

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

J. Isaac Miller¹

*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 possibly nonstationary noise. We motivate the notion of messy data with a nontechnical example using linear interpolation. Even with such a straightforward DGP, we show that the resulting noise is mildly nonstationary. We adopt a unified theoretical approach to avoid strict distributional assumptions and to allow for such nonstationarity. Least squares estimation of the cointegrating vector is consistent under general conditions, even though the estimator is neither asymptotically normal nor unbiased. In order to allow valid statistical inference, we construct a canonical cointegrating regression (CCR) using standard consistent nonparametric variance estimators, and we show that least squares estimation of the CCR provides consistent and asymptotically normal estimation – even with nonstationary disturbances. We briefly examine large- and small-sample properties of the estimator when linear interpolation is the specific driver behind the messiness.

First draft: January 18, 2007

This draft: January 26, 2009

JEL Classification: C13, C14, C32

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

¹This draft prepared for presentation at Exeter, Leeds, Nottingham, and York. 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, and Rice – in particular, Bill Brown, Robert de Jong, Mahmoud El-Gamal, Dave Mandy, Joon Park, Peter Phillips, and Robin Sickles – for helpful comments and suggestions. Thanks also to Wensheng Kang for research assistance.

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

1. Introduction

Data irregularities have plagued researchers across many fields of inquiry as long as data have been collected. Problems such as missingness, temporal aggregation, mixed data frequencies, and measurement error generate “messy” data which have been ubiquitous nuisances at best. More seriously, such messiness frequently creates bias, inconsistency, and non-normality. In a time series setting, messiness may even create nonstationarity. Statistical inference may be invalidated. In this paper, we rigorously show that standard econometric techniques for cointegrating regressions may allow valid inference, even when messy data causing such problems are unavoidable.

The statistics and econometrics literatures (not to mention those of other disciplines) abound in important work on handling messy data. Handbook chapters, such as that composed by Griliches (1986), loose compilations of papers, such as a volume of *Advances in Econometrics* (1998), and many individual papers are dedicated to messy data. Specific types of messiness generally engender specific techniques. The “classical” case of missing (independent) data, for example, is surveyed extensively by Little and Rubin (2002) and Shafer (1997). Imputation techniques such as the EM algorithm and multiple imputation, which necessarily create imputation error, may be appropriate in such circumstances. Analogously, surveys such as that by Fuller (1987) cover the “classical” case of mismeasured survey data. In this case, data need not be imputed, so such analyses typically start with assumptions about the cause of the measurement error, and focus on mitigating the deleterious effects. More recent analyses of survey data, such as Bound et al. (2001) and Hu and Schennach (2007) examine classes of “non-classical” measurement error, which may take nonlinear forms.

Some of the messy data problems encountered in multivariate time series and panel data analyses occur for the same reasons as in cross-sections. Problems such as nonresponse and erratically collected historical data may cause missingness in much the same way. Temporal aggregation and mixed-frequency data are also problematic in these cases. Data from different sources may have been collected at different but regular intervals, or they may be temporally aggregated to different frequencies. Approaches such as Friedman (1962) and Chow and Lin (1971, 1976) use “related” covariates in order to impute a desired high-frequency series from an observed low-frequency series. Recent approaches under the MIDAS moniker (Ghysels et al., 2004) employ spectral methods to estimate mixed frequency models with stationary series. Other approaches to this problem include state space models estimated with the Kalman filter, which may be extended to nonstationary and cointegrated models. 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 time series and model parameters. The filter uses the underlying serial dependence of the data in order to estimate conditional expectations of the missing observations. Mariano and Murasawa (2003) and Seong et al. (2007), have used the Kalman filter to estimate cointegrated models with missing data. Rigorous theory justifying the asymptotic validity of inferences using the Kalman filter with cointegrated data were recently provided by Chang et al. (2006).

Within the framework of cointegrating regressions, we adopt a unified approach to messy

data. This is not the first unified approach. Harvey et al. (1998) proposed using the Kalman filter as a general solution to messy time series data. Like Harvey et al. (1998), we do not propose a fundamentally new technique. Rather, we claim that existing semi-parametric techniques based on least squares, such as the instrumental variables approach of fully-modified least squares (FM-OLS) proposed by Phillips and Hansen (1990) and the construction of a canonical cointegrating regression (CCR) studied by Park (1992), may be used with messy data. We rigorously prove that the latter technique provides consistent and asymptotically normal (CAN) estimates of the long-run relationships between the cointegrated series, under general conditions on the error resulting from messy regressors.

Our estimation procedure compares favorably to likelihood approaches, such as the Kalman filter. Some advantages are simplicity and convenience, no numerical optimization or strict distributional assumptions, and the possibility of nonstationary error sequences. In contrast, the Kalman filter explicitly relies on iid normal errors. The Kalman filter may be modified to allow stationary autoregressive error sequences of known order, but this is still restrictive. Furthermore, the Kalman filter's reliance on numerical optimization frequently makes estimation time-consuming and sometimes intractable.

In the following section, we showcase an example of messy data created by linearly interpolating from a low-frequency series in a mixed-frequency setting. We use this non-technical example to motivate the general econometric concerns of nonstationary error arising from messy data. In Section 3, we describe our econometric model in detail. We outline general sufficient conditions for the noise resulting from messy data. These conditions should be verified in applications. In Sections 4 and 5, we present the main theoretical results of the paper. Specifically, we focus on consistent parameter estimation in Section 4, and we show that CAN estimation of the cointegrating vector is achieved by running least squares on an appropriately constructed CCR in Section 5. In Section 6, we revisit the example from Section 2. We show that our sufficient conditions are satisfied for linearly interpolated mixed-frequency data. Simulations provide insight into the trade-off in small samples between bias and efficiency when interpolation is weighed against data omission. The final section concludes by summarizing the general steps involved in estimation. Two technical appendices contain ancillary lemmas and their proofs, along with proofs of the main results.

Unless otherwise noted, summations are indexed by $t = 1, \dots, n$ 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 motivate messy data, consider a mixed-frequency example in which (x_t) and (y_t) are two time series of interest. The fundamental problem is that although we, as econometricians, observe (y_t) at each $t = 1, \dots, n$, (x_t) may only be observed at a subset of this index set. For concreteness, let the observed subseries of (x_t) be denoted by (x_{τ_p}) with $p = 1, \dots, l$, and each unobserved subseries be denoted by (x_{τ_p+j}) with $j = 1, \dots, m$. There are thus l unobserved subseries of length m . Prime examples from macroeconomics or finance include applications which require monthly measures of series that are only available

quarterly, although in this simple example we abstract from issues arising from aggregation of flow variables that may arise in such settings. If we have, say, 20 years of a quarterly GDP deflator and 20 years of a monthly CPI, then $n = 240$, $l = 80$, and $m = 2$. In other words, there are 80 observed values of the deflator but two unobserved values for every one observed value, so that two-thirds of *monthly* deflator data are not observed.

Econometricians face essentially two choices when handling such series. We may either throw out the subseries (y_{τ_p+j}) corresponding to unobserved (x_{τ_p+j}), or we may try to use the dependence of (x_t) across time and perhaps on an observed covariate (q_t) in order to impute, interpolate, or otherwise fill in unobserved values. Filling in unobserved values creates noise, which is typically called interpolation or imputation error. For the sake of generality, we refer to such noise as *messy-data noise*, of which these are special cases. Throwing out observations obviously creates inefficiency, but adding messy-data noise may create both inefficiency and bias. There is a trade-off between bias and variance, and adding small amounts of bias and inefficiency through such noise may be preferable to the larger inefficiency resulting from data omission.

In many cases, a given time series may have been imputed already by the data collector, in which case econometricians have no choice. The series is already messy. In such cases, messiness may take the same structural form, but is manifested as non-classical measurement error, rather than interpolation error. Both types of messiness are addressed by our general theory following Section 2.

