Inference in Conditional Vector Error-Correction Models with Small Signal-to-Noise Ratio*

Nikolay Gospodinov†
Federal Reserve Bank of Atlanta

Alex Maynard‡
University of Guelph

Elena Pesavento§
Emory University

October 2021

Abstract

It is widely documented that while contemporaneous spot and forward financial prices trace each other extremely closely, their difference is often highly persistent and the conventional cointegration tests may suggest lack of cointegration. This paper studies the possibility of having a cointegrated errors that are characterized simultaneously by high persistence (near unit root behavior) and very small (near zero) variance. The proposed dual parameterization induces the cointegration error process to be stochastically bounded which prevents the variables in the cointegrating system from drifting apart over a reasonably long horizon. More specifically, the paper develops the appropriate asymptotic theory (rate of convergence and asymptotic distribution) for the estimators in unconditional and conditional vector error-correction models when the error correction term is parameterized as a dampened near unit root process (local-to-unity process with local-to-zero variance). The important differences in the limiting behavior of the estimators and their implications for empirical analysis are discussed. Simulation results and an empirical analysis of the forward premium regressions are also provided.

JEL Classification: C12, C15, C22, F31.

Keywords: Cointegration; Vector error correction models; Local-to-unity asymptotics; Local-to-zero variance; Spot and forward exchange rates; Forward premium regression.

---

*This paper is prepared for the conference and festschrift in honor of Joon Y. Park whose work on nonstationarity in economic and financial data had an enormous influence on our research and analytical approach. We would like to thank the Co-Editor and an anonymous referee for insightful suggestions. We are also grateful to Gregory Cox, Christian Gourieroux, Joann Jasiak, and participants at the 2018 IAAE Meetings and the 26th Annual Meeting of the Midwest Econometrics Group for useful comments and discussions on an earlier version of the paper. The authors gratefully acknowledge funding under SSHRC Grant 430417. The views expressed here are the authors' and not necessarily those of the Federal Reserve Bank of Atlanta or the Federal Reserve System.

†Research Department, Federal Reserve Bank of Atlanta, 1000 Peachtree St. N.E., Atlanta, Georgia, 30309–4470, USA; Email: nikolay.gospodinov@atl.frb.org.

‡Department of Economics, Gordon S. Lang School of Business and Economics, University of Guelph, Guelph, Ontario, N1G 2W1, Canada; Email: maynarda@uoguelph.ca.

§Department of Economics, Emory University, Atlanta, Georgia, 30322–2240, USA; Email: epesave@emory.edu.

1 Introduction and Motivation

This paper considers some important inference issues that arise in the analysis of nearly cointegrated processes in the presence of highly persistent cointegrating errors whose variability is only a small fraction of the variance of the original variables. Equivalently, in the vector error correction (VEC) representation of the cointegrated system, the error correction term is near-integrated with low signal-to-noise ratio. Typical examples of this setup include models that study the unbiasedness of forward and futures prices (exchange rates, interest rates, commodity prices) for the expected future spot values. For instance, spot and one-month forward exchange rates trace each other very closely and are virtually indistinguishable from each other as illustrated in the left panel of Figure 1 for the British pound (BP), German mark (DM), Swiss franc (SF) and Canadian dollar (CD) – all against the US dollar. And yet, the spot-forward spread (the difference between the two series) is characterized by high persistence which becomes visible when plotted in isolation in the right panel of Figure 1. In fact, a formal unit root test on the spot-forward spread often cannot reject the null hypothesis of a lack of cointegration. The heuristic reason for this is that the spot-forward spread has a tiny variance compared to the variability of the individual variables and this prevents its near random walk component from forcing spot and forward rates to drift apart in the long run. Similar arguments apply to the time series behavior of cash and futures prices of other asset classes, such as commodities or bond yields.

Figure 1 about here

To accommodate this empirical regularity without compromising the integrity of the cointegrating system, we model this component as a dampened (stochastically bounded) near unit root process. More generally, this parameterization proves to be a useful device in reconciling the internal consistency of the statistical behavior with the widely-held belief that many economic and financial time series are driven by a slowly moving, low-frequency persistent component (Bansal and Yaron, 2004; Gourieroux and Jasiak, 2020;[1] Phillips and Lee, 2013; among others).[2] Consider, for instance, the unobserved component (local level) model

$$\begin{aligned} x_t &= \mu_t + u_t, \\ \mu_t &= \rho\mu_{t-1} + \tau\xi_t, \end{aligned}$$

[1] Gourieroux and Jasiak (2020) employ a model similar to ours in order to explain the long-run predictability puzzle, whereas we focus primarily on explaining puzzles involving spot and forward rates that move very closely together, and yet are barely cointegrated according to traditional cointegration tests.

[2] Müller and Watson (2008, 2018) provide a comprehensive analytical framework for studying low-frequency movements and co-movements in economic and financial time series. In this paper, we consider the possibility that the low-frequency component is not readily detectable, as the explosive behavior arising from its high persistence is offset by its asymptotically vanishing variability.

where μ_t is a highly persistent and possibly unit root component (with ρ near 1), u_t and ξ_t are mutually uncorrelated white noise disturbances with variance σ^2, and τ is the signal-to-noise ratio. The observed variable x_t could be consumption growth, dividend growth or equity returns and the long-run risks associated with μ_{t+1}, which is largely interpreted as capturing the common variation in real activity (Bansal and Yaron, 2004), play a crucial role in explaining the equity premium puzzle. While the above representation is theoretically appealing, the empirical evidence on the existence of such a long-run component in stock returns is rather weak. There are two main questions that drive a wedge between the theoretical setup and the empirical justification of these low-frequency components. First, how come we do not detect this persistence in the data? And second, how can we reconcile the statistical behavior of the model and the data as the sample size increases? After all, the persistent component μ_t has to dominate the dynamics of the observed series as the number of time series observations grows.

To reconcile this tension, it is convenient to adopt a dual localization and model μ_t as a dampened near-unit root process (see Gospodinov, 2009). More specifically, let $\mu_t = \rho_T \mu_{t-1} + \tau_T \xi_t$ denote the low-frequency component. The dual localization involves (a) a local-to-unity parameterization $\rho_T = 1 + c/T$ for some fixed constant $c \leq 0$, and (b) a local-to-zero parameterization for the signal-to-noise ratio $\tau_T = \lambda/\sqrt{T}$ for some fixed constant $\lambda > 0$. This dual localization proves to be instrumental in producing a process that is stochastically bounded and hence consistent with both statistical and economic theory. Unlike regular near-unit root processes that are of order $O_p(T^{1/2})$, the local-to-zero variance localization dampens the stochastic trend behavior of x_t and keeps it stochastically bounded ($O_p(1)$). The dual localization removes the economically unappealing possibility that the low-frequency component can wander off and induce non-stationarity in asset returns or consumption growth. The persistent and noise components of the model now have comparable orders of magnitude as both μ_t and the rest of the variables are stochastically bounded. While this statistical device renders the model congruent, the observed stock returns at the monthly or quarterly frequency are still overwhelmed by noise and the empirical detection of this small low-frequency component remains elusive.

With this background in mind, the paper derives the theoretical implications of the simultaneous presence of high persistence, low variability and endogeneity of the cointegrating errors for the concept of cointegration, the properties of cointegrating regressions, estimation and testing in vector error-correction models, etc. More specifically, we develop the appropriate theory (rate of convergence and asymptotic distributions) for the estimators in VEC and conditional VEC models (Phillips, 1991; Johansen, 1992; Boswijk, 1994) when the error correction term is parameterized as a dampened near unit root process (local-to-unity process with local-to-zero variance). In doing this, we combine the literatures on near cointegration (Zivot, 2000; Jansson and Haldrup, 2002;

Pesavento, 2004; Elliott, Jansson and Pesavento, 2005) and near zero variance regressors (Torous and Valkanov, 2000; Moon, Rubia and Valkanov, 2004; Gospodinov, 2009; Deng, 2014; Gourieroux and Jasiak, 2020). This double local parameterization of the persistence and variance of the cointegration errors provides a powerful tool for deriving limiting results by capturing the salient features of the data in the empirical examples. One important result that emerges from our analysis is that the estimator in the conventional VEC models is characterized by large bias, a reduced rate of convergence and a highly dispersed asymptotic distribution, while its conditional counterpart enjoys a substantially improved asymptotic behavior. The paper provides a detailed investigation of the numerical properties of the estimators in unconditional and conditional VEC models and the empirical size and power of tests for cointegration based on the corresponding test statistics. The practical importance of the analytical results is demonstrated in the context of exchange rate models.

The rest of the paper is organized as follows. Section 2 introduces the analytical setup, modeling assumptions and appropriate limits. It also presents the main representations that characterize the asymptotic behavior of the estimators and their corresponding t-tests. Section 3 contains simulation results while Section 4 reports the results from the empirical application for spot and forward exchange rates. Section 5 concludes.

2 Model and Main Results

2.1 Assumptions and Parameterization

First, we discuss the model setup, assumptions and the proposed dual local parameterization. Suppose that $(x_t', y_t)'$ is a $((k+1) \times 1)$ vector generated by the triangular system[3]

$$\begin{aligned} x_t &= \psi_x + \phi_x t + u_{x,t} \\ y_t &= \psi_y + \phi_y t + \gamma' x_t + u_{y,t} \end{aligned} \tag{1}$$

and

$$\begin{pmatrix} (1-L)u_{x,t} \\ (1-\rho_T L)u_{y,t} \end{pmatrix} = \begin{pmatrix} v_{x,t} \\ \tau_T v_{y,t} \end{pmatrix}$$

with

$$A(L)\, v_t = \varepsilon_t$$

for $t = 1, ..., T$. We make the following assumptions.

