -\Gamma\left(\int QdZ^*(s)' + \Delta_{*q}'\right) C'\psi - \Sigma_{u*}C'\psi$$

using Lemma A.2[a], [f], and [c], respectively. The third and fourth terms of (20) are similarly governed by Lemma A.2[c] and [b], respectively, so that the stated result is obtained.
[b] As in the proof of part [a], we use (1) to write $\sum x_t^* w_t' = \sum x_t w_t' + \sum z_t^* w_t'$, and the stated result follows along similar lines.
[c] Expanding the summation in part [c] yields $\sum w_t v_t^* = \sum w_t v_t - \sum w_t z_t^{*\prime} C'\psi$, and, again, the stated result immediately follows.
[d] Finally, expanding the summation in part [d] reveals a structure similar to part [a]. Specifically,

$$\sum x_t^* x_t^* = \sum x_t x_t' + \sum z_t^* z_t^{*\prime} + \sum x_t z_t^{*\prime} + \sum z_t^* x_t',$$

where the first term has an asymptotic distribution of $n^{-2}\sum x_t x_t' \to_d \Gamma \int QQ'\Gamma'$, which dominates under our conditions. □

Lemma A.4 Let [A1]-[A6], [NED1], [NED4'], and [K1]-[K3] hold. Further, assume that either [NED2]-[NED6] or the results of Lemma A.2 hold, and consider estimators $\hat{\alpha}$, $\hat{\psi}$, $\hat{\mu}$, and $\hat{\Gamma}$ defined by the least squares estimators above. We have
[a] $n^{-1}\sum x_t^{**}v_t^{**} \rightarrow_d \Gamma \int QdV_{\perp Q}^{*}(s)$
[b] $n^{-1}\sum x_t^{**}w_t' \rightarrow_d \Gamma \int QdW(s)'$, and
[c] $n^{-2}\sum x_t^{**}x_t^{**\prime} \rightarrow_d \Gamma \int QQ'\Gamma'$
as $n \rightarrow \infty$.

Proof of Lemma A.4 Again, we use Lemma A.2 rather than [NED1]-[NED6] for the proofs.
[a] The summation may be expanded using (16) and (18) as

$$\sum x_t^* v_t^* - \sum x_t^* \triangle q_t' \hat{\Omega}_{qq}^{-1}\hat{\omega}_{qv} - (\hat{\Gamma}\hat{\Delta}_{bq}' + \hat{\Sigma}_{ub})\hat{\Sigma}_{bb}^{-1}\sum \hat{b}_t^* b_t^{*\prime}\hat{\kappa} \tag{21}$$

using feasible estimators of all parameters. The distribution of the first term of (21) follows from Lemma A.3[a]. The second term of (21) may be written as

$$-\left(\sum x_t \triangle q_t' + \sum z_t^* \triangle q_t'\right)\hat{\Omega}_{qq}^{-1}\hat{\omega}_{qv}, \tag{22}$$

where the variance estimators have a limiting distribution of $\hat{\Omega}_{qq}^{-1}\hat{\omega}_{qv} \rightarrow_p \Omega_{qq}^{-1}(\omega_{qv} - \Omega_{q*}C'\psi)$ by Lemma 4.5. When normalized by $1/n$, the first summation in (22) has an asymptotic distribution given by $\Gamma(\int QdQ(s)' + \Delta_{qq}') + \Sigma_{uq}$, and the probability limit of the second summation in (22) is Σ_{*q} when similarly normalized. To determine the limit of the final term of (21), we need to deal with the limit of $\sum \hat{b}_t^* b_t^{*\prime}$. Expanding this as

$$\sum \hat{b}_t b_t' + \hat{D}\sum z_t^* b_t' + \sum \hat{b}_t z_t^{*\prime} D' + \hat{D}\sum z_t^* z_t^{*\prime} D',$$

it is clear using (11) that this consistently estimates $\Sigma_{b^*b^*}$, as does $\hat{\Sigma}_{bb}$. We may thus write this term as $-(\hat{\Gamma}\hat{\delta}_{vq}' + \hat{\Sigma}_{uv}) + (\hat{\Gamma}\hat{\Delta}_{qq}' + \hat{\Sigma}_{uq})\hat{\Omega}_{qq}^{-1}\hat{\omega}_{qv} + o_p(1)$, where

$$(\hat{\Gamma}\hat{\delta}_{vq}' + \hat{\Sigma}_{uv}) \rightarrow_p \Gamma\left(\delta_{vq}' - \Delta_{*q}'C'\psi\right) + (\sigma_{uv} - \Sigma_{u*}C'\psi) + (\sigma_{*v} - \Sigma_{**}C'\psi)$$

and $(\hat{\Gamma}\hat{\Delta}_{qq}' + \hat{\Sigma}_{uq}) \rightarrow_p \Gamma\Delta_{qq}' + \Sigma_{uq} + \Sigma_{*q}$ as $n \rightarrow \infty$. Combining all of these terms (after appropriate cancellations) yields the stated result.
[b] Using (16), the summation is equal to $\sum x_t^* w_t' - (\hat{\Gamma}\hat{\Delta}_{bq}' + \hat{\Sigma}_{ub})\hat{\Sigma}_{bb}^{-1}\sum \hat{b}_t^* \hat{b}_t^{*\prime}(0, I, 0, 0)'$, where $(0, I, 0, 0)$ is a $p \times R$ matrix with $p \times p$ identity submatrix after the first column, and where the distribution of the first term comes from Lemma A.3[b]. The second term is also $O_p(n)$ by Lemma 4.5, with a probability limit given by $-\left(\Gamma\Delta_{wq}' + \Sigma_{uw} + \Sigma_{*w}\right)$.
[c] Expanding the summation yields

$$\begin{aligned} &\sum x_t^* x_t^{*\prime} + (\hat{\Gamma}\hat{\Delta}_{bq}' + \hat{\Sigma}_{ub})\hat{\Sigma}_{bb}^{-1}\sum \hat{b}_t^* \hat{b}_t^{*\prime}\hat{\Sigma}^{-1}(\hat{\Gamma}\hat{\Delta}_{bq}' + \hat{\Sigma}_{ub})' \\ &- \sum x_t^* \hat{b}_t^{*\prime}\hat{\Sigma}_{bb}^{-1}(\hat{\Gamma}\hat{\Delta}_{bq}' + \hat{\Sigma}_{ub})' - (\hat{\Gamma}\hat{\Delta}_{bq}' + \hat{\Sigma}_{ub})\hat{\Sigma}_{bb}^{-1}\sum \hat{b}_t^* x_t^{*\prime} \end{aligned} \tag{23}$$

where the asymptotics of the first term are derived in Lemma A.3[d]. It remains to show that the other terms are $o_p\left(n^2\right)$. This is clearly true for the second term, which is $O_p\left(n\right)$ as a direct result of Lemma 4.5. The summation $\sum x_t^* \hat{b}_t^{*\prime}$ in the third and fourth terms of (23) may be partitioned as

$$\sum \left(x_t^* \hat{v}_t^*, x_t^* w_t', x_t^* \hat{u}_t^{*\prime}, x_t^* \triangle q_t'\right), \tag{24}$$