2.1 The Lerp Dilemma: Nonstationarity

Linear interpolation is one possible solution to the problem posed by the mixed-frequency example. It is the most straightforward to understand and analyze. The genesis of linear interpolation can be traced back to Babylonian and Greek astronomers. Obviously, there have been many refinements, which may work better or worse in different situations. A recent chronology of interpolation (Meijering, 2002) contains 358 (mostly modern) citations. Linear interpolation, or *lerp*, as it has been affectionately dubbed, pervades many quantitative literatures as an uncomplicated alternative to data omission.

In general, we use (x_t^*) to denote a messy series. In this example, the messy series created by estimating the missing subseries using lerp is ($x_{\tau_{p-1}+j}^*$), where

$$x_{\tau_{p-1}+j}^* = \left(1 - \frac{j}{m+1}\right) x_{\tau_{p-1}} + \frac{j}{m+1} x_{\tau_p}, \quad (1)$$

for $j = 1, \dots, m+1$. We let the index j increase up to $m+1$ to cover the whole interval, with the convention that $\tau_p + (m+1) = \tau_{p+1}$. Under this indexing convention, we have $\tau_l = l(m+1) = n$ and $\tau_0 = 0$, and for concreteness we assume that x_0 is observed.

In general, we define the difference between the observed messy series and the unobserved “clean” series as

$$z_t^* \equiv x_t^* - x_t, \quad (2)$$

which we call *messy-data noise*. In this example, the messy-data noise is the difference between the partially unobserved series (x_t) and a hybrid observed and interpolated series

(x_t^*) . Using (1) and (2), we may write

$$z_t^* = \frac{j}{m+1} \sum_{i=1}^{m+1} \Delta x_{\tau_{p-1}+i} - \sum_{i=1}^j \Delta x_{\tau_{p-1}+i} \quad (3)$$

for $t = \tau_{p-1} + j$, $p = 1, \dots, l$, and $j = 1, \dots, m+1$. Note that, by construction, there is no noise when $j = m+1$ (or, alternatively, when $j = 0$) – i.e., when (x_t) is observed.

Beneath its seemingly simplistic veneer, lerp poses a fundamental problem: nonstationary messy-data noise. This noise explicitly depends on the relative time index j within each interval of missing data. Even under the most optimistic assumptions that (Δx_t) is white noise or $I(0)$, (z_t^*) is clearly nonstationary.

2.2 Why Lerp Might Work for Integrated (x_t)

If (x_t) is $I(1)$, adding an $I(0)$ noise term does not affect the integratedness of (x_t) . However, adding a nonstationary messy-data noise term may fundamentally change some of the properties of the series. In particular, (x_t^*) is *not* $I(1)$ and standard results for integrated time series are off limits without further justification. Impetus for optimism lies in the hope that the messy-data noise is asymptotically dominated by (x_t) , as it is in (3). Exploiting this dominance may allow many of the asymptotic results for integrated time series. With the possibility of a favorable bias/variance trade-off in small samples, these asymptotics may speak well for the durability of ancient astronomers' innovations.

2.3 Making the Most of Messiness

In order to refocus from stationarity (or lack thereof) to asymptotic orders of (x_t) and (z_t^*) , we make extensive use of the concept of near-epoch dependence, a concept that has been defined and refined by *inter alia* Ibragimov (1962), Billingsley (1968), and McLeish (1975).

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

$$\left\| z_t - \mathbf{E}(z_t | \mathcal{F}_{t-K}^{t+K}) \right\|_p \leq d_t \nu_K$$

where $\nu_K \rightarrow 0$ as $K \rightarrow \infty$ such that $\nu_K = O(K^{-\lambda-\varepsilon})$ for $\varepsilon > 0$, (d_t) is a sequence of positive constants, and \mathcal{F}_{t-K}^{t+K} is the σ -field generated by $\eta_{t-K}, \dots, \eta_{t+K}$.

We primarily deal with L_2 -NED sequences in this paper, which are simply described as *near-epoch dependent* (NED). The intuition underlying near-epoch dependence is that the random variable z_t is more dependent on proximate than on distant elements of the sequence (η_t) . 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. Nonstationary dependence and heterogeneity are allowed, as long as the nonstationarity of (z_t^*) does not dominate the nonstationarity of (x_t) .

Before proceeding to more explicit assumptions about the messy-data noise, we discuss the basic assumptions that comprise the foundation of our general regression model.

3. Cointegrating Regression With Messy Regressors

Consider a cointegrating regression given by

$$y_t = \alpha' w_t + \beta' x_t + v_t, \quad (4)$$

where (x_t) is an r -dimensional I(1) series, (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 α and β are conformable vectors of unknown parameters. The regressors may be cointegrated, as long as β is not a cointegrating vector of (x_t) . Under these assumptions, (y_t) is I(1) and cointegrated with (x_t) with cointegrating vector $(1, \beta)'$.

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

$$x_t = \mu + \Gamma q_t + u_t, \quad (5)$$

where (q_t) is a g -dimensional I(1) series of linearly independent stochastic trends common to (x_t) with $1 \leq g \leq r$, (u_t) is an r -dimensional I(0) 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 dimensions, $r - g$ cointegrating vectors, and g common stochastic trends. Moreover, $(y_t, x_t)'$ has $r + 1$ dimensions, $r - g + 1$ cointegrating vectors, and still g common stochastic trends.

More complicated deterministic trends in (4) and (5) may be of interest. However, including these may fundamentally alter the asymptotic results and even necessitate changing the estimation technique, if the deterministic trends dominate the stochastic trends. We focus on asymptotically 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

$$b_t \equiv (v_t, w_t', u_t', \Delta q_t')'$$

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 $4a/(a - 1)$.

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

[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 considered to be 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 $I(0)$ 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 . We assume an invariance principle (IP) for $B_n(s)$, so that

$$[\mathbf{A3}] \quad B_n(s) \rightarrow_d B(s) \equiv (V(s), W(s)', U(s)', Q(s)')',$$

a vector Brownian motion with finite long-run variance Ω_{bb} , such that

$$[\mathbf{A4}] \quad \Omega_{bb} > 0.$$

The latter assumption precludes degeneracy in the variance matrix of the stationary components of the model. 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

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

which allows \sqrt{n} -consistent estimation of α . Assumptions [A1]-[A5] are standard for this type of model.

3.1 Messy Regressors

The heart of this analysis is messy regressors.² As in the example from the previous section, we define the messy-data noise by (2), where (x_t) contains r regressors observed with noise as (x_t^*) . Using (2), feasible analogs of (4) and (5) are thus

$$y_t = \alpha' w_t + \beta' x_t^* + v_t^* \tag{6}$$

and

$$x_t^* = \mu + \Gamma q_t + u_t^*, \tag{7}$$

respectively, where $u_t^* \equiv u_t + z_t^*$ and $v_t^* \equiv v_t - \beta' z_t^*$. Note that the feasible regressors (x_t^*) in (6) are explicitly correlated with the new error term (v_t^*) . Herein lies the potential for asymptotically biased estimation, if not properly addressed.

If the messy-data noise can simply be assumed to be covariance stationary, then the results of Park (1992) or Phillips and Hansen (1990) follow directly. Variance estimation of the stationary series (u_t^*) and (v_t^*) would replace variance estimation of the stationary

²We do not explicitly consider a messy regressand, but this is a straightforward extension. Our subsequent results hold as long as a messy regressand's noise is additive and satisfies the same properties as the regressors' noise.

series (u_t) and (v_t) . However, as the simple example of linear interpolation illustrates, this assumption is unrealistic.

Instead, we set general limitations on the divergence of moments involving (z_t^*) . In particular, we require limit theory for the first two sample moments of (z_t^*) and (b_t) . Letting (z_{it}^*) , with $i = 1, \dots, r$, 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

$$\Sigma_{**}^t \equiv \mathbf{E}z_t^* z_t^{*'} \quad \text{and} \quad \Sigma_{*b}^t \equiv \mathbf{E}z_t^* b_t'$$

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 the types of laws of large numbers (LLN's) and central limit theorems (CLT's) for L_p -NED sequences analyzed by Davidson (1994), Davidson and de Jong (1997), and de Jong and Davidson (1997), among others.

