Economic Modelling 125 (2023) 106352

Contents lists available at ScienceDirect

Economic Modelling

journal homepage: www.journals.elsevier.com/economic-modelling

Testing for integration and cointegration when time series are observed with noise

Angelica Gianfreda [a,c], Paolo Maranzano [b,d], Lucia Parisio [b], Matteo Pelagatti [b,*]

[a] Department of Economics, University of Modena and Reggio Emilia, Via Jacopo Berengario, 51, Modena, 41121, Italy
[b] Department of Economics, Management and Statistics, University of Milano-Bicocca, Piazza Ateneo Nuovo, 1, Milan, 20126, Italy
[c] Energy Markets Group, London Business School, Regent's Park, London, NW1 4SA, United Kingdom
[d] Fondazione Eni Enrico Mattei (FEEM), Corso Magenta, 63, Milan, 20123, Italy

ARTICLE INFO

Dataset link: https://github.com/PaoloMaranzano/AG_PM_MP_LVP_FilterCointStat

Keywords:
Filtering
Cointegration
Unit root
Stationarity
Electricity prices
Pairs trading

ABSTRACT

When time series are observed with noise, the popular Augmented Dickey–Fuller (ADF) unit root test and Johansen's cointegration test are oversized: the ADF tends to conclude for stationarity too often and Johansen's test finds too many cointegrating relations. This fact is well-known but no simple solution has been proposed in the literature. In this work, we show why this happens and prove theoretically and by Monte Carlo simulations how three different filtering approaches can significantly improve the performance of the two tests applied to noisy data without harming their properties when observations are free from noise. We show how conclusions can change when using filtered time series in two real applications: one concerning wholesale electricity prices in European countries, and the second warning against pairs trading strategies based on spurious cointegrating relations among stock prices.

1. Introduction

1.1. Motivation

Tests for integration and cointegration of time series may fail and lead to wrong conclusions when data are observed with noise. We show that the popular Augmented Dickey–Fuller (ADF) and Johansen's tests are significantly oversized and their size gets worse as the amount of noise increases. In this regard, the ADF finds stationarity too often and Johansen's test tends to suggest too many cointegrating relations.

We show why these tests are unable to cope with noisy time series and propose three different filtering approaches as possible solutions to the problem. The first approach is based on reducing the frequency of the time series by simple averaging. The other two approaches exploit a state-space representation of the time series containing a stochastic trend and apply the tests on the trend component extracted by means of the Kalman filter and smoother.

We provide theoretical results for the simple random walk plus noise model that motivates our expectation that both the ADF and Johansen's tests applied to the filtered data should be much more reliable in terms of size. A battery of Monte Carlo experiments confirms that the sizes of the ADF test applied to the filtered time series remain at acceptable levels, the best one being that based on state-space smoothing, while the size of the same test applied to raw data increases quickly towards 1. This robustness has a cost in terms of power. As for Johansen's sequential procedure, our results provide a very strong message: the approach based on the Kalman filter is the one that selects the correct number of cointegrating relations with the highest probability. Furthermore, all simulations show that, in absence of additive noise, the tests applied to the state-space filtered or smoothed time series are equivalent to the ones applied to raw data. This suggests that filtering before testing should become a common practice when using the ADF and Johansen's procedure.

Beyond theoretical results, we provide empirical evidence on the effectiveness of different filtering strategies by applying the technique to real series. In this regard, financial, environmental and electricity markets provide prominent examples of time series characterized by the kind of features we want to deal with. However, it is important to emphasize that the filtering technique is useful in all other cases when data show similar characteristics.