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

$$\sum x_t^* v_t^* + \sum x_t^* w_t' \left(\alpha - \hat{\alpha}\right) + \sum x_t^* x_t^{*\prime} (\psi' C - \hat{\psi}' \hat{C})$$

These are clearly no more than $O_p\left(n\right)$ from Lemma A.3[a], [b], [d], and the fact that $(\psi' C - \hat{\psi}' \hat{C}) = O_p\left(1/n\right)$ (see the proof of Lemma 4.3 below). The second partition is obviously $O_p\left(n\right)$ as a special case of the first. The third partition in (24) admits the expansion

$$\begin{aligned} &\sum x_t u_t' + \sum z_t^* u_t' + \sum x_t \left(\mu - \hat{\mu}\right)' + \sum z_t^* \left(\mu - \hat{\mu}\right)' \\ &+ \sum x_t q_t' (\Gamma - \hat{\Gamma})' + \sum z_t^* q_t' (\Gamma - \hat{\Gamma})' + \sum x_t z_t^{*\prime} + \sum z_t^* z_t^{*\prime} \end{aligned}$$

using (1) and (11). All of these are $o_p\left(n^2\right)$ under our assumptions. The last partition in (24) is $O_p\left(n\right)$ for the same reasons as the second partition. Finally, returning to the third and fourth terms of (23), since $\hat{\Sigma}^{-1}(\hat{\Gamma}\hat{\Delta}_{bq}' + \hat{\Sigma}_{ub})' = O_p\left(1\right)$, the proof is complete. □

Lemma A.5 Let [A1]-[A3] hold and define $\Sigma_{bb}\left(k\right) = \mathbf{E} b_t b_{t-k}'$ to be the k^{th} autocovariance of (b_t). For any i, j,[12] we have $\lim_{l\to\infty} \mathbf{cov}(B_l^{(j)}\left(s\right), B_l^{(j-i)}\left(s\right)) = s\Omega_{bb}^{(i)}$ as $l \to \infty$, where $B_l^{(j)}\left(s\right) \equiv l^{-1/2} \sum_{p=1}^{[ls]} b_{\tau_{p-1}+j}$ and $\Omega_{bb}^{(i)} \equiv \sum_{k=-\infty}^{\infty} \Sigma_{bb}\left(k\left(m+1\right)+i\right)$ are a stochastic process and limiting variance, respectively.

Proof of Lemma A.5 The covariance is equal to

$$\frac{1}{l}\mathbf{E}\left(\sum\nolimits_{p=1}^{[ls]} b_{\tau_p+j} \sum\nolimits_{p=1}^{[ls]} b_{\tau_p+j-i}'\right) = \sum\nolimits_{k=-[ls]}^{[ls]} \left(\frac{[ls]}{l} - \frac{|k|}{l}\right) \Sigma_{bb}\left(k\left(m+1\right)+i\right)$$

by the stationarity of (b_t). The result follows from the Kronecker lemma and the summability of the autocovariances implied by [A1]. □

Lemma A.6 Let [A1]-[A3] hold. Define a Brownian motion

$$Z^*\left(s\right) = \left(m+1\right)^{1/2} U^{(0)}\left(s\right) - U\left(s\right) + \frac{\Gamma}{\left(m+1\right)^{1/2}} \sum\nolimits_{j=1}^{m+1} \sum\nolimits_{i=j+1}^{m+1} Q^{(i)}\left(s\right) - \frac{m}{2}\Gamma Q\left(s\right)$$

and the stochastic processes $Z_n^*\left(s\right)$ as above and with (z_t^*) defined by (19). We have

[12]The result holds for any i, j, although we generally restrict $j \in [1, m+1]$, due to construction of the messy data error.

[a] $B_l^{(j)}(s) \to_d B^{(j)}(s)$, a Brownian motion with variance $\Omega_{bb}^{(0)}$,
[b] $\sum_{j=1}^{m+1} B_l^{(0)}(s) \to_d (m+1) B^{(0)}(s)$,
[c] $\sum_{j=1}^{m+1} B_l^{(j)}(s) \to_d (m+1)^{1/2} B(s)$, and
[d] $Z_n^*(s) \to_d Z^*(s)$
as $n \to \infty$.

Proof of Lemma A.6 The stochastic process $B_l^{(j)}(s)$ is closely related to $B_n(s)$, in that increments of the former form a subset of the set of increments of the latter. Together with our assumption that $B_n(s) \to_d B(s)$, this implies that $B_l^{(j)}(s)$ also converges to a Brownian motion. We only need to show that the limiting variance of the stochastic process is well-defined, which provides the variance of the Brownian motion to which it converges. This is accomplished by setting $i = 0$ in the Lemma A.5, which completes the proof of part [a].

Part [b] follows directly from part [a], by setting $j = 0$ and noting that the summation over the same Brownian motion reduces to multiplication by $m+1$.

For the proof of part [c], note that $n = l(m+1)$. We may write

$$\begin{aligned}\sum_{j=1}^{m+1} B_l^{(j)}(s) &= \left(\frac{m+1}{n}\right)^{1/2} \left(\sum_{j=1}^{m+1}\sum_{p=1}^{[ls]} b_{\tau_{p-1}+j} + \sum_{j=1}^{\bar{m}_s} b_{\tau_{[ls]}+j}\right) - l^{1/2}\sum_{j=1}^{\bar{m}_s} b_{\tau_{[ls]}+j} \\ &= (m+1)^{1/2} B_n(s) - l^{1/2}\sum_{j=1}^{\bar{m}_s} b_{\tau_{[ls]}+j}\end{aligned}$$

with $\bar{m}_s \equiv [ns] - [ls](m+1)$. If $[ls]$ is an integer, then $[ns] - [ls](m+1) = 0$. Otherwise,

$$[ns] - [ls](m+1) = [l(m+1)s] - [ls](m+1) \leq (l(m+1)s - 1) - (ls - 1)(m+1) = m$$

so that the final term is $o_p(1)$. The stated result immediately follows from [A3].