Assumption A *Assume that* $\max_{-p \leq t \leq 0} \left\| \left(u_{x,t}', u_{y,t}\right)' \right\| = O_p(1)$, *where* $\|\cdot\|$ *is the Euclidean norm.*

[3] For notational convenience, we suppress the dependence of u_t, x_t, and y_t on T.

Assumption B *Assume that* $A(L) = I_{k+1} - \sum_{i=1}^{p} A_i L^i$ *is a matrix polynomial of a finite (known) order* p *in the lag operator* L*, with roots that lie outside the unit circle.*

Assumption C *Assume that* $\varepsilon_t = (\varepsilon_{x,t}', \varepsilon_{y,t})'$ *is a homoskedastic martingale difference sequence with a variance matrix* Σ*,* $0 < \|\Sigma\| < \infty$*, and finite fourth moments,* $\max_i E(\varepsilon_{it}^4) < \infty$*.*

Our model resembles the standard triangular model for cointegration (Phillips, 1991; see also Engle and Granger, 1987; Park and Phillips, 1988, 1989; Stock and Watson, 1993; Park, 1992; among many others) but we allow the error in the cointegration regression to be persistent, yet still bounded. For clarity of exposition, the analytical results below are presented for the case of no deterministic terms in model (1); i.e., $\psi_x = 0$, $\psi_y = 0$, $\phi_x = 0$ and $\phi_y = 0$. The generalization to deterministic terms is straightforward to obtain at the expense of additional notation (see Section 2.2 for further discussion).

Assumptions A-C are the same as in Elliott, Jansson and Pesavento (2005). Assumption A states that the initial values are asymptotically negligible while Assumption B implies stationarity. Assumption C ensures that ε_t satisfies the Functional Central Limit Theorem (FCLT) so that

$$\frac{1}{\sqrt{T}} \sum_{s=1}^{[Tr]} \varepsilon_s \Rightarrow \Sigma^{1/2} W(r),$$

where $W(r)$ is a standard vector Brownian motion, $\Rightarrow$ signifies weak convergence and $[\cdot]$ denotes the greatest lesser integer function. Furthermore, Assumptions A-C imply that $v_t = A(L)^{-1} \varepsilon_t$ has the following limit

$$\frac{1}{\sqrt{T}} \sum_{s=1}^{[Tr]} v_s \Rightarrow \Omega^{1/2} W(r),$$

where $\Omega = A(1)^{-1} \Sigma A(1)^{-1\prime}$ is the spectral density at frequency zero of v_t scaled by 2π, $\Omega^{1/2} = \begin{pmatrix} \Omega_{11}^{1/2} & 0 \\ \omega_{21}\Omega_{11}^{-1/2} & \omega_{2.1}^{1/2} \end{pmatrix}$, $\omega_{2.1}^{1/2} = \omega_{22} - \omega_{21}\Omega_{11}^{-1}\omega_{12}$, and $W' = \begin{pmatrix} W_1' & W_2 \end{pmatrix}$ is a vector of independent standard Brownian motions, partitioned conformably to $v_{x,t}$ and $v_{y,t}$.

Next, define the scalar $\theta^2 = \delta'\delta$, where $\delta = \Omega_{11}^{-1/2}\omega_{12}\omega_{22}^{-1/2}$ denotes a vector containing the bivariate correlations at frequency zero of each element of $v_{x,t}$ with $v_{y,t}$. The scalar θ^2 represents the contribution of the right-hand variables in the second equation of (1) and it takes a value of zero when $v_{x,t}$ are not correlated in the long run with the errors from the cointegration regression.

Assumption D *Assume that* $0 \leq \theta^2 < 1$ *and* Ω_{11} *is non-singular.*

Assumption D restricts the squared long-run correlation θ^2 to be strictly less than one for technical reasons (see also Hansen, 1995). Also, the assumption that Ω_{11} is non-singular implies that elements of x_t are not individually cointegrated. Our next assumption follows Gospodinov

(2009) and reparameterizes ρ_T and τ_T as local-to-unity and local-to-zero sequences to account for the possibility of highly persistent errors and low signal-to-noise ratio.

Assumption E *Assume that* $\rho_T = 1 + c/T$ *for some fixed constant* $c \leq 0$, *and* $\tau_T = \lambda/\sqrt{T}$ *for some fixed constant* $\lambda > 0$.

The normalization factors T and $T^{1/2}$ for the local-to-unity and local-to-zero parameterizations are chosen to match the asymptotics of the estimators of ρ_T and τ_T. The local-to-zero parameterization has been used in a predictive regression framework by Torous and Valkanov (2000), Moon, Rubia and Valkanov (2004), Deng (2014) and Gourieroux and Jasiak (2020). In a different context, Ng and Perron (1997) adopt a similar parameterization to study the effect of low signal-to-noise ratio of the regressor on the sampling properties of cointegrating vector estimators.

Thus, under our assumptions, both y_t and x_t in (1) have a unit root but are cointegrated, and the cointegration error $y_t - \gamma' x_t$ is persistent – potentially persistent enough that we would not detect cointegration with standard tests in small samples. Yet, we have that asymptotically τ_T is approaching zero so that the cointegration error remains stochastically bounded even when its persistence parameter ρ_T, that drives its dynamics, is near or at unity.

As pointed out in the introduction, the dual localization is key for ensuring that the cointegration error $u_{y,t} = y_t - \gamma' x_t$ is stochastically bounded and hence consistent with both statistical and economic theory. Unlike regular near-unit root processes that are of order $O_p(T^{1/2})$, the local-to-zero variance localization dampens the stochastic trend behavior of $u_{y,t}$ and keeps it stochastically bounded ($O_p(1)$). More specifically, $u_{y,t}$ converges weakly to an Ornstein-Uhlenbeck process without any normalization that depends on the sample size:[4]

$$\begin{aligned} u_{y,t} &= \lambda T^{-1/2} \sum_{i=1}^{t} (1 + c/T)^{t-i} v_{y,i} \\ &\Rightarrow \omega_{2.1}^{1/2} \lambda J_{12c}(r), \end{aligned}$$

where $J_{12c}(r) = W_{12}(r) + c \int_0^r e^{(r-s)c} W_{12}(s) ds$ with $W_{12}(r) = \sqrt{\frac{\theta^2}{1-\theta^2}} \tilde{W}_1(r) + W_2(r)$, where $\tilde{W}_1(r)$ is an univariate standard Brownian motion independent of $W_2(r)$. The dual localization removes the unappealing possibility for some economic series (spot and forward prices, for example) that the errors $u_{y,t}$ can wander off and preserves the cointegration between y_t and x_t.

2.2 Limiting Distributions

We first consider the standard OLS estimator for the cointegration vector γ obtained from the regression of y_t on x_t. Theorem 1 below presents the limiting distribution of the estimator $\hat{\gamma}$.[5]

[4] The proof of this result is provided in the Appendix and follows closely Pesavento (2004).

[5] For ease of exposition, we follow the usual convention and suppress the (r) from the Brownian motion terms. All integrals are intended to be between 0 and 1, unless stated otherwise.

Theorem 1 *Under Assumptions A-E, and as $T \to \infty$,*

$$\sqrt{T}\left(\hat{\gamma}-\gamma_0\right) \Rightarrow \Omega_{11}^{-1/2\prime}\omega_{2.1}^{1/2}\lambda\left(\int W_1W_1'\right)^{-1}\left(\int W_1J_{12c}\right), \tag{2}$$

$$T\cdot \mathrm{SE}(\hat{\gamma}) \Rightarrow \left[\lambda^2\Omega_{11}^{-1/2}\left(\int W_1W_1'\right)^{-1}\Omega_{11}^{-1/2\prime}\omega_{2.1}\left(\int \tilde{J}_{12c}^2\right)\right]^{1/2}, \tag{3}$$

where $\tilde{J}_{12c} = J_{12c} - \left(\int W_1J_{12c}\right)'\left(\int W_1W_1'\right)^{-1}W_1$, $J_{12c}(r) = W_{12}(r) + c\int_0^r e^{(r-s)c}W_{12}(s)ds$, $W_{12}(r) = \sqrt{\frac{\theta^2}{1-\theta^2}}\tilde{W}_1(r) + W_2(r)$ *and* $\tilde{W}_1(r)$ *and* $W_2(r)$ *are independent univariate standard Brownian motions.*

Proof. See Appendix. ■

Interestingly, the estimator $\hat{\gamma}$ has an asymptotic distribution that resides in between the usual spurious and cointegration regressions. Unlike spurious regressions, $\hat{\gamma}$ is consistent but, in contrast to the usual cointegration regressions, it is not super-consistent as it has a slower ($\sqrt{T}$) rate of convergence. Additionally, while the estimator is consistent, the conventional t-statistic of $H_0 : \gamma = \gamma_0$ diverges at rate $T^{1/2}$ as in a spurious regression. This can be easily seen from the results in Theorem 1, i.e.

$$t_{\gamma=\gamma_0} = \frac{(\hat{\gamma}-\gamma_0)}{\mathrm{SE}(\hat{\gamma})} = \frac{\sqrt{T}\left(\hat{\gamma}-\gamma_0\right)}{T\cdot\mathrm{SE}(\hat{\gamma})}\sqrt{T} \to \pm\infty$$

as $T \to \infty$. Note also that an efficient estimator of γ can be obtained using a control variable approach (Phillips, 1991).