We also require asymptotic results for sample moments involving integrated (q_t) . Without messy data, such asymptotics are straightforward from the IP and other limiting distributions implied by [A1]-[A5]. We define an additional stochastic process

$$Z_n^*(s) \equiv \frac{1}{\sqrt{n}} \sum_{t=1}^{\lfloor ns \rfloor} z_t^* \tag{8}$$

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

[NED5.a] $\mathbf{E}Z_n^* Z_n^{*'}(s) = \Omega_{**}(s)$ 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, $(Z_n^{*'}(s), Q_n'(s))' \rightarrow_d (Z^{*'}(s), Q'(s))'$, where $Z^*(s)$ is a vector Brownian motion with variance $\Omega_{**}(s)$, which plays a key role in our subsequent results.

Finally, we assume that

[NED6.a] $\mathbf{E}z_t^*w'_{t-k} = 0$ for all t, k ,

and that for the stochastic process (with an abuse of notation)

$$Z^*W_{n,j}(s) \equiv \frac{1}{\sqrt{n}} \sum_{t=1}^{[ns]} z_t^*w_{jt},$$

for $j = 1, \dots, p$, the following holds:

[NED6.b] $\mathbf{E}Z^*W_{n,j}(Z^*W_{n,j})'(s) = \Omega_{*w_j*w_j}$ with $\Omega_{*w_j*w_j} < \infty$.

The prohibition of contemporaneous and serial correlation between the stationary regressors and the messy-data noise is not strictly required to obtain consistent estimates of β . It is, however, 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 of this paper. 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 theoretical result.

3.2 Singularity and Unobservable Trends

Setting aside messy data, this model differs from the model assumed by the standard CCR and FM-OLS approaches, in that we allow for cointegrated regressors.³ Such singularity is synonymous with (asymptotic) collinearity. 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 . Asymptotically negligible terms may allow invertibility in finite samples, but these terms create bias and inconsistency. 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) and estimate a $g \times 1$ vector ψ in place of β . We may then make inferences about β' using $\psi'C$.

How might we choose C ? If the regressors are *not* cointegrated, then $g = r$ and we may choose $C = I$. Otherwise, a natural choice for C would be a matrix such that $CT = I$,

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

so that ψ may be meaningfully interpreted as the vector of marginal effects of the trends themselves. Since $g \leq r$, a logical choice for C is

$$C = (\Gamma'\Gamma)^{-1}\Gamma', \quad (9)$$

which is simply the Moore-Penrose generalized inverse of Γ . 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 Γ does not cointegrate (q_t).

Choosing C in this way necessitates observing (q_t) and estimating Γ . Indeed, our theoretical results implicitly assume that the trends are observable and rely on increments of (q_t) rather than (x_t). In practice, stochastic trends are rarely observable. However, estimation of our model does not necessarily require observing the trends directly. If C is chosen arbitrarily, then we only need to observe the (messy) regressors. Even if C is chosen as in (9), it may be possible to estimate Γ using series cointegrated with the trends as proxies. For example, in the panel of metropolitan 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).

If C is defined in terms of Γ , then it must be estimated. We assume that such an estimator \hat{C} satisfies

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

If \hat{C} is chosen to be $(\hat{\Gamma}'\hat{\Gamma})^{-1}\hat{\Gamma}'$, then as long as $(\hat{\Gamma} - \Gamma) = o_p(n^{-1})$ – which we show below for the least squares estimator of Γ when (q_t) or cointegrated proxies are observed – [A6] follows as a result and is a superfluous assumption. Obviously, [A6] is also superfluous if C is chosen so that it does not have to be estimated.

4. Consistent Estimation

We turn our attention to consistent estimation of the parameters described by the feasible system given by (6) and (7), with the substitution of $\psi'\hat{C}$ for β' .

4.1 Consistent Estimation of ψ

Least squares estimation of (6) provides a consistent (but neither asymptotically normal nor unbiased) estimator of ψ . We may write the least squares estimator as

$$(\hat{\psi}_{LS} - \psi) = (\hat{C}N_n^*\hat{C}')^{-1}\hat{C}M_n^*$$

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

$$N_n^* \equiv \sum x_t^* x_t^{*'} - \sum x_t^* w_t' \left(\sum w_t w_t' \right)^{-1} \sum w_t x_t^{*'},$$

using a partitioned regression.

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

$$N(s) \equiv \Gamma \int Q Q' \Gamma'$$

and

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

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

$$n(\hat{\psi}_{LS} - \psi) \rightarrow_d (CN(s)C')^{-1}CM^*(s)$$

as $n \rightarrow \infty$.

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.⁴ Also, note that naively choosing \hat{C} to be a unit row vector – as long as the unit vector does not cointegrate of (x_t) – yields a consistent estimator of ψ for that particular linear combination.

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 . Moreover, if C is defined in terms of Γ , a consistent estimator of Γ is certainly required. In fact, the least squares estimator,

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

where $\bar{q}_n \equiv n^{-1} \sum q_t$ is the sample mean of (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 $n \rightarrow \infty$.

⁴As may be seen in the proof of Lemma A.3, relaxing [A5] and [NED6] leads to a different distribution, but the estimator would still be consistent.

In addition to Γ , we also need consistent estimators of α and μ in order to estimate covariances consistently. 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

$$\hat{\alpha}_{LS} = \left(\sum w_t w_t' \right)^{-1} \sum w_t (y_t - \hat{\psi}_{LS}' \hat{C} x_t^*) \quad (10)$$

for α and

$$\hat{\mu}_{LS} = \frac{1}{n} \sum (x_t^* - \hat{\Gamma}_{LS} q_t) \quad (11)$$

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

$$\text{[a] } (\hat{\alpha}_{LS} - \alpha) = O_p(n^{-1/2}), \text{ and}$$

$$\text{[b] } (\hat{\mu}_{LS} - \mu) = O_p(n^{-1/2})$$

as $n \rightarrow \infty$.

Theorem 4.1 and Lemma 4.3[a] jointly establish consistency with appropriate rates of convergence for least squares estimation of (6), while Lemma 4.2 and Lemma 4.3[b] accomplish the same for (7).⁵ 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.

4.3 Consistent Covariance Estimation

Appropriate long-run variances and covariances must be estimated consistently. To this end, let

$$b_t^* \equiv b_t + D z_t^*$$

where D is an $R \times r$ matrix defined by

$$D \equiv (-C' \psi, 0, I, 0)',$$

where 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) and $(D z_t^*)$ are observable. In this case, the long-run variance estimator is

$$\tilde{\Omega}_{b^* b^*} = \frac{1}{n} \sum \sum_{s=1}^n b_t^* b_s^{*'} \pi((t-s)/\ell_n) \quad (12)$$

⁵In practice, a problem may arise for some imputation techniques. If data are missing from (x_t) , we need to impute (or omit) data before running least squares, but we may need to 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 . If the resulting imputation error satisfies our conditions above, then Γ may then be re-estimated consistently.

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), and Hansen (1992), *inter alia*, have addressed this problem under stationarity and/or mixing assumptions.