To prove part [d], first note that

$$n^{-1/2}\sum_{t=1}^{[ns]} z_t^* = n^{-1/2}\sum_{j=1}^{m+1}\sum_{p=1}^{[ls]} z_{\tau_{p-1}+j}^* + n^{-1/2}\sum_{j=1}^{\bar{m}_s} z_{\tau_{[ls]}+j}^*$$

with $\bar{m}$ defined as in part [c]. Similarly, the second term is $o_p(1)$. The dominant term may be expanded as

$$\begin{aligned} n^{-1/2}\sum_{j=1}^{m+1}\sum_{p=1}^{[ls]} z_{\tau_p+j}^* = {} & n^{-1/2}\sum_{j=1}^{m+1}\left(1-\frac{j}{m+1}\right)\sum_{p=1}^{[ls]} u_{\tau_{p-1}} \\ & + n^{-1/2}\sum_{j=1}^{m+1}\frac{j}{m+1}\sum_{p=1}^{[ls]} u_{\tau_p} - n^{-1/2}\sum_{j=1}^{m+1}\sum_{p=1}^{[ls]} u_{\tau_{p-1}+j} \\ & + \frac{\Gamma}{n^{1/2}}\sum_{j=1}^{m+1}\left(\frac{j}{m+1}-1\right)\sum_{i=1}^{m+1}\sum_{p=1}^{[ls]} \triangle q_{\tau_{p-1}+i} \\ & + \frac{\Gamma}{n^{1/2}}\sum_{j=1}^{m+1}\sum_{i=j+1}^{m+1}\sum_{p=1}^{[ls]} \triangle q_{\tau_{p-1}+i} \end{aligned} \tag{25}$$

where the first two terms of (25) may be rewritten as

$$n^{-1/2}\sum_{j=1}^{m+1}\sum_{p=1}^{[ls]}u_{\tau_{p-1}}+o_p\left(1\right)=\left(m+1\right)^{-1/2}\sum_{j=1}^{m+1}U_l^{(0)}\left(s\right)+o_p\left(1\right)$$

for large l. This has a limiting distribution given by $\left(m+1\right)^{1/2}U^{(0)}\left(s\right)$ as a direct result of part [b]. The third term is $\left(m+1\right)^{-1/2}\sum_{j=1}^{m+1}U_l^{(j)}\left(s\right)$, which has a limiting distribution of $U\left(s\right)$ according to part [c]. Similarly, using part [c], the limiting distribution of the fourth term of (25) is $-(m/2)\Gamma Q\left(s\right)$. Finally, the distribution of the last term of (25) is $\Gamma\left(m+1\right)^{-1/2}\sum_{j=1}^{m+1}\sum_{i=j+1}^{m+1}Q^{(i)}\left(s\right)$ using part [a] of the lemma. □

Appendix B: Proofs of the Main Results

Proof of Theorem 4.1 Under our assumptions, all four results of Lemma A.3 hold. Moreover, since we assume that [A5] and either [NED6] or the result of Lemma A.2[e] hold, we have $\sigma_{wv}-\Sigma_{w*}C'\psi=0$ in Lemma A.3[b]. Consequently, using Lemma A.3, [A6], and the continuous mapping theorem, $n^{-1}M_n^*\to_d M^*$ as $n\to\infty$. Similarly, $n^{-2}N_n^*=\frac{1}{n^2}\sum x_t^*x_t^{*\prime}+o_p\left(1\right)\to_d N$, so that the stated result is obtained. The choice of C ensures that CNC' is invertible, even though N is not. □

Proof of Lemma 4.2 Since [NED1]-[NED2], [NED4'], and [NED5.a]-[NED5.b] are sufficient for Lemma A.2[d] and [f], we prove the lemma using the latter. Using the definition of u_t^*, we may rewrite the first summation as $\sum u_tq_t'-\sum u_t\bar{q}_n'+\sum z_t^*q_t'-\sum z_t^*\bar{q}_n'$. The first term is $O_p\left(n\right)$ under [A1]-[A3], using standard asymptotics for integrated series. The second term is also $O_p\left(n\right)$, since $n^{-1/2}\bar{q}_n\to_d\int Q$ and since our assumption that $\mathbf{E}u_t=0$ allows a central limit theorem for $n^{-1/2}\sum u_t$. The third and fourth terms are $O_p\left(n\right)$ by Lemma A.2[f] and [d], respectively. It remains to show that the second summation in the estimator is $O_p\left(n^2\right)$. This is straightforward, since $n^{-2}\sum q_tq_t'-n^{-3/2}\sum q_tn^{-3/2}\sum q_t'\to_d\int QQ'-\int Q\int Q'$ as $n\to\infty$. This is invertible since the trends (q_t) are distinct. □

Proof of Lemma 4.3 Since [NED1]-[NED6] are sufficient for Lemma A.2, we prove the lemma using the latter.

[a] The estimator $(\hat{\alpha}_{LS}-\alpha)$ may be rewritten as $(\sum w_tw_t')^{-1}\sum w_t((\psi'C-\hat{\psi}_{LS}'\hat{C})x_t+(v_t-\hat{\psi}_{LS}'\hat{C}z_t^*))'$ by substituting (1) and our definition of (v_t^*) into (7). Under our assumptions, $\sum w_tw_t'=O_p\left(n\right)$ and invertible. Under [A5] and Lemma A.2[e], we have $\sum w_t(v_t-z_t^{*\prime}\hat{C}'\hat{\psi}_{LS})=O_p(n^{1/2})$. 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 at most $O_p\left(1/n\right)$ under our assumptions. Thus, $\sum w_tx_t'(C'\psi-\hat{C}'\hat{\psi}_{LS})=O_p\left(1\right)$, and whole estimator is $O_p\left(n^{-1/2}\right)$.