In what follows, we assume that γ is known which is the case in our empirical application. We briefly discuss the setup when γ is estimated after we present our main results in Theorems 2 and 3 below. Consider the VEC representation of the model given by

$$\begin{pmatrix} \Delta x_t \\ \Delta y_t \end{pmatrix} = (\rho - 1)\begin{pmatrix} 0_k & 0 \\ -\gamma' & 1 \end{pmatrix}\begin{pmatrix} x_{t-1} \\ y_{t-1} \end{pmatrix} + \begin{pmatrix} I_k & 0 \\ \gamma' & 1 \end{pmatrix}\begin{pmatrix} v_{x,t} \\ \tau_T v_{y,t} \end{pmatrix}. \tag{4}$$

Premultiplying by $A(L)$ and using $A(L) = I_{k+1} - \sum_{i=1}^p A_iL^i = I_{k+1} - \left(A_1 + A_2L + ... + A_pL^{p-1}\right)L = I_{k+1} - A^*(L)L$ and $A(L) = A(1) + (1-L)\bar{A}(L)$, where $\bar{A}(L)$ is another $(p-1)$-order lag polynomial, we obtain

$$\begin{pmatrix} \Delta x_t \\ \Delta y_t \end{pmatrix} = (\rho - 1)A(1)\begin{pmatrix} 0_k \\ u_{y,t-1} \end{pmatrix} + \tilde{A}(L)\begin{pmatrix} \Delta x_{t-1} \\ \Delta y_{t-1} \end{pmatrix} + A(L)\begin{pmatrix} v_{x,t} \\ \tau_T v_{y,t} + \gamma' v_{x,t} \end{pmatrix},$$

where $\tilde{A}(L) = A^*(L) + \bar{A}(L)(\rho-1)\begin{pmatrix} 0_k & 0 \\ -\gamma' & 1 \end{pmatrix}$.

Here, we restrict our analysis to a single equation error correction model by imposing that the first to the second-to-last elements of the last column of $A(1)$ are equal to zero (see our discussion after Theorem 2 below); i.e., $A(1) = \begin{pmatrix} A_{11}(1) & 0_k \\ A_{21}(1) & a_{22}(1) \end{pmatrix}$. In this case, the conditional ECM for Δy_t is given by

$$\Delta y_t = (\rho - 1)a_{22}(1)u_{y,t-1} + \tilde{A}_{21}(L)\Delta x_{t-1} + \tilde{a}_{22}(L)\Delta y_{t-1} + (A_{21}(L) + a_{22}(L)\gamma')v_{x,t} + a_{22}(L)\tau_T v_{y,t}$$

or, using that $\Delta x_t = v_{x,t}$ and $\tau_T a_{22}(L)v_{y,t} = \tau_T \varepsilon_{y,t} - A_{21}(L)\tau_T \Delta x_t$,

$$\Delta y_t = \beta u_{y,t-1} + \gamma\Delta x_t + \pi_1^*(L)'\Delta x_{t-1} + \pi_2^*(L)\Delta y_{t-1} + \gamma'\Delta x_t + \tau_T\varepsilon_{y,t},$$

where $\beta = (\rho-1)a_{22}(1)$ and $\pi_1^*(L)$ and $\pi_2^*(L)$ are lag polynomials of order $p-1$ that are functions of $A(L)$ and $\tilde{A}(L)$. Furthermore, define $\eta_t = \Sigma^{-1/2}\varepsilon_t$ such that $\varepsilon_{x,t} = \Sigma_{11}^{1/2}\eta_{x,t}$ and $\varepsilon_{y,t} = \sigma_{21}\Sigma_{11}^{-1/2}\eta_{xt} + \sigma_{2.1}^{1/2}\eta_{yt}$. Then, substituting $\eta_{x,t} = \Sigma_{11}^{-1/2}\varepsilon_{x,t}$ and noting that the terms in $\frac{\lambda}{\sqrt{T}}\sigma_{21}\Sigma_{11}^{-1/2}(\Delta x_t - \varepsilon_{x,t})$ can be expressed in terms of the lags of Δx_t and Δy_t, whose coefficients can be absorbed into $\pi_1^*(L)$ and $\pi_2^*(L)$, the conditional ECM for Δy_t takes the form

$$\Delta y_t = \beta u_{y,t-1} + \varphi\Delta x_t + \pi_1(L)'\Delta x_{t-1} + \pi_2(L)\Delta y_{t-1} + e_t, \tag{5}$$

where $\varphi = \gamma' + \tau_T\sigma_{21}\Sigma_{11}^{-1/2}$ and $e_t = \tau_T\sigma_{2.1}^{1/2}\eta_{yt}$.

Notice that given our assumptions, the stationary dependent variable in (5) is explained by a stationary regressor Δx_t, whose influence is not dominated by $u_{y,t-1}$, even when $u_{y,t-1}$ is persistent. Next, let $\tilde{\beta}$ be the OLS estimator in the conditional VECM (5), $\tilde{t}_{\beta=\beta_0}$ be the t-test of $H_0 : \beta = \beta_0$ and $\tilde{t}_{\beta=0}$ denote the t-test of the null hypothesis $H_0 : \beta = 0$ (or $\rho = 1$).

Theorem 2 *Suppose that Assumptions A-E hold. In addition, assume that $A_{12}(1) = 0_k$. Then, as $T \to \infty$,*

$$T\left(\tilde{\beta} - \beta_0\right) \Rightarrow \left(\int J_{12c}^2\right)^{-1}\left(\int J_{12c}dW_2\right), \tag{6}$$

$$\tilde{t}_{\beta=\beta_0} \Rightarrow \left(\int J_{12c}^2\right)^{-1/2}\left(\int J_{12c}dW_2\right), \tag{7}$$

$$\tilde{t}_{\beta=0} \Rightarrow \left(\int W_{12}^2\right)^{-1/2}\left(\int W_{12}dW_2\right), \tag{8}$$

where $J_{12c}(r)$ and $W_{12}(r)$ are defined in Theorem 1.

Proof. See Appendix. ■

Theorem 2 shows that the estimator in the conditional error-correction equation converges at rate T – i.e., it is super-consistent – and has a non-standard distribution which is an implicit function of the long-run correlation θ^2 and c. The asymptotic behavior of the estimator $\tilde{\beta}$ bears some similarities to the limiting representations derived in other contexts; see, for example, Hansen

(1995), Zivot (2000), and Pesavento (2004). As expected, it depends on the long-run correlation θ^2 and c. By controlling for Δx_t, we remove the noisy component of the error term and both $u_{y,t-1}$ and e_t have shrinking innovations variances, even though $u_{y,t-1}$ is allowed to be persistent (local to unity). Importantly, the limiting distributions of the estimator and the t-statistic do not depend on the signal-to-noise ratio through the localizing constant λ.

The assumption $A_{12}(1) = 0_k$ warrants some remarks. It is imposed in this model (see also Zivot, 2000) to ensure that $H_0 : \beta = 0$ can be interpreted as a test for cointegration in the conditional ECM (5) by ignoring the information contained in the marginal model for Δx_t. While this assumption simplifies the asymptotic representations for $T\left(\tilde{\beta} - \beta_0\right)$ and $\tilde{t}_{\beta=\beta_0}$ in Theorem 2, it should be stressed that the condition $A_{12}(1) = 0_k$ is not required for establishing the limiting behavior of the estimator and the t-test of $H_0 : \beta = \beta_0$ and the limiting expressions in (6) and (7) can be readily modified by relaxing this assumption. Of course, this assumption is automatically satisfied in the case of no seral correlation; i.e., $A(L) = I_{k+1}$. For further discussion on the trade-off between the weak exogeneity assumption in single-equation ECM and the system-based approach to testing for cointegration, see Elliott, Jansson and Pesavento (p. 36, 2005). These remarks also apply to the results in Theorem 3 below.

It is often the case that the VECM is defined (for predictive purposes, for instance) as

$$\Delta y_t = \beta u_{y,t-1} + \pi_1(L)'\Delta x_{t-1} + \pi_2\left(L\right)\Delta y_{t-1} + \xi_t, \tag{9}$$

where $\xi_t = \varphi\Delta x_t + e_t$.[6] We refer to model (9) as the unconditional VECM. Let $\hat{\beta}$ denote the OLS estimator in the unconditional VECM, $\hat{t}_{\beta=\beta_0}$ be the t-test of $H_0 : \beta = \beta_0$ based on the estimator $\hat{\beta}$ and $\hat{t}_{\beta=0}$ be the t-test of the null hypothesis $H_0 : \beta = 0$ (or $\rho = 1$). We then have the following result.