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

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

where $\varpi(\xi)$ is the Fourier transform of $\pi(z)$. For covariance estimation using (12) and throughout the paper, we assume that the kernel function and lag truncation parameter satisfy

$$[\mathbf{K1}] \lim_{n \rightarrow \infty} (1/\ell_n + \ell_n/n) = 0,$$

$$[\mathbf{K2}] \pi \in \mathcal{K}, \text{ and}$$

$$[\mathbf{K3}] n^{-1/2} \sum_{k=0}^n \pi(k/\ell_n) = o(1).$$

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

$$[\text{a}] \tilde{\Sigma}_{b^*b^*} \rightarrow_p \Sigma_{b^*b^*},$$

$$[\text{b}] \tilde{\Delta}_{b^*b^*} \rightarrow_p \Delta_{b^*b^*}, \text{ and}$$

$$[\text{c}] \tilde{\Omega}_{b^*b^*} \rightarrow_p \Omega_{b^*b^*}$$

as $n \rightarrow \infty$, 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, only (w_t) and (Δq_t) are observed, 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

$$\hat{v}_t^* = y_t - \hat{\alpha}' w_t - \hat{\psi}' \hat{C} x_t^* \quad \text{and} \quad \hat{u}_t^* = x_t^* - \hat{\mu} - \hat{\Gamma} q_t, \quad (13)$$

or simply by

$$\hat{v}_t^* = \hat{v}_t - \hat{\psi}' \hat{C} z_t^* \quad \text{and} \quad \hat{u}_t^* = \hat{u}_t + z_t^*,$$

where

$$\hat{v}_t \equiv v_t + (\alpha - \hat{\alpha})' w_t + (\psi' C - \hat{\psi}' \hat{C}) x_t \quad \text{and} \quad \hat{u}_t \equiv u_t + (\mu - \hat{\mu}) + (\Gamma - \hat{\Gamma}) q_t. \quad (14)$$

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

$$\hat{b}_t^* \equiv \hat{b}_t + \hat{D} z_t^* \quad (15)$$

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

$$\hat{\Omega}_{b^*b^*} \equiv \frac{1}{n} \sum \sum_{s=1}^n \hat{b}_t^* \hat{b}_s^{*'} \pi((t-s)/\ell_n) \quad (16)$$

may replace the infeasible estimator in (12). 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

$$\hat{\Delta}_{b^*b^*} \equiv \frac{1}{n} \sum_{k=0}^n \pi(k/\ell_n) \sum_{t=k+1}^n \hat{b}_t^* \hat{b}_{t-k}^{*'} \quad \text{and} \quad \hat{\Sigma}_{b^*b^*} \equiv \frac{1}{n} \sum \hat{b}_t^* \hat{b}_t^{*'} \quad (17)$$

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

$$[\text{a}] \quad \hat{\Sigma}_{b^*b^*} \rightarrow_p \Sigma_{b^*b^*},$$

$$[\text{b}] \quad \hat{\Delta}_{b^*b^*} \rightarrow_p \Delta_{b^*b^*}, \text{ and}$$

$$[\text{c}] \quad \hat{\Omega}_{b^*b^*} \rightarrow_p \Omega_{b^*b^*}$$

as $n \rightarrow \infty$.

The natural feasible estimators described by (16) and (17) are consequently consistent.

5. CAN Estimation of β Using a CCR

With consistent estimators of all of the requisite parameters in hand, we now turn to asymptotically normal estimation of the parameter vector of interest. For a prototypical cointegrated system, the most common choices are CCR, FM-OLS, or maximum likelihood. The first two are semiparametric. Being more robust to dependence, heterogeneity, and non-normality, which may be characteristic of messy-data noise, they are naturally more attractive in this setting. It is well-known that both CCR and FM-OLS estimation of the prototypical model considered by Phillips and Hansen (1990) and Park (1992) provide CAN estimators under standard assumptions. The theoretical obstacle imposed by messy regressors is not trivial, since these approaches were designed to accommodate stationary disturbances.

Nevertheless, a properly constructed CCR may achieve CAN estimation with messy regressors – even with 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 a feasible estimator is illuminated.

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 (Δq_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 (Δq_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

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

so that we estimate

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

in place of (4), where $v_t^{**} \equiv b_t' \kappa$.

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

$$y_t^{**} - \hat{\alpha}' w_t = \psi' \hat{C} x_t^{**} + v_t^{**} \tag{18}$$

to estimate ψ . It is not difficult to show that when (x_t) is not messy and the true values of all nuisance parameters are known, running least squares on (18) provides a CAN estimator 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

$$x_t^{**} \equiv x_t^* - (\hat{\Gamma} \hat{\Delta}'_{bq} + \hat{\Sigma}_{ub}) \hat{\Sigma}_{bb}^{-1} \hat{b}_t^* \quad (19)$$

and

$$y_t^{**} \equiv y_t - \hat{\psi}' \hat{C} (\hat{\Gamma} \hat{\Delta}'_{bq} + \hat{\Sigma}_{ub}) \hat{\Sigma}_{bb}^{-1} \hat{b}_t^* - \hat{\omega}_{vq} \hat{\Omega}_{qq}^{-1} \Delta q_t, \quad (20)$$

where all variances and covariances are estimated simultaneously using a single nonparametric procedure with kernel and lag truncation parameter discussed above. The error term in (18) is now

$$v_t^{**} = b_t^{*'} \hat{\kappa}, \quad (21)$$

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 the least squares estimator of this feasible CCR.

Define the 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 $Q(s)$. This Brownian motion has a long-run variance of

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

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

$$n(\hat{\psi}_{CCR} - \psi) \rightarrow_d (CN(s)C')^{-1}C\Gamma \int QdV_{\perp Q}^*(s)$$

as $n \rightarrow \infty$.

Since

$$\text{var} \left(C\Gamma \int QdV_{\perp Q}^*(s) \right) = \text{lrvar}(V_{\perp Q}^*(s))CN(s)C',$$

the asymptotic distribution may be rewritten as

$$n(\hat{\psi}_{CCR} - \psi) \rightarrow_d \sqrt{\text{lrvar}(V_{\perp Q}^*(s))} (CN(s)C')^{-1/2} \mathbf{N}(0, I_g),$$

where $(CN(s)C')^{-1/2}$ is the Cholesky decomposition of the inverse of $CN(s)C'$. 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 for $\hat{\beta}_{CCR} = \hat{C}'\hat{\psi}_{CCR}$, which may be written as

$$n(\hat{\beta}_{CCR} - \beta) \rightarrow_d \sqrt{\text{lrvar}(V_{\perp Q}^*(s))} (C'(CN(s)C')^{-1}C)^{1/2} \mathbf{N}(0, I_r),$$

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

$$\sqrt{\text{lrvar}(V_{\perp Q}^*(s))} \left(C' \left(\int QQ' \right)^{-1} C \right)^{1/2} \mathbf{N}(0, I_r),$$

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

Using the feasible system given by (6) and (7), we may write the univariate messy-data noise from linear interpolation in (3) as

$$\begin{aligned} z_t^* &= (u_{\tau_{p-1}} - u_{\tau_{p-1}+j}) + \frac{j}{m+1}(u_{\tau_p} - u_{\tau_{p-1}}) \\ &\quad - \Gamma \left(\sum_{i=1}^j \Delta q_{\tau_{p-1}+i} - \frac{j}{m+1} \sum_{i=1}^{m+1} \Delta q_{\tau_{p-1}+i} \right) \end{aligned} \quad (22)$$

We briefly present some large- and small-sample results for this specific but common form of messy-data noise.

6.1 Large Sample Results

The large sample results of Section 4 and 5 hold if either all NED assumptions of Section 3.1 hold or the results of Lemma A.2, [NED1], and [NED4'] hold. The link between these general asymptotic results and the specific practice of linear interpolating mixed-frequency data is solidified in the proof of the following proposition.

Proposition 6.1 Let [A1]-[A4] hold. Define (z_t^*) by (22) 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 some existing econometric techniques – CCR, in particular – when interpolated data are observed in place of integrated data.

6.2 Small Sample Results

In order to evaluate the small-sample performance of the CCR estimator under alternative assumptions, we simulate 16 models 10,000 times each, differentiated by the number of regressors $r = 1, \dots, 16$. In each model, the first regressor is observed every third observation, but the remaining regressors and regressand are observed every observation. This setup is consistent with a macro model in which only one regressor – say, the GDP deflator – is observed at a lower frequency. We compare linearly interpolation of the unobserved data in this regressor with omission of the corresponding data in the remaining regressors and regressand, which do not have unobserved data. As more and more regressors are included, omission – although unbiased – creates more and more inefficiency.