[b] 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 (6). The first term is $(\hat{\Gamma}_{LS}-\Gamma)O_p(n^{1/2})$, which is $O_p(n^{-1/2})$ by Lemma 4.2. The second term is $O_p(n^{-1/2})$ using a CLT, as is the third term by Lemma A.2[d]. □

Proof of Lemma 4.4 The estimator (9) may be written as

$$\tilde{\Omega}_{b^*b^*} = (I, D)\left\{\frac{1}{n}\sum\sum\nolimits_{s=1}^{n}\zeta_t\zeta_s'\pi\left(\frac{t-s}{\ell_n}\right)\right\}(I, D)', \tag{26}$$

with $R\times R$ identity submatrix and known D. The expression inside the curly brackets is an estimator of the long-run variance of (ζ_t), which is a vector of NED sequences by Lemma A.1[c]. We need only show that this estimator is consistent for the result to hold. Note that for the $(R+r)$-dimensional random vector $\zeta_t - \mathbf{E}\left(\zeta_t|\mathcal{F}_{t-K}^{t+K}\right)$,

$$\sum\nolimits_i \mathbf{E}\left|\zeta_{it} - \mathbf{E}\left(\zeta_{it}|\mathcal{F}_{t-K}^{t+K}\right)\right|^2 \leq (R+r)\sup\nolimits_i \mathbf{E}\left|\zeta_{it} - \mathbf{E}\left(\zeta_{it}|\mathcal{F}_{t-K}^{t+K}\right)\right|^2$$

for $i = 1, \ldots, R+r$ so that

$$\left\|\zeta_t - \mathbf{E}\left(\zeta_t|\mathcal{F}_{t-K}^{t+K}\right)\right\|_2 \leq (R+r)^{1/2}\sup\nolimits_i\left\|\zeta_{it} - \mathbf{E}\left(\zeta_{it}|\mathcal{F}_{t-K}^{t+K}\right)\right\|_2,$$

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. The latter is NED with sequences $(d_t^{b,z})$ and $(\nu_K^{b,z})$ defined in terms of the respective sequences (d_t) and (ν_K) implicitly defined in Lemma A.1[c]. Specifically, let $d_t^{b,z} \equiv (R+r)^{1/2}\max_i d_{it}$ and $\nu_K^{b,z} \equiv \max_i \nu_{iK}$, and since the sequences (d_{it}) are bounded and there are a finite number of regressors in (3), $(d_t^{b,z})$ is also bounded. Along similar lines, $\sup_t \|\zeta_t\|_{2a/(a-1)} \leq (R+r)^{1/2}\sup_{i,t}\|\zeta_{it}\|_{2a/(a-1)}$, which is finite by [NED4']. Conditions [NED1], [A1], [K1] and [K2] are jointly sufficient with the above inequality for Theorem 2.1 of de Jong and Davidson (2000), so that the estimator inside the curly brackets in (26) is consistent. □

Proof of Lemma 4.5 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 messy-data noise. Consider $\hat{\Delta}_{b^*b^*}$. If we can show that $\hat{\Delta}_{b^*b^*} = \tilde{\Delta}_{b^*b^*} + o_p(1)$, 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 the difference between the two estimators of $\Delta_{b^*b^*}$ is

$$\left|\hat{\Delta}_{b^*b^*} - \tilde{\Delta}_{b^*b^*}\right| \leq \frac{1}{\sqrt{n}}\sum_{k=0}^{n}\left|\pi\left(\frac{k}{\ell_n}\right)\right|\left(\begin{array}{c}\frac{1}{\sqrt{n}}\sum_{t=k+1}^{n}\left|\hat{b}_t^*(\hat{b}_{t-k}^* - b_{t-k}^*)'\right| \\ +\frac{1}{\sqrt{n}}\sum_{t=k+1}^{n}\left|(\hat{b}_t^* - b_t^*)b_{t-k}^{*\prime}\right|\end{array}\right) \tag{27}$$

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/ℓ_n is $o\left(n^{1/2}\right)$ by [K3], the result holds if we can show that the remaining sums in the majorant of (27) are both $O_p\left(n^{1/2}\right)$.

Consider the second summation in the majorant of (27). The ij^{th} element is

$$\frac{1}{\sqrt{n}}\sum_{t=k+1}^{n}\left|e_i'\hat{b}_t^*(\hat{b}_{t-k}^* - b_{t-k}^*)'e_j\right| \leq \left(\frac{1}{n}\sum\left|e_i'\hat{b}_t^*\hat{b}_t^{*\prime}e_i\right|\right)^{1/2}$$
$$\times\left(\sum\left|e_j'(\hat{b}_t^* - b_t^*)(\hat{b}_t^* - b_t^*)'e_j\right|\right)^{1/2} \tag{28}$$

using the Cauchy-Schwarz inequality and the non-negativity of k. We may expand the first summation in the majorant of (28) (up to premultiplication by e_i' and postmultiplication by e_i) as

$$\frac{1}{n}\sum \hat{b}_t\hat{b}_t' + \frac{1}{n}\sum \hat{D}z_t^* z_t^{*\prime}\hat{D}' + \frac{1}{n}\sum \hat{b}_t z_t^{*\prime}\hat{D}' + \frac{1}{n}\sum \hat{D}z_t^*\hat{b}_t' \tag{29}$$

using (12). The stochastic boundedness of $\frac{1}{n}\sum \hat{b}_t\hat{b}_t'$ follows by the same reasoning as that employed by Park (1992, Lemma 4.3), since (w_t) and $(\triangle q_t)$ are stationary, and $(\hat{v}_t)$ and $(\hat{u}_t)$ consistently estimate stationary series (v_t) and (u_t) by way of (11) and the rates of convergence of $\hat{\psi}-\psi$, etc. already established under our assumptions. The second term of (29) is $O_p(1)$ as a direct result of Lemma A.2[b] and the consistency of the estimator $\hat{D}$. The third and fourth terms of (29) are slightly more complicated. They involve sums of both outer products of (w_t) and $(\triangle q_t)$ with (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 $\sum v_t z_t^{*\prime} + (\alpha-\hat{\alpha})'\sum w_t z_t^{*\prime} + (\psi' C - \hat{\psi}'\hat{C})\sum x_t z_t^{*\prime}$ and $\sum u_t z_t^{*\prime} + (\mu-\hat{\mu})\sum z_t^{*\prime} + (\Gamma-\hat{\Gamma})\sum q_t z_t^{*\prime}$, which are $O_p(n)$ by [A5], Lemma A.2[a], [c], and [f], Theorem 3.1, and Lemmas 3.2 and 3.3. The last two terms of (29) are therefore $O_p(1)$.