Theorem 3 *Suppose that Assumptions A-E hold. In addition, assume that* $A_{12}(1) = 0_k$ *and* $\gamma \neq 0$. *Then, as* $T \to \infty$,

$$\sqrt{T}\left(\hat{\beta} - \beta_0\right) \Rightarrow \lambda^{-1}\omega_{2.1}^{-1/2}(\gamma'\Omega_{11}\gamma)^{1/2}\left(\int J_{12c}^2\right)^{-1}\left(\int J_{12c}d\tilde{W}_1 + \Lambda^*\right), \tag{10}$$

$$\hat{t}_{\beta=\beta_0} \Rightarrow \frac{(\gamma'\Omega_{11}\gamma)^{1/2}}{(\gamma'\Gamma_{0,xx}\gamma)^{1/2}}\left(\int J_{12c}^2\right)^{-1/2}\left(\int J_{12c}d\tilde{W}_1 + \Lambda^*\right), \tag{11}$$

$$\hat{t}_{\beta=0} \Rightarrow \frac{(\gamma'\Omega_{11}\gamma)^{1/2}}{(\gamma'\Gamma_{0,xx}\gamma)^{1/2}}\left(\int W_{12c}^2\right)^{-1/2}\left(\int W_{12c}d\tilde{W}_1 + \Lambda^*\right), \tag{12}$$

[6]Strictly speaking, the parameters in model (9) should be denoted differently than those in the conditional ECM (5) since the unconditional model (9) is misspecified as it omits the term $\varphi\Delta x_t$. For notational simplicity, we do not index the parameters in the conditional and unconditional specifications in the theoretical part but we do so in the empirical application.

where $\Lambda^* = \omega_{21}^{-1/2}\lambda^{-1}(\gamma'\Omega_{11}\gamma)^{-1/2}\gamma'\Lambda_{y,x}$, $\Lambda_{y,x} = \sum_{h=1}^{\infty}\Gamma'_{h,yx}$,

$$\Gamma_h = \begin{pmatrix} E\left(v_{x,t-h}v'_{x,t}\right) & E\left(v_{x,t-h}v_{y,t}\right) \\ E\left(v_{y,t-h}v'_{x,t}\right) & E\left(v_{y,t-h}v_{y,t}\right) \end{pmatrix} = \begin{pmatrix} \Gamma_{h,xx} & \Gamma_{h,xy} \\ \Gamma_{h,yx} & \Gamma_{h,yy} \end{pmatrix},$$

$\tilde{W}_1$ *is an univariate Brownian motion independent of* W_2*, and* $J_{12c}(r)$ *and* $W_{12}(r)$ *are defined in Theorem 1.*

Proof. See Appendix. ■

Unlike the estimator $\tilde{\beta}$ in the conditional VECM, the estimator $\hat{\beta}$ has a slower (root-T) rate of convergence and its limiting distribution depends inversely on λ so that low values of λ make the estimator highly volatile. The asymptotic distributions of the estimator and its t-statistic are still non-standard. While they are also functionals of Brownian motions as for the estimator in the conditional VECM, there is a sharp contrast in the limiting behavior of these two estimators and their corresponding t-tests. In particular, because we are not conditioning on Δx_t, the errors in the unconditional error correction equation will be serially correlated and there will be extra parameters for the short- and long-run variances that will enter the asymptotic distribution. When there is no serial correlation , i.e. $A(L) = I_{k+1}$, the limit distribution of the t-statistics in the unconditional VECM are

$$\hat{t}_{\beta=\beta_0} \Rightarrow \left(\int J_{12c}^2\right)^{-1/2}\left(\int J_{12c}d\tilde{W}_1\right) \tag{13}$$

and

$$\hat{t}_{\beta=0} \Rightarrow \left(\int W_{12}^2\right)^{-1/2}\left(\int W_{12}d\tilde{W}_1\right). \tag{14}$$

For example, when $\theta^2 = 0$ (and $W_{12}(r) = W_2(r)$), the asymptotic distribution of the t-statistic for $H_0 : \beta = 0$ in the unconditional VECM reduces to the standard normal distribution while the limit of the t-test in the conditional VECM is characterized by the Dickey-Fuller distribution.

Table 1 about here

Critical values at the 5% significance level for the limiting distributions of the t-tests $\tilde{t}_{\beta=0}$ and $\hat{t}_{\beta=0}$ when there is no serial correlation ($p = 0$) are presented in Table 1. This is the setup of our empirical example and simulation experiment where these critical values are directly applicable. They are tabulated for different values of the scalar θ^2 that characterizes the degree of endogeneity in the model and determines implicit weights assigned to the standard normal and Dickey-Fuller distribution. The critical values for the case with no deterministic terms are obtained from the asymptotic representations in Theorems 2 and 3. For comparison purposes, we only use the case when there is no serial correlation and no nuisance parameters that need to be estimated, which is also the relevant case in our empirical application. For the cases with deterministic terms in (5) and (9) ("constant, no trend" and "constant and trend"), the standard Brownian motion is replaced by its demeaned and detrended analogs. The critical values are obtained by simulation using $200,000$ replications and $T = 30,000$.

Finally, while the results in Theorems 2 and 3 assume that the cointegration vector γ is known, the corresponding asymptotic representations can also be characterized when γ is estimated by OLS and $\hat{u}_{y,t-1}$ is used in the conditional and unconditional regressions. Despite the more complex form of these asymptotic distributions, the limiting behavior of the main quantities of interest is qualitatively unchanged: $\hat{u}_{y,t}$ is still stochastically bounded, $\tilde{\beta}$ is super-consistent and has an asymptotic distribution that does not depends on λ, and $\hat{\beta}$ continues to converge at a slower (root-T) rate with a limiting representation that depends on the signal-to-noise ratio.[7]

3 Simulation Results

To gain further understanding of the combined effect of low signal-to-noise ratio and persistent cointegration errors, and to quantify the cost of using the unconditional VECM in this context, we simulate data from a bivariate version of (1) with no serial correlation:

$$\begin{aligned} x_t &= u_{x,t} \\ y_t &= \gamma x_t + u_{y,t} \end{aligned} \tag{15}$$

and

$$\begin{pmatrix} (1-L)u_{x,t} \\ (1-\rho_T L)u_{y,t} \end{pmatrix} = \begin{pmatrix} \varepsilon_{x,t} \\ \tau_T \varepsilon_{y,t} \end{pmatrix}$$

with $\rho_T = 1 + c/T$ and $\tau_T = \lambda/\sqrt{T}$. Several values of λ, c, θ^2, and T are considered. In each case we estimate the conditional and unconditional VECM and compute the bias and standard error of $\tilde{\beta}$ and $\hat{\beta}$ together with the rejection rates for $\tilde{t}_{\beta=0}$ and $\hat{t}_{\beta=0}$ using $100,000$ Monte Carlo replications. We first choose a value for $\lambda = 0.05$ to match the implicit values for λ in the empirical application. The parameter θ^2 is set to 0, 0.3, or 0.7. In the bivariate model, θ^2 is the square of the correlation between $\varepsilon_{x,t}$ and $\varepsilon_{y,t}$, so that we can think of these values for θ^2 as corresponding to low, medium and high degrees of endogeneity. We set c to equal $0, -5$, and -10: values that correspond to ρ between 1 and 0.95 depending on the sample size. All reported tests have a nominal level of 5%.

Table 2 about here

[7] A sketch of this result is present in the proof for Theorem 1. The limiting results for the t-tests when γ is estimated follow directly and are available from the authors upon request.

When $c = 0$ ($\rho = 1$) and $\theta^2 = 0$, both the conditional and unconditional VEC estimators have negligible bias and correct size, although the standard error of the unconditional estimator can be large. As the degree of endogeneity increases, the size of both tests remains close to the nominal 5% level but the estimator from the unconditional regression, $\hat{\beta}$, has a large negative bias that increases as θ^2 increases.

For negative values of c ($\rho < 1$), we move away from the null hypothesis. In this case, we can see that both the bias and the standard error of $\hat{\beta}$ increase. Importantly, the probability of rejecting the null hypothesis in the conditional VECM correctly increases as ρ moves away from unity and as the degree of endogeneity increases. By contrast, the rejection probability of the t-test in the unconditional VECM remains around the 5% nominal level indicating a lack of power. As Theorem 3 shows, these asymptotic results are driven by the low signal-to-noise ratio: the asymptotic distribution of the unconditional VECM estimator $\hat{\beta}$ and, in particular, its variance[8] depend on λ, while the asymptotic distribution of the conditional VECM estimator $\tilde{\beta}$ is invariant to the value of λ. This suggests that the bias of the unconditional VECM estimator and the power of its test for significance will not improve even with large samples. This is in line with the theoretical results in Theorem 3. For example, the only difference that we observe between sample sizes of 200 and 400 is that, as T gets larger, the unconditional VECM estimates tend to have slightly smaller standard errors, although they are substantially larger than the standard errors for its conditional model counterpart.

As Theorem 3 highlights, the performance of the estimators from the unconditional VEC regression is inversely related to the localizing constant λ. A small value for λ makes the signal-to-noise negligible and induces high volatility in $\hat{\beta}$. To provide further evidence of the effect of λ on the behavior of the estimators and tests of significance, Table 3 presents results for a bigger value of λ ($\lambda = 10$) that renders the signal-to-noise ratio relatively large.

Table 3 about here

Table 3 reveals that when λ is larger, the estimator based on the unconditional regression performs better with the bias and standard error of $\hat{\beta}$ being much smaller than those for a small signal-to-noise ratio ($\lambda = 0.05$). However, the conditional VECM estimator continues to be more efficient which results in non-trivial power for its t-test. In contrast, while the power of the t-test in the unconditional VECM is improved compared to the case in Table 2, it is still dominated by the conditional VECM test.

[8]See the Appendix for details on the asymptotic variance of the estimators.