For simplicity, we generate regressors with serially dependent but contemporaneously independent increments. Specifically, $(v_t, \Delta x_t')$ is generated as a simple VAR(1) with Σ an $(r + 1) \times (r + 1)$ diagonal matrix with $1/2$ along the diagonals. We set β to be an $r \times 1$ vector of ones. Each model has a sample size of 240 representing, say, 20 years of monthly data. The first regressor has only $l = 80$ and $m = 2$, representing 80 quarters. Variances and covariances are estimated using a Bartlett kernel.

We use root mean-squared error (RMSE) (averaged across the simulations) in order to evaluate the in-sample fit of the estimated model. The figure illustrates average RMSE when the first regressor has 160 lerp observations and when all regressors and regressand have 160 omitted observations, *relative* to average RMSE with no mixed-frequency problem. The relative loss of fit using lerp is much slower than that using omission. Additionally, we use root mean squared one-step-ahead forecast error (RMSFE) across simulations to examine out-of-sample fit, again *relative* to RMSFE with no mixed-frequency problem. The out-of-sample fit using lerp and omission are roughly comparable.

We may conclude from these simulations that when the number of regressors observed at a higher frequency is large relative to those observed at a lower frequency (and the integrated regressors are cointegrated with the regressand), lerp may be superior to omission in small samples.

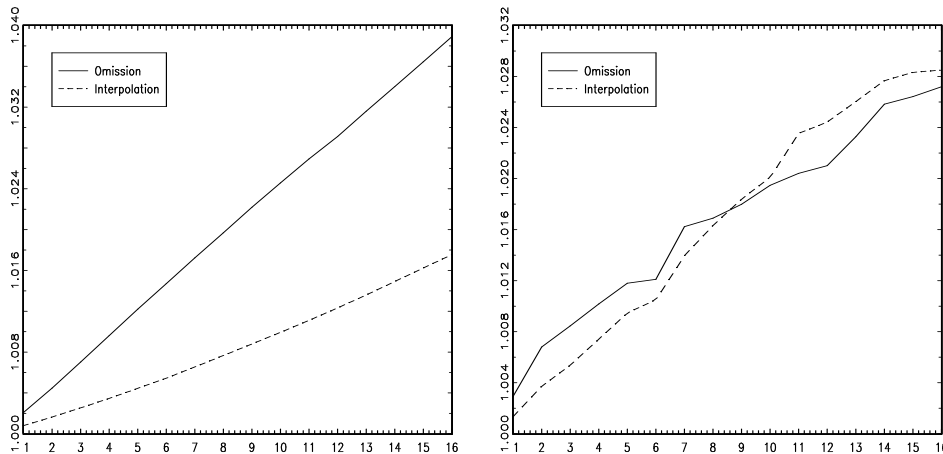


Figure 1: Relative average RMSE and RMSFE. Horizontal axis is total number of regressors.

7. Synopsis of the Estimation Technique

We conclude this analysis with a brief summary of the straightforward steps involved in feasibly estimating β in the cointegrated system given by (4) and (5) in a way that allows valid statistical analysis in large samples.

- [1] **Verify the Conditions for Theorem 5.1.** If the messy-data noise are driven by an explicit structural mechanism, such as imputation or interpolation, these conditions must be verified. As we showed in Section 6, these conditions are satisfied for messy data driven by linear interpolation in a general mixed-frequency context.
- [2] **Estimate α , ψ , μ , and Γ .** Least squares estimation of (7) provides consistent estimates of μ and Γ . If the regressors (x_t) are not cointegrated with each other, let $\hat{C} = I$. Otherwise, let $\hat{C} = (\hat{\Gamma}'\hat{\Gamma})^{-1}\hat{\Gamma}'$. Least squares estimation of (6) provides consistent estimates of α and ψ .
- [3] **Estimate Ω , Σ , and Δ .** Use observable (w_t) and (Δq_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 should use a kernel and a lag truncation parameter satisfying [K1]-[K3].
- [4] **Create (x_t^{**}) and (y_t^{**}) .** Use (19) and (20) with consistent parameter estimates.
- [5] **Re-Estimate ψ .** Least squares estimation of the CCR given by (18) provides consistent *and asymptotically normal* estimates.
- [6] **Identify β .** Use $\hat{\beta}_{CCR} = \hat{C}'\hat{\psi}_{CCR}$ and $\mathbf{Var}(\hat{\beta}_{CCR}) = \hat{C}'\mathbf{Var}(\hat{\psi}_{CCR})\hat{C}$ to identify $\hat{\beta}$ and its standard errors.

Other than the initial verification step, each of these steps are required for using a CCR to estimate a standard cointegrating regression with cointegrated regressors and well-identified parameters.

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.
- Bound, J., C. Brown, and N. Mathiowetz (2001). "Measurement Error in Survey Data," in *Handbook of Econometrics*, vol. V, J.J. Heckman and E. Leamer eds., Amsterdam: Elsevier.
- Chang, Y., J.I. Miller, and J.Y. Park (2006). "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.
- Fischer, A.M. (1990). "Cointegration and I(0) Measurement Error Bias," *Economics Letters*, 34, 255-9.
- Friedman, M. (1962). "The Interpolation of Time Series by Related Series," *Journal of the American Statistical Association*, 57, 729-57.

- Fuller, W.A. (1987). *Measurement Error Models*. New York: John Wiley & Sons.
- 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 (2000). "A New Historical Database for the NYSE 1815 to 1925: Performance and Predictability," Yale ICF Working Paper No. 00-13; Yale SOM Working Paper No. ICF-00-13.
- Gomez, V. and A. Maravall (1994). "Estimation, Prediction, and Interpolation for Non-stationary Series with the Kalman Filter," *Journal of the American Statistical Association*, 89, 611-24.
- Griliches, Z. (1986). "Economic Data Issues." In Z. Griliches and M. Intriligator, eds., *Handbook of Econometrics*, 3, Amsterdam: North-Holland.
- 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.
- Hu, Y. and S.M. Schennach (2007). "Instrumental Variable Treatment of Nonclassical Measurement Error Models," *Econometrica*, forthcoming.
- 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.
- Little, R.J.A. and D.B. Rubin (2002). *Statistical Analysis with Missing Data*. Hoboken, New Jersey: John Wiley & Sons.
- 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.
- Meijering, E. (2002). "A Chronology of Interpolation: From Ancient Astronomy to Modern Signal and Image Processing," *Proceedings of the IEEE*, 90, 319-42.
- 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.
- Seong, B., S.K. Ahn, and P.A. Zadorozny (2007). “Cointegration Analysis with Mixed-Frequency Data,” CESifo Working Paper No. 1939.
- Shafer, J.L. (1997). *Analysis of Incomplete Multivariate Data*. New York: Chapman & Hall.

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 $\zeta_t \equiv (b'_t, z_t^{*'})'$ be an $(R+r) \times 1$ vector.

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

- [a] $(z_t^* z_t^{*'})$ 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 $(z_t^* z_t^{*'})$ 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) .⁶ 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. \square

⁶The reader is referred to Example 17.3 in Davidson (1994).

Lemma A.2 Let [NED1] hold. Then

- [a] $n^{-1} \sum z_t^* = o_p(1)$ under [NED2], [NED4'],
- [b] $n^{-1} \sum (z_t^* z_t^{*'} - \Sigma_{**}) = o_p(1)$ 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] $n^{-1} \sum q_t z_t^{*'} \rightarrow_d \int QdZ^*(s)' + \Delta'_{*q}$ under [A1]-[A4], [NED2], [NED4'], [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 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 $d_{it}^z = O(\|z_{it}^*\|_2)$ as $n \rightarrow \infty$, 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

$$t^{-1} \|z_{it}^*\|_2 = O(t^{-5/6-\varepsilon})$$

for some $\varepsilon > 0$,⁷ 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 $(z_t^* z_t^{*'})$, 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

$$\|z_{i,t}^* z_{j,t}^{*'} - e_i' \Sigma_{**} e_j\|_2 \leq \|z_{it}^*\|_4 \|z_{jt}^{*'}\|_4 + e_i' \Sigma_{**} e_j$$

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

⁷The 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 $a > 1$.