An expansion of the second summation in the majorant of (28) (up to premultiplication by e_j' and postmultiplication by e_j) yields

$$\begin{aligned}&\sum(\hat{b}_t-b_t)(\hat{b}_t-b_t) + (\hat{D}-D)\sum z_t^* z_t^{*\prime}(\hat{D}-D)' \\ &+\sum(\hat{b}_t-b_t)z_t^{*\prime}(\hat{D}-D)' + (\hat{D}-D)\sum z_t^*(\hat{b}_t-b_t)',\end{aligned} \tag{30}$$

the first term of which is $O_p(1)$ – again, as a straightforward extension of the proof in Park's proof (1992, Lemma 4.3). The summation $\sum z_t^* z_t^{*\prime}$ 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 and fourth terms of (30), the vector $(\hat{b}_t - b_t)$ contains zeros for observable series (w_t) and $(\triangle q_t)$. For the subseries $(\hat{v}_t)$ and $(\hat{u}_t)$ of estimates, we must contend with summations $(\alpha-\hat{\alpha})'\sum w_t z_t^{*\prime}(\hat{D}-D)' + (\psi' C-\hat{\psi}'\hat{C})\sum x_t z_t^{*\prime}(\hat{D}-D)$ and $(\mu-\hat{\mu})\sum z_t^{*\prime}(\hat{D}-D)' + (\Gamma-\hat{\Gamma})\sum q_t z_t^{*\prime}(\hat{D}-D)'$, which are both $O_p(1)$ under our assumptions. Again, since elements of the matrix $(\hat{D}-D)$ are either identically 0 or $O_p(1/n)$, the third term of (30) is $O_p(1)$. Consequently, the second summation in the majorant of (28) is $O_p(1)$, so that the entire majorant is $O_p(1)$.

Finally, we must show that the third summation in the majorant of (27) is stochastically bounded. This follows in exactly the same way as the second summation, except that i and j are exchanged in the majorant of (28). This completes the proof for $\hat{\Delta}_{b^*b^*}$. □

Proof of Theorem 5.1 The estimator may be written as $(\hat{\psi}_{CCR}-\psi) = (C\sum x_t^{**}x_t^{**\prime}C')^{-1}(C\sum x_t^{**}w_t'(\alpha-\hat{\alpha}) + C\sum x_t^{**}v_t^{**}) + o_p(1)$ since $\hat{C}\to_p C$. Now, since $(\alpha-\hat{\alpha}) = o_p(1)$, Lemma A.4[b] implies that $\sum x_t^{**}w_t'(\alpha-\hat{\alpha}) = o_p(n)$. The resulting distribution follows directly from Lemma A.4[a] and [c]. □

Proof of Proposition 6.1 The proof proceeds by verifying the conditions for Lemma A.2 in addition to [NED1] and [NED4'].

Verification of Lemma A.2[a] First, note that we may write

$$\frac{1}{n}\sum z_t^* = \frac{1}{m+1}\sum_{j=1}^{m+1}\left(1-\frac{j}{m+1}\right)\frac{1}{l}\sum_{p=1}^{l}u_{\tau_{p-1}} + \frac{1}{(m+1)^2}\sum_{j=1}^{m+1}j\frac{1}{l}\sum_{p=1}^{l}u_{\tau_p} \qquad (31)$$

$$- \frac{1}{m+1}\sum_{j=1}^{m+1}\frac{1}{l}\sum_{p=1}^{l}u_{\tau_{p-1}+j} - \frac{1}{m+1}\Gamma\sum_{j=1}^{m+1}\sum_{i=1}^{j}\frac{1}{l}\sum_{p=1}^{l}\triangle q_{\tau_{p-1}+i}$$

$$+ \frac{1}{(m+1)^2}\Gamma\sum_{j=1}^{m+1}j\sum_{i=1}^{m+1}\frac{1}{l}\sum_{p=1}^{l}\triangle q_{\tau_{p-1}+i}$$

since $z_t^* = 0$ for $t = \tau_p$ with $p = 1, \ldots, l$ and $n/l = m+1$. We can apply an LLN to the final summation in each of these terms. The second summations in each of the first two terms of (31) obey LLN's (with mean zero), so that both terms are simply $o_p(1)\, m/2\,(m+1) = o_p(1)$. Even though $u_{\tau_{p-1}+j}$ in the third term of (31) depends on j, an LLN similarly applies to the second summation in this term. Again, the whole term is $o_p(1)$ since $m < \infty$. The fourth and fifth terms of (31) would be trickier to deal with if m were increasing with the sample size. Since that is not the case, we may again apply an LLN to the last summations in each term to see that both terms are $O(m)\, o_p(1) = o_p(1)$.

Verification of Lemma A.2[b] We must show that the probability limit of $\frac{1}{n}\sum z_t^* z_t^{*\prime}$ is finite and independent of t. An expansion of this matrix using (19) reveals 25 terms: 5 symmetric matrices, 10 cross-products, and 10 transposes. We examine in detail only the most complicated of these 25, which is the cross product of the last two terms of (19). The transpose of this cross-product has the same asymptotics, and the remaining 23 terms may be analyzed along similar lines. Summing across t, dividing by n, and using (19) and the fact that $n/l = m+1$, this representative term is

$$\frac{1}{(m+1)^2}\sum_{j=1}^{m}j\sum_{k=1}^{j}\sum_{i=1}^{m+1}\frac{1}{l}\sum_{p=1}^{l}\triangle q_{\tau_{p-1}+k}\triangle q'_{\tau_{p-1}+i} \to_p \frac{1}{(m+1)^2}\sum_{j=1}^{m}j\sum_{k=1}^{j}\sum_{i=1}^{m+1}\Sigma_{qq}(k-i)$$

which does not depend on time (only on m) due to the stationarity of $(\triangle q_t)$ and the summation over the index j. It remains to show that this term is finite.

$$\frac{1}{(m+1)^2}\sum_{j=1}^{m}j\sum_{k=1}^{j}\sum_{i=1}^{m+1}|\Sigma_{qq}(k-i)| \leq \frac{1}{m+1}\sum_{j=1}^{m}j^2\Sigma_{qq} = \frac{1}{6}m(2m+1)\Sigma_{qq} = O\left(m^2\right)$$

which is finite by construction. The remaining 23 terms of the expansion of $\frac{1}{n}\sum z_t^* z_t^{*\prime}$ reveal similar asymptotics.

Verification of Lemma A.2[c] The proof parallels that of part [b] and is therefore omitted.