4 Empirical Application: Forward Premium Model

The main motivation for the theoretical analysis developed above has been some puzzling results in the forward premium regression models that link spot currency future returns to current forward premium, defined as the difference between the forward and spot exchange rates. The main properties of the data for the British pound (BP), German mark (DM), Swiss franc (SF) and Canadian dollar (CD) – all against the US dollar – are visualized in Figure 1 in the introduction. The left charts illustrate the extremely small signal-to-noise ratio of the forward premium regression while the right panel of charts presents the near-unit root dynamics of the spot-forward spread. These data features strongly suggest that the conventional forward premium regression attempts to explain a noisy but stationary dependent variable with a small, but persistent regressor. Consequently, the estimator is likely to exhibit non-standard finite-sample and asymptotic behavior.

The two regression specifications that we consider are based on the unconditional and conditional VECM, for a cointegrating vector $\gamma = (1, -1)'$, and take the following form

$$\Delta s_t = \alpha_U + \beta_U (s_{t-1} - f_{t-1}) + \xi_t, \tag{16}$$

$$\Delta s_t = \alpha_C + \beta_C (s_{t-1} - f_{t-1}) + \varphi_C \Delta f_t + u_t, \tag{17}$$

where s_t denotes the log spot exchange rate and f_t is its corresponding one-month log forward rate. The parameters are indexed by U and C to signify their association with the unconditional and conditional VECM, respectively. Equation (16) has been used extensively in the forward premium literature folowing Fama (1984),[9] but note that to be consistent with our setup and notation in the methodological section, the error-correction term is defined as $(s_{t-1} - f_{t-1})$ and not as $(f_{t-1} - s_{t-1})$ as in the forward premium literature. We should also stress that our setup and hypothesis of interest differ from the tests for forward rate unbiasedness in the forward premium puzzle (for the analysis of the forward premium puzzle, see Maynard and Phillips, 2001; Gospodinov, 2009; among others). Specifically, we test $\beta = \rho - 1 = 0$ whereas the unbiasedness hypothesis is a test of $\beta_U = 1$.

The data consist of monthly observations for the four exchange rates (BP, DM, SF, CD), mentioned above, for the period January 1975 – May 2006 and the Japanese yen (JY) for the period August 1978 – May 2006. The monthly spot rates are constructed by taking the observation on the last business day of each month (daily mid-market observation from Datastream). One-month forward rates are constructed from end-of-the month Eurocurrency rates for US, UK, Germany, Japan, Canada and Switzerland obtained from Datastream, using the covered interest parity.

[9] The vast majority of this literature does not include lags of Δs_t and Δf_t and we follow this tradition. We have also tried estimating the model with lagged differences included but found the lags to be insignificant (at 5% significance level) for BP, DM, SF and CD. Even in the JY regression where they are only borderline significant, the coefficients on the lagged Δs_t and Δf_t offset each other (with opposite signs and similar magnitude so that the sum of the coefficients is near zero). For this reason, we decided to maintain the specification of the forward premium regression that is commonly used in practice (with no lags of Δs_t and Δf_t).

Table 4 about here

Table 4 presents the regression estimates and their associated Newey-West standard errors (with 12 lags) from the model specifications (16) and (17), along with the R^2 from these regressions. But the table starts by reporting some salient features of the data that justify our dual parameterization and the system approach to estimation and inference. The ratio of volatilities between the regressor $(s_{t-1} - f_{t-1})$ and the dependent variable Δs_t in the unconditional VECM is very low (close to zero) with implied values of λ ranging between 0.026 and 0.052. At the same time, the spot-forward spread $(s_{t-1} - f_{t-1})$ is a highly persistent process with AR(1) coefficients near one. The unit root test (augmented Dickey-Fuller test for a model with a drift and 12 lags) cannot reject the null of a unit root at 5% significance level. This leads to the counter-intuitive conclusion that spot and forward rates drift apart in the long-run despite being visually indistinguishable from each other in Figure 1. Overall, the combination of these two data characteristics (low signal-to-noise ratio and high persistence) highlights the importance of using limiting distributions that are explicit functions of these parameters.

As our theory and simulation results suggest, the estimate of β_U appears to be highly volatile (large standard error) and thus has a substantial probability of being far away from its implied value under the null. Given the very low signal-to-noise ratio of this regression model, the explanatory power of $(s_{t-1} - f_{t-1})$ is very low which is reflected in values of R^2 ranging between 0.005 and 0.045 for the different currencies. As Theorem 3 highlights, the precision of these estimates is inversely related to the localizing constant λ whose proximity to zero makes the signal-to-noise ratio negligibly small and induces high volatility in the OLS estimate of β_U. Since the degree of endogeneity in this model (measured by the long-run correlation θ^2) is somewhat low, the large downward biases reported in the simulations (for large θ^2) are not expected to be an issue here. Instead, the sampling behavior of the unconditional VECM estimator of β_U is driven by the negligible signal-to-noise ratio and the incompatibility between the dependent and independent variables in terms of both their scale and persistence.

By contrast, the estimates of β_C in model specification (17) are in line with those predicted by theory (one minus the implied value of the persistence parameter ρ) with significantly reduced variability. The meaningful improvement in the sampling properties of the estimator in (17) arises from "balancing" the stationary, but noisy, dependent variable with the inclusion of an additional regressor, Δf_{t+1}, with similar scale and persistence.[10] This recalibrates the signal and noise components

[10] The traditional definition of an unbalanced regression is that the regressand and the regressor are of different orders of integration. In our context, there can also be imbalance between the innovation variances, when one is fixed and the other shrinking, or between the overall scale or magnitude of the regressor and regressand, which depends jointly on both the integration order and innovation variance. Given these multiple notions of balance, we avoid the generic use of the term "balanced/unbalanced" regression and instead clarify the notion of "balance/unbalance"

to put them on a more equal footing.

The OLS estimates of β_C vary between 0.036 (Swiss franc) and 0.141 (Japanese yen) that map well within the spectrum of plausible values for the implied persistence of the forward premium. Using the critical values in Table 1 for values of θ^2 close to zero, these results lend support to the alternative hypothesis that the estimate of β_C is statistically significant (and smaller than 0); i.e., the persistent parameter of the forward premium is close to but strictly less than unity.

The large values of R^2 for this model provide additional evidence that all relevant information about Δs_t is reflected in the regressors $(s_{t-1} - f_{t-1})$ and Δf_t. It also reflects the good fit of our theoretical model to the exchange rate data. The high R^2 results from the strong correlation between Δs_t and Δf_t. This, in turn, is an implication of our modelling framework. Specifically, it is implied by the near-zero variance for $s_t - f_t$. The small variance of $s_t - f_t = s_{t-1} - f_{t-1} + (\Delta s_t - \Delta f_t)$ requires that Δs_t and Δf_t are typically close in value (i.e., highly correlated).

In summary, the magnitudes of the estimates from model (16) and their tests for significance should be interpreted with caution given their highly volatile behavior arising from the low signal-to-noise ratio of the regressor. On the other hand, model (17) is statistically balanced and its estimates and statistics are characterized by more appealing sampling properties.

5 Concluding Remarks

In this paper we proposed and studied a model of a nonstationary levels relationship in which the residual follows a local-unity process with a shrinking innovation variance. This setup captures empirical applications, such as spot and forward exchange rate and commodity prices, where the levels relationship appears tight despite a persistent, yet small residual. The asymptotic theory that we develop in the paper offers some interesting insights. The limiting behavior of the levels regression lies in between the cointegrating and spurious regression cases. The estimated coefficients remain consistent, but not super-consistent, and their corresponding t-tests diverge with the sample size.

We also analyzed the vector error-correction specifications of this model. Unfortunately, the unconditional VEC model is characterized by an imbalance between a small but persistent error correction term and a large stationary component in its error term. This imbalance is reflected in a low signal to noise ratio, resulting in highly variable coefficient estimates. Conversely, the conditional VEC specification addresses this imbalance by explicitly controlling for the high variance component of the residual in the unconditional VECM. This is manifested in a higher signal to noise ratio and a super-consistent error-correction coefficient estimate. The asymptotic distribution is

intended at each point in which we use the term.

non-standard, but the t-test depends on only a single endogeneity term, which can be consistently estimated and used to adjust the critical value.

Our simulations confirmed the superiority of the conditional VEC specification. While the unconditional and conditional VEC models perform similarly in the standard cointegration setting, the relative performance of the unconditional VECM strongly deteriorates when error variance of the levels residual is small and persistent. By contrast, the conditional VECM continues to exhibit excellent size and power properties.

We illustrated the practical relevance of our theoretical results in the context of spot-forward exchange rate regressions. Our analytical framework rationalizes the otherwise conflicting observations that (i) the spot and forward rates move closely together and (ii) their difference, the forward premium, is highly persistent. Common spot return forward premium regressions correspond to the unconditional VEC model. As predicted by the theory, this regression is imbalanced both in terms of its persistence and in terms of the magnitude of its innovation variance. Not surprisingly, the resulting estimates are imprecise with large standard errors. By contrast, the conditional VEC model produces more precise estimates with tighter standard estimates. Using the conditional VECM, we can reject the hypothesis of an exact unit root in the spot-forward spread, providing additional support for the tight levels relationship observed between spot and forward exchange rates.