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

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

which under [NED5.a] is in fact unity for each i .

[e] The proof proceeds as in part [d], by looking the random vector $n^{-1/2}\Omega_{**w_j^*w_j}^{-1/2}z_t^*w'_te_j$, which corresponds to the j^{th} column of $z_t^*w'_t$. To consider the i^{th} element of this vector, we look at $n^{-1/2}e'_i\Omega_{**w_j^*w_j}^{-1/2}z_t^*w'_te_j$, which has an L_2 -norm of

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

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

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

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 $j = 1, \dots, R+r$. 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] 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, $\mathbf{E}Q_n Q_n'(s)$, $\mathbf{E}Z_n^* Q_n'(s)$, and $\mathbf{E}Z_n^* Z_n^{*'}(s)$ must have finite limits. \square

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] $\frac{1}{n} \sum x_t^* v_t^* \rightarrow_d M^*(s)$,
- [b] $\frac{1}{n} \sum x_t^* w_t' \rightarrow_d \Gamma(\int Q dW(s)' + \Delta'_{wq}) + \Sigma_{uw} + \Sigma_{*w}$,
- [c] $\frac{1}{n} \sum w_t v_t^* \rightarrow_p \sigma_{wv} - \Sigma_{w*} C' \psi$, and
- [d] $\frac{1}{n^2} \sum x_t^* x_t^{*'} \rightarrow_d N(s)$

as $n \rightarrow \infty$.

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 (2) and our definition of (v_t^*) . Expanding the product yields

$$\sum x_t^* v_t^* = \sum x_t v_t - \sum x_t z_t^{*'} C' \psi + \sum z_t^* v_t - \sum z_t^* z_t^{*'} C' \psi, \quad (23)$$

and to find the limiting distribution of the first term of (23), we may further expand this term using the data generating process of (x_t) given by (5). We obtain

$$\frac{1}{n} \sum (\mu + \Gamma q_t + u_t) v_t \rightarrow_d \Gamma \left(\int Q dV(s) + \delta'_{vq} \right) + \sigma_{uv}$$

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 (23) has a distribution given by

$$-\frac{1}{n} \sum (\mu + \Gamma q_t + u_t) z_t^{*'} C' \psi \rightarrow_d -\Gamma \left(\int Q dZ^*(s)' + \Delta'_{*q} \right) C' \psi - \Sigma_{u*} C' \psi$$

using Lemma A.2[a], [f], and [c], respectively. The third and fourth terms of (23) 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 (2) 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

$$\sum w_t v_t^* = \sum w_t v_t - \sum w_t z_t^{*'} C' \psi,$$

and, again, the stated result immediately follows.

[d] Finally, expanding the summation in part [d] reveals a structure similar to part [a]. Specifically,

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

where the first term has an asymptotic distribution of

$$\frac{1}{n^2} \sum x_t x_t' \rightarrow_d \Gamma \int Q Q' \Gamma',$$

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

$$[a] \frac{1}{n} \sum x_t^{**} v_t^{**} \rightarrow_d \Gamma \int Q dV_{\perp Q}^*(s)$$

$$[b] \frac{1}{n} \sum x_t^{**} w_t' \rightarrow_d \Gamma \int Q dW(s)', \text{ and}$$

$$[c] \frac{1}{n^2} \sum x_t^{**} x_t^{**'} \rightarrow_d \Gamma \int Q Q' \Gamma'$$

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 (19) and (21) as

$$\sum x_t^* v_t^* - \sum x_t^* \hat{\omega}_{vq} \hat{\Omega}_{qq}^{-1} \Delta q_t - (\hat{\Gamma} \hat{\Delta}'_{bq} + \hat{\Sigma}_{ub}) \hat{\Sigma}_{bb}^{-1} \sum \hat{b}_t^* \hat{b}_t^{*'} \hat{\kappa} \quad (24)$$

using feasible estimators of all parameters. The distribution of the first term of (24) follows from Lemma A.3[a]. The second term of (24) may be written as

$$- \left(\sum x_t \Delta q'_t + \sum z_t^* \Delta q'_t \right) \hat{\Omega}_{qq}^{-1} \hat{\omega}_{qv}, \quad (25)$$

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 (25) has an asymptotic distribution given by

$$\Gamma \left(\int Q dQ(s)' + \Delta'_{qq} \right) + \Sigma_{uq},$$

and the probability limit of the second summation in (25) is Σ_{*q} when similarly normalized. To determine the limit of the final term of (24), we need to deal with the limit of $\sum \hat{b}_t^* \hat{b}_t^{*'}$. Expanding this as

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

it is clear using (14) 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

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

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 (19), the summation is equal to

$$\sum x_t^* w_t' - (\hat{\Gamma} \hat{\Delta}'_{bq} + \hat{\Sigma}_{ub}) \hat{\Sigma}_{bb}^{-1} \sum \hat{b}_t^* \hat{b}_t^{*'} (0, I, 0, 0)',$$

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 $-(\Gamma \Delta'_{wq} + \Sigma_{uw} + \Sigma_{*w})$.

[c] Expanding the summation yields

$$\begin{aligned} & \sum x_t^* x_t^{*'} + (\hat{\Gamma} \hat{\Delta}'_{bq} + \hat{\Sigma}_{ub}) \hat{\Sigma}_{bb}^{-1} \sum \hat{b}_t^* \hat{b}_t^{*'} \hat{\Sigma}^{-1} (\hat{\Gamma} \hat{\Delta}'_{bq} + \hat{\Sigma}_{ub})' \\ & - \sum x_t^* \hat{b}_t^{*'} \hat{\Sigma}_{bb}^{-1} (\hat{\Gamma} \hat{\Delta}'_{bq} + \hat{\Sigma}_{ub})' - (\hat{\Gamma} \hat{\Delta}'_{bq} + \hat{\Sigma}_{ub}) \hat{\Sigma}_{bb}^{-1} \sum \hat{b}_t^* x_t^{*'} \end{aligned} \quad (26)$$

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 4.5. The summation $\sum x_t^* \hat{b}_t^{*'}$ in the third and fourth terms of (26) may be partitioned as

$$\sum (x_t^* v_t^*, x_t^* w_t', x_t^* u_t^{*'}, x_t^* \Delta q_t'), \quad (27)$$

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

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

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 $O_p(n)$ as a special case of the first. The third partition in (27) admits the expansion

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

using (2) and (14). All of these are $o_p(n^2)$ under our assumptions. The last partition in (27) is $O_p(n)$ for the same reasons as the second partition. Finally, returning to the third and fourth terms of (26), since $\hat{\Sigma}^{-1} (\hat{\Gamma} \hat{\Delta}'_{bq} + \hat{\Sigma}_{ub})' = O_p(1)$, the proof is complete. \square

Lemma A.5 Let [A1]-[A3] hold and define $\Sigma_{bb}(k) = \mathbf{E} b_t b_{t-k}'$ to be the k^{th} autocovariance of (b_t) . For any i, j ,⁸ we have

$$\lim_{l \rightarrow \infty} \mathbf{cov}(B_l^{(j)}(s), B_l^{(j-i)}(s)) = s \Omega_{bb}^{(i)}$$

as $l \rightarrow \infty$, where

$$B_l^{(j)}(s) \equiv l^{-1/2} \sum_{p=1}^{\lfloor ls \rfloor} b_{\tau_{p-1}+j} \quad \text{and} \quad \Omega_{bb}^{(i)} \equiv \sum_{k=-\infty}^{\infty} \Sigma_{bb}(k(m+1)+i)$$

are a stochastic process and limiting variance, respectively.