Verification of Lemma A.2[d] Set $s = l/n$ in Lemma A.6[d] to obtain the stated result.

Verification of Lemma A.2[e] The proof parallels that of part [d], due to the contemporaneous and serial uncorrelatedness of (w_t) with (u_t) and (q_t), and is therefore omitted.

Verification of Lemma A.2[f] The sample moment $\frac{1}{n}\sum q_t z_t^{*\prime}$ whose distribution we must verify may be expanded as

$$-\frac{1}{n}\sum q_t u_t' + \frac{1}{n}\sum_{j=1}^{m+1}\sum_{p=1}^{l} q_{\tau_{p-1}+j}u_{\tau_{p-1}}' - \frac{1}{n}\sum_{j=1}^{m+1}\sum_{i=1}^{m+1}\sum_{p=1}^{l}\triangle q_{\tau_{p-1}+j+i}u_{\tau_p}' + o_p(1) \tag{32}$$
$$+\frac{1}{n}\sum_{j=1}^{m+1}\left(\frac{j}{m+1}-1\right)\sum_{i=1}^{m+1}\sum_{p=1}^{l} q_{\tau_{p-1}+j}\triangle q_{\tau_{p-1}+i}'\Gamma' + \frac{1}{n}\sum_{j=1}^{m+1}\sum_{i=j+1}^{m+1}\sum_{p=1}^{l} q_{\tau_{p-1}+j}\triangle q_{\tau_{p-1}+i}'\Gamma'$$

since $\sum_{p=1}^{l} q_{\tau_{p-1}+j}u_{\tau_p}' = \sum_{p=1}^{l} q_{\tau_p+j}u_{\tau_p}' - \sum_{p=1}^{l}\sum_{i=1}^{m+1}\triangle q_{\tau_{p-1}+j+i}u_{\tau_p}'$ and $\sum_{p=1}^{l} q_{\tau_p+j}u_{\tau_p}' = \sum_{p=1}^{l} q_{\tau_{p-1}+j}u_{\tau_{p-1}}' - q_j u_0' + q_{\tau_l+j}u_{\tau_l}'$. The first term of (32) has a limiting distribution of $\int QdU(s)' + \Delta_{uq}'$ using standard asymptotic theory. The second term is only slightly more complicated. The asymptotics are similar, with the primary difference being that the series $(u_{\tau_{p-1}})$ contains $m+1$ multiples of l members, with a total of n members. Consequently, it is still on the same clock as $(q_{\tau_{p-1}+j})$. The limiting distribution of this term is $n^{-1/2}\sum_{j=1}^{m+1}\sum_{p=1}^{[ls]} u_{\tau_{p-1}}' \to_d (m+1)^{1/2}U^{(0)}(s)$ using Lemma A.6[b]. The limiting distribution of the second term of (32) is therefore $(m+1)^{1/2}\int QdU^{(0)}(s)' + \Delta_{uq}' + \sum_{i=1}^{j}\Sigma_{uq}'(-i)$ along similar lines as the first. The limiting distribution of the third term of (32) is simply

$$\frac{1}{m+1}\sum_{j=1}^{m+1}\sum_{i=1}^{m+1}\mathbf{E}\triangle q_{\tau_{p-1}+j+i}u_{\tau_p}' = \frac{1}{m+1}\sum_{j=1}^{m+1}\sum_{i=1}^{m+1}\Sigma_{uq}'(m+1-(j+i))$$

using an LLN. The summation over i in the fourth term of (32) may be expanded as

$$\frac{1}{n}\sum_{j=1}^{m+1}\left(\frac{j}{m+1}-1\right)\sum q_t\triangle q_t'\Gamma' \tag{33}$$
$$+\frac{1}{n}\sum_{j=1}^{m+1}\left(\frac{j}{m+1}-1\right)\sum_{p=1}^{l}\sum_{i=1}^{j-1}\sum_{k=i+1}^{j}\triangle q_{\tau_{p-1}+k}\triangle q_{\tau_{p-1}+i}'\Gamma'$$
$$-\frac{1}{n}\sum_{j=1}^{m+1}\left(\frac{j}{m+1}-1\right)\sum_{p=1}^{l}\sum_{i=j+1}^{m+1}\sum_{k=j+1}^{i}\triangle q_{\tau_{p-1}+k}\triangle q_{\tau_{p-1}+i}'\Gamma'$$

The limiting distribution of the first term of (33) $-(m/2)(\int QdQ(s)' + \Delta_{qq}')\Gamma'$follows from standard asymptotic theory. The limits of the third and fourth terms are clearly

$$\sum_{j=1}^{m+1}\left(\frac{j}{m+1}-1\right)\left(\sum_{i=1}^{j-1}\sum_{k=i+1}^{j}\Sigma_{qq}'(i-k) + \sum_{i=j+1}^{m+1}\sum_{k=j+1}^{i}\Sigma_{qq}'(i-k)\right)\Gamma'$$

using an LLN. The fifth term of (32) has more complicated indices. The limit may be deduced by focusing on the summation with l summands, since this is the only one with infinite summands in the limit. The difficulty lies in the fact that the sequence $(q_{\tau_{p-1}+j})$ may have up to n – not l – members. Hence $(q_{\tau_{p-1}+j})$ and $(\triangle q_{\tau_{p-1}+i})$ reside in a finer partition of the state space. In order to use standard asymptotics to obtain the limit, we may rewrite these as stochastic processes with the same time clock as the summation. Specifically, we may rewrite

$$\frac{1}{l}\sum_{p=1}^{l} q_{\tau_{p-1}+j}\triangle q'_{\tau_{p-1}+i} = \sum_{p=1}^{l}\left(\sum_{k=1}^{m+1} Q_l^{(k)}\left(\frac{p-l}{l}\right)+\sum_{k=1}^{j}\frac{\triangle q_{\tau_{p-1}+k}}{l^{1/2}}\right)\left(Q_l^{(i)}\left(\frac{p}{l}\right)-Q_l^{(i)}\left(\frac{p-l}{l}\right)\right)'.$$