Appendix: Proofs of Main Results

A.1 Preliminary Lemma

Lemma A1. *Under Assumptions A-E, we have that*

1. $\Omega_{11}^{-1/2}\frac{1}{T^2}\sum_{t=1}^{T}x_t x_t'\Omega_{11}^{-1/2\prime} \Rightarrow \int W_1 W_1'$,

2. $\Omega_{11}^{-1/2}\omega_{2.1}^{-1/2}\frac{1}{T^{\frac{3}{2}}}\sum_{t=1}^{T}x_t u_{y,t} \Rightarrow \lambda \int W_1 J_{12c}$,

3. $\omega_{2.1}^{-1}\frac{1}{T}\sum_{t=1}^{T}u_{y,t}^2 \Rightarrow \lambda^2 \int J_{12c}^2$,

4. $\omega_{2.1}^{-1/2}\frac{1}{T^{\frac{1}{2}}}\sum_{t=1}^{T}u_{y,t-1}(\Sigma^{-1/2}\varepsilon_t) \Rightarrow \lambda \int J_{12c}dW$,

5. $\Omega_{11}^{-1/2}\frac{1}{T}\sum_{t=1}^{T}x_{t-1}(\Sigma^{-1/2}\varepsilon_t) \Rightarrow \int W_1 dW$,

where $J_{12c}(r) = W_{12}(r) + c\int_0^r e^{(r-s)c}W_{12}(s)ds$, $W_{12}(r) = \sqrt{\frac{\theta^2}{1-\theta^2}}\tilde{W}_1(r) + W_2(r)$ *and* $\tilde{W}_1(r)$ *is a univariate standard Brownian motion.*

Proof. Under our assumptions, we have $\frac{1}{\sqrt{T}}\sum_{s=1}^{[Tr]}v_s \Rightarrow \Omega^{1/2}W(r)$ which implies that

$$\frac{1}{\sqrt{T}}\sum_{t=1}^{[Tr]}v_{x,t} \Rightarrow \Omega_{11}^{1/2}W_1(r)$$

and

$$\frac{1}{\sqrt{T}}\sum_{t=1}^{[Tr]}v_{y,t} \Rightarrow \omega_{21}\Omega_{11}^{-1/2}W_1(r) + \omega_{2.1}^{1/2}W_2(r).$$

Recall that $\theta^2 = \delta'\delta$, where $\delta = \Omega_{11}^{-1/2}\omega_{12}\omega_{22}^{-1/2}$ is a vector containing the bivariate zero frequency correlations of each element of $v_{x,t}$ with $v_{y,t}$. Define $\overline{\delta}' = \omega_{2.1}^{-1/2}\omega_{21}\Omega_{11}^{-1/2}$ so that $\overline{\delta}'\overline{\delta} = \frac{\theta^2}{1-\theta^2}$. We can then see that

$$\frac{1}{\sqrt{T}}\sum_{t=1}^{[Tr]}\overline{\delta}'\Omega_{11}^{-1/2}v_{x,t} \Rightarrow \sqrt{\frac{\theta^2}{1-\theta^2}}\tilde{W}_1(r),$$

where $\tilde{W}_1$ is an univariate standard Brownian motion independent of W_2 and

$$\frac{1}{\sqrt{Tr}}\sum_{t=1}^{[Tr]}v_{y,t} \Rightarrow \omega_{2.1}^{1/2}\left[\omega_{2.1}^{-1/2}\omega_{21}\Omega_{11}^{-1/2}W_1(r) + W_2(r)\right] = \omega_{2.1}^{1/2}\left[\sqrt{\frac{\theta^2}{1-\theta^2}}\tilde{W}_1(r) + W_2(r)\right].$$

Using these limiting expressions, all results in Lemma A1 follow from FCLT and Continuous Mapping Theorem. ■

A.2 Proof of Theorem 1

The results follow directly from Lemma A1 and the fact that

$$\sqrt{T}\left(\hat{\gamma}-\gamma\right)=\left(\frac{1}{T^{2}}\sum_{t=1}^{T}x_{t}x_{t}'\right)^{-1}\left(\frac{1}{T^{3/2}}\sum_{t=1}^{T}x_{t}u_{y,t}\right)$$

and

$$T\cdot\mathrm{SE}(\hat{\gamma})=\left[\left(\frac{1}{T^{2}}\sum_{t=1}^{T}x_{t}x_{t}'\right)^{-1}\left(\frac{1}{T}\sum_{t=1}^{T}\hat{u}_{y,t}^{2}\right)\right]^{1/2},$$

where $\hat{u}_{y,t}=u_{y\,t}-\left(\hat{\gamma}-\gamma\right)'x_{t}$. Notice that

$$\omega_{2.1}^{-1/2}\hat{u}_{y,t}=\omega_{2.1}^{-1/2}u_{y\,t}-\omega_{2.1}^{-1/2}\sqrt{T}\left(\hat{\gamma}-\gamma\right)\frac{x_{t}}{\sqrt{T}}$$

and $\hat{u}_{y,t}$ is also $O_p(1)$, so that

$$\begin{aligned}\omega_{2.1}^{-1/2}\hat{u}_{y,t} &\Rightarrow \lambda J_{12c}-\omega_{2.1}^{-1/2}\lambda\left(\int W_{1}J_{12c}\right)'\left(\int W_{1}W_{1}'\right)^{-1}\omega_{2.1}^{1/2}\Omega_{11}^{-1/2}\Omega_{11}^{1/2}W_{1}= \\ &= \lambda J_{12c}-\lambda\left(\int W_{1}J_{12c}\right)'\left(\int W_{1}W_{1}'\right)^{-1}W_{1}=\lambda\tilde{J}_{12c}^{2}.\end{aligned}$$

Therefore,

$$\omega_{2.1}^{-1}\left(\frac{1}{T}\sum_{t=1}^{T}\hat{u}_{y,t}^{2}\right)\Rightarrow\lambda^{2}\int\tilde{J}_{12c}^{2}.$$

A.3 Proof of Theorem 2

To simplify the intuition of the proof we assume no deterministic terms. Recall our conditional VECM (5)

$$\Delta y_{t}=\beta u_{y,t-1}+\varphi\Delta x_{t}+\pi_{1}(L)\Delta x_{t-1}+\pi_{2}\left(L\right)\Delta y_{t-1}+e_{t}$$

This can be written in a compact form as

$$\Delta y_{t}=X_{t}'\Pi+e_{t},$$

where

$$\Pi=\left(\begin{array}{cccccccc}\beta & \varphi & \pi_{11} & \cdots & \pi_{1p} & \pi_{21} & \cdots & \pi_{2p}\end{array}\right)',$$

$$X_{t}'=\left(\begin{array}{cccccccc}u_{y,t-1} & \Delta x_{t} & \Delta x_{t-1} & \cdots & \Delta x_{t-p} & \Delta y_{t-1} & \cdots & \Delta y_{t-p}\end{array}\right),$$

with $\beta=a_{22}(1)\left(\rho-1\right)$, $\varphi=\gamma'+\frac{\lambda}{\sqrt{T}}\sigma_{21}\Sigma_{11}^{-1/2}$ and $e_{t}=\frac{\lambda}{\sqrt{T}}\sigma_{2.1}^{1/2}\eta_{yt}$. Defining, $\hat{\Pi}$ as the OLS estimator of Π, we have[11]

$$T\left(\hat{\Pi}-\Pi\right)=\left(\frac{1}{T}\sum X_{t}'X_{t}\right)^{-1}\left(\sum X_{t}'e_{t}\right).$$

[11] To simplify the notation, we will assume that the summations are over all the available data which will depend on the lag length.

Then, invoking the limiting results in Lemma A1, we have

$$\frac{1}{T}\sum u_{y,t-1}^2 \Rightarrow \omega_{2.1}\lambda^2 \int J_{12c}^2(r),$$

$$\begin{aligned}\sum u_{y,t-1}e_t &= \frac{\lambda}{\sqrt{T}}\sigma_{2.1}^{1/2}\sum u_{y,t-1}\eta_{yt} \\ &\Rightarrow \omega_{2.1}^{1/2}\sigma_{2.1}^{1/2}\lambda^2\int J_{12c}dW_2,\end{aligned}$$

$$\frac{1}{T}\sum u_{y,t-1}\Delta x_{t-i} \to 0$$

for $i = 0, 1, ..., p-1$ since $\frac{1}{\sqrt{T}}\sum u_{y,t-1}\Delta x_{t-1}$ is $O_p(1)$, and

$$\frac{1}{T}\sum u_{y,t-1}\Delta y_{t-i} = \gamma'\frac{1}{T}\sum u_{y,t-1}\Delta y_{t-i} + c\frac{1}{T^2}\sum u_{y,t-1}^2 + \frac{\lambda}{T^{3/2}}\sum u_{y,t-1}v_{yt-i} \to 0,$$

where $\to$ denotes convergence in probability. The result in (6) follows directly from the asymptotic block diagonality of $\left(\frac{1}{T}\sum X_t'X_t\right)^{-1}$. Furthermore, note that $\sum \Delta x_{t-i}e_t = \lambda\sigma_{2.1}^{1/2}\frac{1}{\sqrt{T}}\sum \Delta x_{t-i}\eta_{y,t}$ and $\sum \Delta y_{t-i}e_t = \lambda\sigma_{2.1}^{1/2}\frac{1}{\sqrt{T}}\sum \Delta y_{t-i}\eta_{y,t}$ are $O_p(1)$ so that $T(\tilde{\varphi}-\varphi)$, $T(\hat{\pi}_{2i}-\pi_{2i})$ and $T(\hat{\pi}_{1i}-\pi_{1i})$ will also be $O_p(1)$.