Proof of Lemma A.5 The covariance is equal to

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

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

⁸The result holds for any i, j , although we generally restrict $j \in [1, m+1]$, due to construction of the messy data error.

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

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

and the stochastic processes $Z_n^*(s)$ as in (8) and with (z_t^*) defined by (22). We have

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

as $n \rightarrow \infty$.

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 $B_n(s) \rightarrow_d B(s)$, 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

$$\begin{aligned}
n^{-1/2} \sum_{j=1}^{m+1} \sum_{p=1}^{[ls]} z_{\tau_p+j}^* &= n^{-1/2} \sum_{j=1}^{m+1} \left(1 - \frac{j}{m+1}\right) \sum_{p=1}^{[ls]} u_{\tau_{p-1}} \\
&+ n^{-1/2} \sum_{j=1}^{m+1} \frac{j}{m+1} \sum_{p=1}^{[ls]} u_{\tau_p} - n^{-1/2} \sum_{j=1}^{m+1} \sum_{p=1}^{[ls]} u_{\tau_{p-1}+j} \\
&+ \frac{\Gamma}{n^{1/2}} \sum_{j=1}^{m+1} \left(\frac{j}{m+1} - 1\right) \sum_{i=1}^{m+1} \sum_{p=1}^{[ls]} \Delta q_{\tau_{p-1}+i} \\
&+ \frac{\Gamma}{n^{1/2}} \sum_{j=1}^{m+1} \sum_{i=j+1}^{m+1} \sum_{p=1}^{[ls]} \Delta q_{\tau_{p-1}+i}
\end{aligned} \tag{28}$$

where the first two terms of (28) may be rewritten as

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

for large l . This has a limiting distribution given by $(m+1)^{1/2} U^{(0)}(s)$ as a direct result of part [b]. The third term is

$$(m+1)^{-1/2} \sum_{j=1}^{m+1} U_l^{(j)}(s)$$

which has a limiting distribution of $U(s)$ according to part [c]. Similarly, using part [c], the limiting distribution of the fourth term of (28) is $-\frac{m}{2}\Gamma Q(s)$. Finally, the distribution of the last term of (28) is

$$\frac{\Gamma}{(m+1)^{1/2}} \sum_{j=1}^{m+1} \sum_{i=j+1}^{m+1} Q^{(i)}(s)$$

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,

$$\frac{1}{n} M_n^* \rightarrow_d M^*(s)$$

as $n \rightarrow \infty$. Similarly,

$$\frac{1}{n^2} N_n^* = \frac{1}{n^2} \sum x_t^* x_t^{*'} + o_p(1) \rightarrow_d N(s),$$

so that the stated result is obtained. The choice of C ensures that $CN(s)C'$ is invertible, even though $N(s)$ is not. \square

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

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

The first term is $O_p(n)$ under [A1]-[A3], using standard asymptotics for integrated series. The second term is also $O_p(n)$, since

$$n^{-1/2} \bar{q}_n \rightarrow_d \int Q,$$

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 $O_p(n)$ by Lemma A.2[f] and [d], respectively. It remains to show that the second summation in the estimator is $O_p(n^2)$. This is straightforward, since

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

as $n \rightarrow \infty$. This is invertible since the trends (q_t) are distinct. \square

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

$$\left(\sum w_t w_t' \right)^{-1} \sum w_t \left((\psi' C - \hat{\psi}'_{LS} \hat{C}) x_t + (v_t - \hat{\psi}'_{LS} \hat{C} z_t^*) \right)$$

by substituting (2) and our definition of (v_t^*) into (10). Under our assumptions, $\sum w_t w_t' = O_p(n)$ and invertible. Under [A5] and Lemma A.2[e], we have

$$\sum w_t (v_t - z_t^{*'} \hat{C}' \hat{\psi}_{LS}) = O_p(n^{1/2}).$$

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 $O_p(1/n)$ under our assumptions. Thus,

$$\sum w_t x_t' (C' \psi - \hat{C}' \hat{\psi}_{LS}) = O_p(1),$$

and whole estimator is $O_p(n^{-1/2})$.

[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 (11) and (7). 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]. \square

Proof of Lemma 4.4 The estimator (12) may be written as

$$\tilde{\Omega}_{b^*b^*} = (I, D) \left\{ \frac{1}{n} \sum \sum_{s=1}^n \zeta_t \zeta_s' \pi \left(\frac{t-s}{\ell_n} \right) \right\} (I, D)', \quad (29)$$

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 $\zeta_t - \mathbf{E}(\zeta_t | \mathcal{F}_{t-K}^{t+K})$,

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

for $i = 1, \dots, R$ so that

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

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

$$d_t^{b,z} = (R+r)^{1/2} \max_i d_{it} \quad \text{and} \quad \nu_K^{b,z} = \max_i \nu_{iK},$$

and since the sequences (d_{it}) are bounded and there are a finite number of regressors in (4), $(d_t^{b,z})$ 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 (29) is consistent. \square

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

$$\left| \hat{\Delta}_{b^*b^*} - \tilde{\Delta}_{b^*b^*} \right| \leq \frac{1}{\sqrt{n}} \sum_{k=0}^n \left| \pi \left(\frac{k}{\ell_n} \right) \right| \left(\frac{1}{\sqrt{n}} \sum_{t=k+1}^n \left| \hat{b}_t^* (\hat{b}_{t-k}^* - b_{t-k}^*)' \right| + \frac{1}{\sqrt{n}} \sum_{t=k+1}^n \left| (\hat{b}_t^* - b_t^*) b_{t-k}^{*'} \right| \right) \quad (30)$$

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 $o(n^{1/2})$ by [K3], the result holds if we can show that the remaining sums in the majorant of (30) are both $O_p(n^{1/2})$.

Consider the second summation in the majorant of (30). The ij^{th} element of this matrix is

$$\begin{aligned} \frac{1}{\sqrt{n}} \sum_{t=k+1}^n \left| e_i' \hat{b}_t^* (\hat{b}_{t-k}^* - b_{t-k}^*)' e_j \right| &\leq \left(\frac{1}{n} \sum \left| e_i' \hat{b}_t^* \hat{b}_t^{*'} e_i \right| \right)^{1/2} \\ &\times \left(\sum \left| e_j' (\hat{b}_t^* - b_t^*) (\hat{b}_t^* - b_t^*)' e_j \right| \right)^{1/2} \end{aligned} \quad (31)$$

using the Cauchy-Schwarz inequality and the non-negativity of k . We may expand the first summation in the majorant of (31) (up to premultiplication by e_i' and postmultiplication by e_i) as

$$\frac{1}{n} \sum \hat{b}_t \hat{b}_t' + \frac{1}{n} \sum \hat{D} z_t^* z_t^{*'} \hat{D}' + \frac{1}{n} \sum \hat{b}_t z_t^{*'} \hat{D}' + \frac{1}{n} \sum \hat{D} z_t^* \hat{b}_t' \quad (32)$$

using (15). 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 (Δq_t) are stationary, and (\hat{v}_t) and (\hat{u}_t) consistently estimate stationary series (v_t) and (u_t) by way of (14) and the rates of convergence of $\hat{\psi} - \psi$, etc. already established under our assumptions. The second term of (32) 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 (32) are slightly more complicated. They involve sums of both outer products of (w_t) and (Δq_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

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

and

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

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 (32) are therefore $O_p(1)$.