Note that

$$\sum_{p=1}^{l}\sum_{k=1}^{m+1} Q_l^{(k)}\left(\frac{p-l}{l}\right)\left(Q_l^{(i)}\left(\frac{p}{l}\right)-Q_l^{(i)}\left(\frac{p-l}{l}\right)\right) \to_d (m+1)^{1/2}\int QdQ^{(i)}(s)'+\sum_{k=1}^{m+1}\Delta_{qq}^{(k-i)\prime}$$

by standard limit theory and Lemma A.6[c], where $\Delta_{qq}^{(k-i)} \equiv \sum_{r=1}^{\infty}\Sigma_{qq}\left(r\left(m+1\right)+k-i\right)$ comes from the covariance of increments of $Q_l^{(k)}$ and $Q_l^{(i)}$. Thus,

$$\frac{1}{n}\sum_{j=1}^{m+1}\sum_{i=j+1}^{m+1}\sum_{p=1}^{l} q_{\tau_{p-1}+j}\triangle q'_{\tau_{p-1}+i}\Gamma' \to_d (m+1)^{-1/2}\sum_{j=1}^{m+1}\sum_{i=j+1}^{m+1}\left(\int QdQ^{(i)}(s)'+\sum_{k=1}^{m+1}\Delta_{qq}^{(k-i)\prime}\right)\Gamma'.$$

Moreover,

$$\frac{1}{n}\sum_{j=1}^{m+1}\sum_{i=j+1}^{m+1}\sum_{p=1}^{l}\sum_{k=1}^{\bar{m}}\triangle q_{\tau_{p-1}+k}\triangle q'_{\tau_{p-1}+i}\Gamma' \to_p \frac{1}{m+1}\sum_{j=1}^{m+1}\sum_{i=j+1}^{m+1}\sum_{k=1}^{j}\Sigma_{qq}\left(k-i\right)\Gamma'$$

using an LLN, since the increments are stationary and mixing by [A1]. Collecting terms provides a distribution of $\int QdZ^*(s)'+\Delta_{q*}$ with $Z^*(s)$ defined as in Lemma A.6 and Δ_{q*} defined implicitly by the remaining (nonstochastic) limits, which are not time-dependent, even though they generally depend on the length m of each missing interval.

Verification of [**NED1**] Since (u_t) and $(\triangle q_t)$ are stationary, they have Wold representations, which we generically denote by $\sum_{l=0}^{\infty}\varphi_l\varepsilon_{t-l}+c_t$, where (ε_t) is a generic sequence of white noise, (c_t) is a generic predictable sequence, and (φ_k) is a generic sequence of absolutely summable coefficients. Similarly to Davidson (1994, Example 17.3), we have

$$\begin{aligned}\left\|u_{\tau_{p-1}+j}-\mathbf{E}\left(u_{\tau_{p-1}+j}|\mathcal{F}_{\tau_{p-1}+j-K}^{\tau_{p-1}+j+K}\right)\right\|_2 &= \left\|\sum_{k=K+1}^{\infty}\varphi_k\left(\varepsilon_{\tau_{p-1}+j-k}-\mathbf{E}\left(\varepsilon_{\tau_{p-1}+j-k}|\mathcal{F}_{\tau_{p-1}+j-K}^{\tau_{p-1}+j+K}\right)\right)\right\|_2\\ &\le \sup_{s\le\tau_{p-1}+j}\|\varepsilon_s\|_2\sum_{k=K+1}^{\infty}|\varphi_k|\end{aligned}$$

since the difference is zero for $k\le K$. Analogously, for $(u_{\tau_{p-1}})$ and (u_{τ_p}), we have

$$\begin{aligned}\left\|u_{\tau_{p-1}}-\mathbf{E}\left(u_{\tau_{p-1}}|\mathcal{F}_{\tau_{p-1}+j-K}^{\tau_{p-1}+j+K}\right)\right\|_2 &= \left\|\sum_{k=K+1-j}^{\infty}\varphi_k\left(\varepsilon_{\tau_{p-1}-k}-\mathbf{E}\left(\varepsilon_{\tau_{p-1}-k}|\mathcal{F}_{\tau_{p-1}+j-K}^{\tau_{p-1}+j+K}\right)\right)\right\|_2\\ &\le \sup_{s\le\tau_{p-1}}\|\varepsilon_s\|_2\sum_{k=K+1-(m+1)}^{\infty}|\varphi_k|\end{aligned}$$

since the difference is zero for $k \leq K - j$ and since $j \leq m + 1$, and

$$\left\| u_{\tau_p} - \mathbf{E}\left(u_{\tau_p} | \mathcal{F}_{\tau_{p-1}+j-K}^{\tau_{p-1}+j+K} \right) \right\|_2 = \left\| \sum_{k=K+1-j+(m+1)}^{\infty} \varphi_k \left(\varepsilon_{\tau_p - k} - \mathbf{E}\left(\varepsilon_{\tau_p - k} | \mathcal{F}_{\tau_{p-1}+j-K}^{\tau_{p-1}+j+K} \right) \right) \right\|_2$$
$$\leq \sup_{s \leq \tau_p} \|\varepsilon_s\|_2 \sum_{k=K+1}^{\infty} |\varphi_k|$$

since the difference is zero for $k \leq K - j + (m+1)$ and again since $j \leq m + 1$. Clearly, $\sup_{s \leq t} \|\varepsilon_s\|_2 < \infty$ for any t by the covariance stationarity of (u_t), and the summations of coefficient above are finite and go to zero as K increases. Multiplying any of these terms by $j/(m+1)$ does not fundamentally alter these results, since $j/(m+1) \leq 1$. Note that the summability of (φ_k) implies near-epoch dependence of size $-\infty$ (which is also size -1). Consequently, [NED1] is verified for the terms of (19) involving (u_t).

Similarly, since the series $(\triangle q_t)$ is stationary, we may write the norm of the difference between this term and its conditional expectation as

$$\left\| \sum_{k=(K+1)-j+i}^{\infty} \varphi_k \left(\varepsilon_{\tau_{p-1}+i-k} - \mathbf{E}\left(\varepsilon_{\tau_{p-1}+i-k} | \mathcal{F}_{\tau_{p-1}+j-K}^{\tau_{p-1}+j+K} \right) \right) \right\|_2 \leq \sup_{s \leq \tau_{p-1}+i} \|\varepsilon_s\|_2 \sum_{k=(K+1)-(m+1)}^{\infty} |\varphi_i|$$

since the difference is zero for $k \leq K - j + i$ and since $j - i \leq m + 1$. Since the terms involving $(\triangle q_t)$ are simply linear combinations of $(\triangle q_{\tau_{p-1}+i})$, and since the properties of NED processes are preserved under such transformations, these terms are also NED, which means the entire messy-data noise given by (19) is NED with the required properties.