For the variance of the estimator $\tilde{\beta}$

$$T^2 \cdot \mathrm{Var}\left(\tilde{\beta}\right) = \left(\frac{1}{T^2}\sum u_{y,t-1}^2\right)^{-1}\left(\sum \hat{e}_t^2\right),$$

we have from Lemma A1 that

$$\frac{1}{T}\sum_{t=1}^{[T\cdot]} u_{y,t}^2 \Rightarrow \lambda^2\omega_{2.1}\int J_{12c}^2.$$

We can write

$$\hat{e}_t = \Delta y_t - X_t'\hat{\Pi} = e_t - \frac{X_t'}{T}T(\hat{\Pi}-\Pi).$$

Since $T(\hat{\Pi}-\Pi)$ are $O_p(1)$, $\hat{e}_t$ will converge in the limit to $e_t = \frac{\lambda}{\sqrt{T}}\sigma_{2.1}^{1/2}\eta_{yt}$ and

$$\sum \hat{e}_t^2 \to \lambda^2\sigma_{2.1}$$

since $\frac{1}{T}\sum \eta_{y,t}^2 \to 1$. Therefore,

$$T \cdot \mathrm{SE}\left(\tilde{\beta}\right) \Rightarrow \omega_{2.1}^{-1/2}\sigma_{2.1}^{1/2}\left(\int J_{12c}^2\right)^{-1/2}.$$

Finally, using that $\beta = \rho - 1 = \frac{c}{T}$, the t-test for $H_0 : \beta = \beta_0$ has the following limiting representation

$$\tilde{t}_{\beta=\beta_0} = \frac{T\left(\tilde{\beta}-\beta_0\right)}{T\cdot \mathrm{SE}(\tilde{\beta})} \Rightarrow \frac{\left(\int J_{12c}^2\right)^{-1}\left(\int J_{12c}dW_2\right)}{\left(\int J_{12c}^2\right)^{-1/2}}.$$

For the test of the hypothesis $H_0 : \beta = 0$, we have $c = 0$ and $J_{12c} = W_{12}$ in the above limiting expression.

A.4 Proof of Theorem 3

The unconditional VECM is given by

$$\Delta y_t = \beta u_{y,t-1} + \pi_1(L)\Delta x_{t-1} + \pi_2(L)\Delta y_{t-1} + \xi_t.$$

This can be written in a more compact form as $\Delta y_t = X_t'\Pi + \xi_t$, where the notation matches that of the proof of Theorem 2, except that

$$X_t' = \begin{pmatrix} u_{y,t-1} & \Delta x_{t-1} & \cdots & \Delta x_{t-p} & \Delta y_{t-1} & \cdots & \Delta y_{t-p} \end{pmatrix}$$

is redefined to omit Δx_t, $\Pi = \begin{pmatrix} \beta & \pi_{11} & \cdots & \pi_{1p} & \pi_{21} & \cdots & \pi_{2p} \end{pmatrix}'$, and the new error $\xi_t = \varphi\Delta x_t + e_t$ thus includes the omitted $\varphi\Delta x_t$. Substituting for $\varphi = \gamma' + \frac{\lambda}{\sqrt{T}}\sigma_{21}\Sigma_{11}^{-1/2}$ and $e_t = \frac{\lambda}{\sqrt{T}}\sigma_{2.1}^{1/2}\eta_{yt}$, we have

$$\xi_t = \left(\gamma' + \frac{\lambda}{\sqrt{T}}\sigma_{21}\Sigma_{11}^{-1/2}\right)\Delta x_t + \frac{\lambda}{\sqrt{T}}\sigma_{2.1}^{1/2}\eta_{yt}.$$

The OLS estimator now converges at rate $\sqrt{T}$ since

$$T\left(\hat{\Pi} - \Pi\right) = \left(\frac{1}{T}\sum X_t'X_t\right)^{-1}\left(\frac{1}{\sqrt{T}}\sum X_t'\xi_t\right)$$

where

$$\frac{1}{\sqrt{T}}\sum u_{y,t-1}\xi_t = \frac{1}{\sqrt{T}}\sum u_{y,t-1}\gamma'\Delta x_t + \lambda\sigma_{21}\Sigma_{11}^{-1/2}\frac{1}{T}\sum u_{y,t-1}\Delta x_t + \sigma_{2.1}^{1/2}\lambda\frac{1}{T}\sum u_{y,t-1}\eta_{yt}.$$

The last two terms converge to zero while

$$\begin{aligned}\frac{1}{\sqrt{T}}\sum u_{y,t-1}\gamma'\Delta x_t = \frac{1}{\sqrt{T}}\sum u_{y,t-1}\gamma' v_{xt} \quad &\Rightarrow \quad \omega_{2.1}^{1/2}\lambda(\gamma'\Omega_{11}\gamma)^{1/2}\int J_{12c}d\widetilde{W}_1 + \gamma'\Lambda_{y,x} \\ &= \quad \omega_{2.1}^{1/2}\lambda(\gamma'\Omega_{11}\gamma)^{1/2}\left(\int J_{12c}d\widetilde{W}_1 + \Lambda^*\right).\end{aligned}$$

where $\Lambda_{y,x}$ and Λ^* are defined in Theorem 3. Noting that $\left(\frac{1}{T}\sum X_t'X_t\right)^{-1}$ is again asymptotically block diagonal and $\frac{1}{T}\sum u_{y,t-1}^2 \Rightarrow \omega_{2\cdot 1}\lambda^2\int J_{12c}^2(r)$, we have

$$\begin{aligned}\sqrt{T}\left(\hat{\beta} - \beta_0\right) \quad &= \quad \left(\frac{1}{T}\sum u_{y,t-1}^2\right)^{-1}\left(\frac{1}{\sqrt{T}}\sum u_{y,t-1}\xi_t\right) \\ &\Rightarrow \quad \lambda^{-1}\omega_{2.1}^{-1/2}(\gamma'\Omega_{11}\gamma)^{1/2}\left(\int J_{12c}^2\right)^{-1}\left(\int J_{12c}d\widetilde{W}_1 + \Lambda^*\right)\end{aligned}$$

For the variance of the estimator, we proceed similarly as in the proof in Theorem 2. Since

$$\hat{\xi}_t = \Delta y_t - X_t'\hat{\Pi} = \xi_t - \frac{X_t'}{\sqrt{T}}\sqrt{T}\left(\hat{\Pi} - \Pi\right),$$

where $\xi_t = \left(\gamma' + \frac{\lambda}{\sqrt{T}}\sigma_{21}\Sigma_{11}^{-1/2}\right)\Delta x_t + \frac{\lambda}{\sqrt{T}}\sigma_{2.1}^{1/2}\eta_{yt} = \gamma'\Delta x_t + o_p(1)$, we have

$$\frac{1}{T}\hat{\xi}'\hat{\xi} = \frac{1}{T}\xi'\xi + o_p(1) = \frac{1}{T}\sum \gamma' v_{x,t}v_{x,t}'\gamma + o_p(1) \to \gamma'\Gamma_{0,xx}\gamma.$$

The variance of the estimator can then be expressed as

$$T \cdot \text{Var}\left(\hat{\beta}\right) = \left(\frac{1}{T^2}\sum u_{y,t-1}^2\right)^{-1}\left(\frac{1}{T}\sum \hat{\xi}_t^2\right) \Rightarrow \gamma'\mathit{\Gamma}_{0,x}\gamma'\left(\omega_{2.1}\lambda^2\int J_{12c}^2\right)^{-1}.$$

Thus,

$$\hat{t}_{\beta=\beta_o} = \frac{\sqrt{T}\left(\hat{\beta}-\beta_0\right)}{\sqrt{T}\cdot \text{SE}(\hat{\beta})} \Rightarrow \frac{(\gamma'\Omega_{11}\gamma)^{1/2}}{(\gamma'\Gamma_{0,x}\gamma)^{1/2}}\left(\int J_{12c}^2\right)^{-1/2}\left(\int J_{12c}d\widetilde{W}_1 + \Lambda^*\right).$$

As before, when $\beta = 0,\ c = 0$ and $J_{12c} = W_{12}$.

References

[1] Bansal, R. and A. Yaron (2004), "Risks for the long run: a potential resolution of asset pricing puzzles", *The Journal of Finance* 59, 1447–1950

[2] Boswijk, H. P. (1994), "Testing for an unstable root in conditional and structural error correction models," *Journal of Econometrics* 63, 37–60.

[3] Deng, A. (2014), "Understanding spurious regression in financial economics," *Journal of Financial Econometrics* 12, 122–150.

[4] Elliott, G, M. Jansson, and E. Pesavento (2005), "Optimal power for testing potential cointegrating vectors with known parameters for nonstationarity," *Journal of Business and Economic Statistics* 23, 34–48.

[5] Engle, R. F., and C. W. J. Granger (1987), "Co-integration and error-correction: Representation, estimation and testing," *Econometrica* 55, 251–276.

[6] Fama, E. (1984), "Forward and spot exchange rates," *Journal of Monetary Economics* 14, 319–338.

[7] Gourieroux, C. and J. Jasiak (2020), "Inference for noisy long run component process", Unpublished manuscript, York University.

[8] Gospodinov, N. (2009), "A new look at the forward premium puzzle," *Journal of Financial Econometrics* 7, 312–338.

[9] Hansen, B. E. (1995), "Rethinking the univariate approach to unit root tests: How to use covariates to increase power," *Econometric Theory* 11, 1148–1171.

[10] Jansson, M., and N. Haldrup (2002), "Regression theory for nearly cointegrated time series," *Econometric Theory* 18, 1309–1335.