An expansion of the second summation in the majorant of (31) (up to premultiplication by e_j' and postmultiplication by e_j) yields

$$\begin{aligned} &\sum (\hat{b}_t - b_t) (\hat{b}_t - b_t) + (\hat{D} - D) \sum z_t^* z_t^{*'} (\hat{D} - D)' \\ &+ \sum (\hat{b}_t - b_t) z_t^{*'} (\hat{D} - D)' + (\hat{D} - D) \sum z_t^* (\hat{b}_t - b_t)', \end{aligned} \quad (33)$$

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^{*'} 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 (33) is also $O_p(1)$. Turning to the third and fourth terms of (33), the vector $(\hat{b}_t - b_t)$ contains zeros for observable series (w_t) and (Δq_t) . For the subseries (\hat{v}_t) and (\hat{u}_t) of estimates, we must contend with summations$

$$(\alpha - \hat{\alpha})' \sum w_t z_t^{*'} (\hat{D} - D)' + (\psi' C - \hat{\psi}' \hat{C}) \sum x_t z_t^{*'} (\hat{D} - D)$$

and

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

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 (33) is $O_p(1)$. Consequently, the second summation in the majorant of (31) 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 (30) 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 (31). This completes the proof for $\hat{\Delta}_{b^* b^*}$. \square

Proof of Theorem 5.1 The estimator $\hat{\psi}_{CCR}$ may be rewritten as

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

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

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

using the Slutsky theorem. 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 of $(\hat{\psi}_{CCR} - \psi)$ follows directly from Lemma A.4[a] and [c]. \square

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

$$\begin{aligned} \frac{1}{n} \sum z_t^* &= \frac{1}{m+1} \sum_{j=1}^{m+1} \left(1 - \frac{j}{m+1} \right) \frac{1}{l} \sum_{p=1}^l u_{\tau_{p-1}} + \frac{1}{(m+1)^2} \sum_{j=1}^{m+1} j \frac{1}{l} \sum_{p=1}^l u_{\tau_p} \\ &\quad - \frac{1}{m+1} \sum_{j=1}^{m+1} \frac{1}{l} \sum_{p=1}^l u_{\tau_{p-1}+j} - \frac{1}{m+1} \Gamma \sum_{j=1}^{m+1} \sum_{i=1}^j \frac{1}{l} \sum_{p=1}^l \Delta q_{\tau_{p-1}+i} \\ &\quad + \frac{1}{(m+1)^2} \Gamma \sum_{j=1}^{m+1} j \sum_{i=1}^{m+1} \frac{1}{l} \sum_{p=1}^l \Delta q_{\tau_{p-1}+i} \end{aligned} \quad (34)$$

since $z_t^* = 0$ for $t = \tau_p$ with $p = 1, \dots, l$ 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 (34) 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 (34) 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 (34) 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 $\frac{1}{n} \sum z_t^* z_t^{*'}$ is finite and independent of t . An expansion of this matrix using (22) 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 (22). 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 (22) and the fact that $n/l = m + 1$, this representative term is

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

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

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

which is finite by construction. The remaining 23 terms of the expansion of $\frac{1}{n} \sum z_t^* z_t^{*'}$ 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 $\frac{1}{n} \sum q_t z_t^{*'}$ whose distribution we must verify may be expanded as

$$\begin{aligned} & -\frac{1}{n} \sum q_t u'_t + \frac{1}{n} \sum_{j=1}^{m+1} \sum_{p=1}^l q_{\tau_{p-1}+j} u'_{\tau_{p-1}} - \frac{1}{n} \sum_{j=1}^{m+1} \sum_{i=1}^{m+1} \sum_{p=1}^l \Delta q_{\tau_{p-1}+j+i} u'_{\tau_p} + o_p(1) \quad (35) \\ & + \frac{1}{n} \sum_{j=1}^{m+1} \left(\frac{j}{m+1} - 1 \right) \sum_{i=1}^{m+1} \sum_{p=1}^l q_{\tau_{p-1}+j} \Delta q'_{\tau_{p-1}+i} \Gamma' + \frac{1}{n} \sum_{j=1}^{m+1} \sum_{i=j+1}^{m+1} \sum_{p=1}^l q_{\tau_{p-1}+j} \Delta q'_{\tau_{p-1}+i} \Gamma' \end{aligned}$$

since

$$\sum_{p=1}^l q_{\tau_{p-1}+j} u'_{\tau_p} = \sum_{p=1}^l q_{\tau_p+j} u'_{\tau_p} - \sum_{p=1}^l \sum_{i=1}^{m+1} \Delta q_{\tau_{p-1}+j+i} u'_{\tau_p}$$

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 (35) has a limiting distribution of

$$\int Q dU(s)' + \Delta'_{uq}$$

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

$$n^{-1/2} \sum_{j=1}^{m+1} \sum_{p=1}^{\lfloor ls \rfloor} u'_{\tau_{p-1}} \rightarrow_d (m+1)^{1/2} U^{(0)}(s)$$

using Lemma A.6[b]. The limiting distribution of the second term of (35) is therefore

$$(m+1)^{1/2} \int Q dU^{(0)}(s)' + \Delta'_{uq} + \sum_{i=1}^j \Sigma'_{uq}(-i)$$

along similar lines as the first. The limiting distribution of the third term of (35) is simply

$$\frac{1}{m+1} \sum_{j=1}^{m+1} \sum_{i=1}^{m+1} \mathbf{E} \Delta q_{\tau_{p-1}+j+i} u'_{\tau_p} = \frac{1}{m+1} \sum_{j=1}^{m+1} \sum_{i=1}^{m+1} \Sigma'_{uq}(m+1-(j+i))$$

using an LLN. The summation over i in the fourth term of (35) may be expanded as

$$\begin{aligned} & \frac{1}{n} \sum_{j=1}^{m+1} \left(\frac{j}{m+1} - 1 \right) \sum q_t \Delta q'_t \Gamma' \\ & + \frac{1}{n} \sum_{j=1}^{m+1} \left(\frac{j}{m+1} - 1 \right) \sum_{p=1}^l \sum_{i=1}^{j-1} \sum_{k=i+1}^j \Delta q_{\tau_{p-1}+k} \Delta q'_{\tau_{p-1}+i} \Gamma' \\ & - \frac{1}{n} \sum_{j=1}^{m+1} \left(\frac{j}{m+1} - 1 \right) \sum_{p=1}^l \sum_{i=j+1}^{m+1} \sum_{k=j+1}^i \Delta q_{\tau_{p-1}+k} \Delta q'_{\tau_{p-1}+i} \Gamma' \end{aligned} \quad (36)$$

The limiting distribution of the first term of (36)

$$-\frac{m}{2} \left(\int Q dQ(s)' + \Delta'_{qq} \right) \Gamma'$$

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 (35) 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 $(\Delta q_{\tau_{p-1}+i})$ 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

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

Note that

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

by standard limit theory and Lemma A.6[c], where

$$\Delta_{qq}^{(k-i)} \equiv \sum_{r=1}^{\infty} \Sigma_{qq}(r(m+1) + k - i)$$

comes from the covariance of increments of $Q_l^{(k)}$ and $Q_l^{(i)}$. Thus,

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

Moreover,

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

using an LLN, since the increments are stationary and mixing by [A1]. Collecting terms provides a distribution of

$$\int Q dZ^*(s)' + \Delta_{q^*}$$

with $Z^*(s)$ defined as in Lemma A.6 and Δ_{q^*} defined implicitly by the remaining (non-stochastic) 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 (Δq_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

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

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

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

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

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

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 (22) involving (u_t) .

Similarly, since the series (Δq_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 (Δq_t) are simply linear combinations of $(\Delta q_{\tau_{p-1}+i})$, 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 (22) is NED with the required properties.

Verification of [NED4'] The Minkowski inequality allows

$$\begin{aligned} \sup_{p \leq l, j \leq m} \left\| z_{\tau_{p-1}+j}^* \right\|_{2a/(a-1)} &\leq \sup_{p \leq l, j \leq m} \left\| u_{\tau_{p-1}+j} \right\|_{2a/(a-1)} + \sup_{p \leq l} \left\| u_{\tau_{p-1}} \right\|_{2a/(a-1)} \\ &\quad + \sup_{p \leq l} \left\| u_{\tau_p} \right\|_{2a/(a-1)} \\ &\quad + 2\Gamma(m+1) \sup_{p \leq l, j \leq m} \left\| \Delta q_{\tau_{p-1}+j} \right\|_{2a/(a-1)} \end{aligned}$$

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