Verification of [NED4'] The Minkowski inequality allows

$$\sup_{p \leq l, j \leq m} \left\| z^*_{\tau_{p-1}+j} \right\|_{2a/(a-1)} \leq \sup_{p \leq l, j \leq m} \left\| u_{\tau_{p-1}+j} \right\|_{2a/(a-1)} + \sup_{p \leq l} \left\| u_{\tau_{p-1}} \right\|_{2a/(a-1)}$$
$$+ \sup_{p \leq l} \left\| u_{\tau_p} \right\|_{2a/(a-1)}$$
$$+ 2\Gamma(m+1) \sup_{p \leq l, j \leq m} \left\| \triangle q_{\tau_{p-1}+j} \right\|_{2a/(a-1)}$$

since $j \leq m + 1$. The stated result immediately follows from [A1]. □

References

Andrews, D.W.K. (1991). "Heteroskedasticity and Autocorrelation Consistent Covariance Matrix Estimation," *Econometrica*, 59, 817-58.

Billingsley, P. (1968). *Convergence of Probability Measures.* New York: John Wiley.

Chang, Y., J.I. Miller, and J.Y. Park (2009). "Extracting a Common Stochastic Trend: Theory with Some Applications," *Journal of Econometrics*, forthcoming.

Chang, Y. and W. Rhee (2005). "Cost of Living Index and More Evidence on Convergence across States," mimeograph, Department of Economics, Rice University.

Chow, C.C. and A.-L. Lin (1971). "Best Linear Unbiased Interpolation, Distribution, and Extrapolation of Time Series by Related Series," *The Review of Economics and Statistics*, 53, 372-5.

Chow, C.C. and A.-L. Lin (1976). "Best Linear Unbiased Estimation of Missing Observations in an Economic Time Series," *Journal of the American Statistical Association*, 71, 719-21.

Davidson, J. (1994). *Stochastic Limit Theory: An Introduction for Econometricians*. Oxford: Oxford University Press.

Davidson, J. and R.M. de Jong (1997). "Strong Laws of Large Numbers for Dependent Heterogeneous Processes: A Synthesis of Recent and New Results," *Econometric Reviews*, 16, 251-79.

de Jong, R.M. (1997). "Central Limit Theorems for Dependent Heterogeneous Random Variables," *Econometric Theory*, 13, 353-67.

de Jong, R.M. and J. Davidson (2000). "Consistency of Kernel Estimators of Heteroscedastic and Autocorrelated Covariance Matrices," *Econometrica*, 68, 407-23.

Friedman, M. (1962). "The Interpolation of Time Series by Related Series," *Journal of the American Statistical Association*, 57, 729-57.

Ghysels, E., P. Santa-Clara, and R. Valkanov (2004). "The MIDAS Touch: Mixed Data Sampling Regression Models," CIRANO Working Paper 2004s-20.

Goetzmann, W.N., R.G. Ibbotson, and L. Peng (2001). "A New Historical Database for the NYSE 1815 to 1925: Performance and Predictability," *Journal of Financial Markets*, 4, 1-32.

Gomez, V. and A. Maravall (1994). "Estimation, Prediction, and Interpolation for Nonstationary Series with the Kalman Filter," *Journal of the American Statistical Association*, 89, 611-24.

Gonzalo, J. and C.W.J. Granger (1995). "Estimation of Common Long-Memory Components in Cointegrated Systems," *Journal of Business & Economic Statistics*, 13, 27-35.

Hansen, B.E. (1991). "GARCH(1,1) Processes Are Near Epoch Dependent," *Economics Letters*, 36, 181-6.

Hansen, B.E. (1992). "Consistent Covariance Matrix Estimation for Dependent Heterogeneous Processes," *Econometrica*, 60, 967-72.

Harvey, A.C., S.J. Koopmans, and J. Penzer (1998). "Messy Time Series: A Unified Approach," *Advances in Econometrics*, 13, 103-43.

Harvey, A.C. and R.G. Pierse (1984). "Estimating Missing Observations in Economic Time Series," *Journal of the American Statistical Association*, 79, 125-31.

Ibragimov, I.A. (1962). "Some Limit Theorems for Stationary Processes," *Theory of Probability and Its Applications*, 7, 349-82.

Kohn, R. and C.F. Ansley (1986). "Estimation, Prediction, and Interpolation for ARIMA Models with Missing Data, *Journal of the American Statistical Association*, 81, 751-61.

Kurozumia, E., and K. Hayakawa (2009). "Asymptotic Properties of the Efficient Estimators for Cointegrating Regression Models with Serially Dependent Errors," *Journal of Econometrics*, 149, 118-35.

McLeish, D.A. (1975). "A Maximal Inequality and Dependent Strong Laws," *Annals of Probability*, 3, 329-39.

Mariano, R.S. and Y. Murasawa (2003). "A New Coincident Index of Business Cycles Based on Monthly and Quarterly Series," *Journal of Applied Econometrics*, 18, 427-43.

Müller, U.K., and M.W. Watson (2008a) "Testing Model of Low-Frequency Variability," *Econometrica*, 76, 979-1016.

Müller, U.K., and M.W. Watson (2008b) "Low-Frequency Robust Cointegration Testing," mimeograph, Department of Economics, Princeton University.

Newey, W.K. and K.D. West (1987). "A Simple, Positive Semi-Definite, Heteroskedasticity and Autocorrelation Consistent Covariance Matrix," *Econometrica*, 55, 703-08.

Park, J.Y. (1992). "Canonical Cointegrating Regressions," *Econometrica*, 60, 119-43.

Phillips, P.C.B. (1995). "Fully Modified Least Squares and Vector Autoregression," *Econometrica*, 63, 1023-78.

Phillips, P.C.B. and B.E. Hansen (1990). "Statistical Inference in Instrumental Variables Regression with I(1) Processes," *The Review of Economic Studies*, 57, 99-125.

Phillips, P.C.B. and S. Ouliaris (1990). "Asymptotic Properties of Residual-Based Tests for Cointegration," *Econometrica*, 58, 165-93.

Seong, B., S.K. Ahn, and P.A. Zadorozny (2007). "Cointegration Analysis with Mixed-Frequency Data," CESifo Working Paper No. 1939.

Wright, J.H. (2000). "Alternative Variance Ratio Tests Using Ranks and Signs," *Journal of Business and Economic Statistics*, 18, 1-9.