[11] Johansen, S. (1992), "Cointegration in partial systems and the efficiency of single-equation analysis," *Journal of Econometrics* 52, 389–402.

[12] Maynard, A., and P. C. B. Phillips (2001), "Rethinking an old empirical puzzle: Econometric evidence on the forward discount anomaly," *Journal of Applied Econometrics* 16, 671–708.

[13] Moon, R., A. Rubia, and R. Valkanov (2004), "Long-horizon regressions when the predictor is slowly varying," Unpublished manuscript, University of Southern California.

[14] Müller, U. K., and M. W. Watson (2008), "Testing models of low-frequency variability," *Econometrica* 76, 979–1016.

[15] Müller, U. K., and M. W. Watson (2018), "Long-run covariability," *Econometrica* 86, 775–804.

[16] Ng, S., and P. Perron (1997), "Estimation and inference in nearly unbalanced nearly cointegrated systems," *Journal of Econometrics* 79, 53–81.

[17] Park, J. Y. (1992), "Canonical cointegrating regressions" *Econometrica* 60, 119–143.

[18] Park, J. Y., and P. C. B. Phillips (1989), "Statistical inference in regressions with integrated processes: Part 2," *Econometric Theory* 5, 95–131.

[19] Pesavento, E. (2004), "Analytical evaluation of the power of tests for the absence of cointegration, " *Journal of Econometrics* 122, 349–384.

[20] Phillips, P. C. B. (1991), "Optimal inference in cointegrated systems," *Econometrica* 59, 283–306.

[21] Phillips, P. C. B., and J. H. Lee (2013), "Predictive regression under various degrees of persistence and robust long-horizon regression," *Journal of Econometrics* 177, 250–264.

[22] Stock, J. H., and M. W. Watson (1993), "A simple estimator of cointegrating vectors in higher order integrated systems," *Econometrica* 61, 783–820.

[23] Torous, W., and R. Valkanov (2000), "Boundaries of predictability: Noisy predictive regressions," Unpublished manuscript, UCLA.

[24] Zivot, E. (2000), "The power of single equation tests for cointegration when the cointegrating vector is prespecified," *Econometric Theory* 16, 407–439.

Table 1. Asymptotic critical values for $\tilde{t}_{\beta=0}$ (conditional VECM) and $\hat{t}_{\beta=0}$ (unconditional VECM) at 5% significance level.

θ^2	$\tilde{t}_{\beta=0}$	$\hat{t}_{\beta=0}$	$\tilde{t}_{\beta=0}$	$\hat{t}_{\beta=0}$	$\tilde{t}_{\beta=0}$	$\hat{t}_{\beta=0}$
	no determ. terms		constant, no trend		constant and trend	
0	-1.941	-1.645	-2.863	-1.645	-3.413	-1.645
0.2	-1.939	-1.819	-2.775	-2.278	-3.274	-2.544
0.3	-1.927	-1.857	-2.721	-2.403	-3.192	-2.725
0.5	-1.900	-1.902	-2.584	-2.584	-2.995	-2.995
0.7	-1.853	-1.921	-2.398	-2.721	-2.730	-3.192
0.8	-1.822	-1.932	-2.274	-2.778	-2.548	-3.280
0.9	-1.173	-1.938	-2.098	-2.826	-2.301	-3.351

Notes: Critical values are computed by simulating the asymptotic distributions with $200,000$ replications, $T = 30,000$ and $p = 0$. For the case of no deterministic terms ("no determ. terms"), the critical values are obtained from the limiting distributions in Theorems 2 and 3. For the other two cases, the standard Brownian motion in the limiting distributions is replaced by its demeaned and detrended analogs.

Table 2. Simulation results for $\lambda = 0.05$, and various values of ρ, θ^2 and T.

	$\theta^2 = 0$			$\theta^2 = 0.3$			$\theta^2 = 0.7$		
	bias	s.d.	t-test	bias	s.d.	t-test	bias	s.d.	t-test
$c = 0$									
$T = 200$ ($\rho = 1$)									
CVECM	-0.007	0.008	0.052	-0.005	0.007	0.053	-0.002	0.004	0.050
UVECM	0.003	2.276	0.052	-1.035	2.279	0.051	-1.583	2.281	0.051
$T = 400$ ($\rho = 1$)									
CVECM	-0.004	0.004	0.052	-0.003	0.004	0.053	0.001	0.002	0.051
UVECM	0.000	1.771	0.052	-0.806	1.774	0.051	-1.240	1.774	0.050
$c = -5$									
$T = 200$ ($\rho = 0.97$)									
CVECM	-0.009	0.018	0.368	-0.006	0.015	0.566	-0.003	0.010	0.880
UVECM	-0.05	5.002	0.053	-1.396	5.011	0.044	-2.136	5.015	0.041
$T = 400$ ($\rho = 0.99$)									
CVECM	-0.004	0.009	0.364	-0.003	0.007	0.565	-0.001	0.005	0.880
UVECM	0.014	3.554	0.052	-0.981	3.557	0.043	-1.508	3.561	0.040
$c = -10$									
$T = 200$ ($\rho = 0.95$)									
CVECM	-0.009	0.024	0.793	-0.006	0.020	0.918	-0.003	0.013	0.994
UVECM	-0.011	6.671	0.052	-1.422	6.681	0.041	-2.174	6.684	0.039
$T = 400$ ($\rho = 0.97$)									
CVECM	-0.005	0.012	0.792	-0.003	0.010	0.918	-0.001	0.007	0.994
UVECM	0.020	4.742	0.053	-1.009	4.746	0.041	-1.556	4.751	0.037

Notes: The table presents the average bias and standard deviations of the β estimates as well as the rejection probabilities of the t-test for $H_0 : \beta = 0$. The results are based on 30,000 Monte Carlo replications.

Table 3. Simulation results for $\lambda = 10$, and various values of ρ, θ^2 and T.

	$\theta^2 = 0$			$\theta^2 = 0.3$			$\theta^2 = 0.7$		
	bias	s.d.	t-test	bias	s.d.	t-test	bias	s.d.	t-test
$c = 0$									
$T = 200$ ($\rho = 1$)									
CVECM	-0.007	0.008	0.052	-0.005	0.007	0.053	-0.002	0.004	0.050
UVECM	-0.007	0.014	0.079	-0.012	0.017	0.061	-0.015	0.019	0.054
$T = 400$ ($\rho = 1$)									
CVECM	-0.004	0.004	0.052	-0.003	0.004	0.053	-0.001	0.002	0.051
UVECM	-0.004	0.010	0.073	-0.008	0.012	0.059	-0.010	0.013	0.052
$c = -5$									
$T = 200$ ($\rho = 0.97$)									
CVECM	-0.009	0.018	0.368	-0.006	0.015	0.566	-0.003	0.010	0.880
UVECM	-0.009	0.031	0.248	-0.016	0.038	0.139	-0.020	0.041	0.109
$T = 400$ ($\rho = 0.99$)									
CVECM	-0.004	0.009	0.364	-0.003	0.007	0.565	-0.001	0.005	0.880
UVECM	-0.004	0.020	0.182	-0.009	0.024	0.110	-0.012	0.026	0.088
$c = -10$									
$T = 200$ ($\rho = 0.95$)									
CVECM	-0.009	0.024	0.793	-0.006	0.020	0.918	-0.003	0.013	0.994
UVECM	-0.009	0.041	0.390	-0.016	0.050	0.219	-0.020	0.055	0.169
$T = 400$ ($\rho = 0.97$)									
CVECM	-0.005	0.012	0.792	-0.003	0.010	0.918	-0.001	0.007	0.994
UVECM	-0.004	0.027	0.269	-0.010	0.032	0.158	-0.012	0.034	0.125

Notes: The table presents the average bias and standard deviations of the β estimates as well as the rejection probabilities of the t-test for $H_0 : \beta = 0$. The results are based on 30,000 Monte Carlo replications.

Table 4. Estimation results for unconditional and conditional VEC models (16) and (17).

	BP	DM	SF	CD	JY
dual param.					
λ	0.0402	0.0444	0.0519	0.0261	0.0407
ρ	0.9188 (0.0239)	0.9585 (0.0221)	0.9635 (0.0161)	0.8869 (0.0413)	0.8579 (0.0297)
ADF p-value	0.0740	0.3855	0.4614	0.0619	0.0750
model (16)					
β_U	1.7395 (0.9785)	0.9848 (0.7968)	1.4326 (0.6829)	1.1368 (0.5197)	3.3247 (0.7189)
R^2	0.0141	0.0050	0.0119	0.0100	0.0450
model (17)					
β_C	−0.0811 (0.0271)	−0.0433 (0.0211)	−0.0358 (0.0166)	−0.1098 (0.0382)	−0.1410 (0.0292)
φ_C	0.9999 (0.0025)	1.0018 (0.0019)	0.9995 (0.0020)	0.9974 (0.0026)	0.9997 (0.0023)
R^2	0.9993	0.9996	0.9996	0.9984	0.9989

Notes: OLS estimates with Newey-West standard errors with 12 lags are in parentheses. The ADF test for the null of a unit root is based on a model with a drift and 12 lags. The table reports its p-value. R^2 denotes the goodness-of-fit R^2 statistic.

Figure 1: The left charts plot (on the same scale) the spot rate, one-month forward rate and their difference (spot-forward spread) for four (British pound (BP), German mark (DM), Swiss franc (SF) and Canadian dollar (CD)) currencies. The right charts zoom in on the dynamics of the spot-forward spread for these four currencies.