1.2. Related work

The size distortion of the ADF test in the presence of measurement errors or a moving average component in the data-generating process (the first condition implies the second) has been known and enquired about for many years, even though a simple general solution to the

* Corresponding author.
E-mail address: matteo.pelagatti@unimib.it (M. Pelagatti).

https://doi.org/10.1016/j.econmod.2023.106352
Received 6 September 2022; Received in revised form 4 April 2023; Accepted 6 May 2023
Available online 10 May 2023

problem has never been found. Indeed, Said and Dickey (1984) proved that if the number of lags k in the ADF auxiliary regression grows with n, being n the length of the time series, but not too quickly (i.e., $k = o(n^{1/3})$), the asymptotic distribution of the ADF statistic under the null is the usual one, even if the data generating process has a moving average component. Schwert (1989) shows using Monte Carlo experiments that in finite samples, especially when some roots of the MA operators are close to unity, the size distortion of the ADF can be substantial and the nonparametric correction proposed by Phillips and Perron (1988) does not solve the problem. Gonzalo and Pitarakis (1998) reach similar conclusions by deriving the distortion of the first two moments of the ADF statistic when an MA(1) component is present. Galbraith and Zinde-Walsh (1999) address the problem both theoretically and by Monte Carlo simulation but, again, they conclude that the size distortions are particularly difficult to control when the moving average part contains a root near unity.

The article by Fischer (1990) is the first one to assess by Monte Carlo methods the actual size of residual-based cointegration tests when data are observed with measurement errors. They find that the tests are oversized and, therefore, cointegration is found too often. Haug (1996) considers a wider range of tests, including Johansen's, and he concludes that there is a trade-off between power and size-distortion.[1] Hassler and Kuzin (2009) derive the asymptotic distribution of Johansen's test when the process is observed with measurement error and propose a nonparametric correction to the statistic that leads to the usual asymptotic distribution of the test under the null. However, the simulations they carry out to assess the finite-sample behaviour of their test consider only relatively small noise-to-signal ratios (an MA component with roots bounded away from the unit circle). Our conjecture is that being their approach similar to Phillips and Perron (1988), it should show similar size distortion issues as found by Schwert (1989). More recently, Hong et al. (2016) carried out a Monte Carlo simulation exercise that evaluates the finite-sample behaviour of Johansen's sequential procedure when time series are observed with error and they conclude that "the cointegrating rank is more likely overestimated with measurement errors". Finally, Habimana et al. (2021) find size distortion in Johansen's test in the presence of observational noise and propose to apply Johansen's procedure after wavelet-denoising to correct the size. Their method depends on a tuning parameter that has to be fixed by the user and they propose a rule.

Thus, the size-distortion problems of ADF and Johanses's test in the presence of measurement errors are well known in the literature. The only two solutions that have been proposed are based on nonparametric adjustments of the test statistics and, very recently, wavelet denoising. Note that the first solution seems to fail when the measurement error is strong or, equivalently, when the root of the MA component is close to the unit circle.

The solution proposed in this work is also based on prefiltering. However, we use linear filters that should be more familiar to the applied economist or econometrician and do not need any tuning by the user. One of the three filtering techniques we propose has one tuning parameter whose meaning is straightforward and, often, has a natural value we can assign. However, we also provide an optimal way to assign a value to this tuning parameter.

The idea of improving inference by eliminating a fixed number of noisy Fourier frequencies goes back to Hannan (1969, point (c) on page 584) and has been enquired in various contexts by Watson (1993), Diebold et al. (1998) and Christiano and Vigfusson (2003). Proietti (2008) analyses the effect of band spectral estimation on signal extraction. However, to the best of our knowledge, there is no published work on using band-filtered time series to improve the performance of the unit root and cointegration tests.

[1] This is not surprising as robustness usually comes with a cost in terms of efficiency. The solutions we present in this work are no exception.

1.3. Some fields of application

Tests for integration and cointegration of time series are applied in many real-world cases to highlight some long-run behaviour that can be useful in designing appropriate investment and hedging strategies. The issues of liberalized electricity markets as well as strategic behaviour in trading strategies are both extremely relevant in times of uncertainties and political instabilities. Then, the two considered case studies strongly show the relevance of the developments proposed in this paper.

The organized market platforms for the exchange of electricity are probably one of the best examples of market architectures influencing price time series. Since the beginning of the '90, they have become the standard mechanism used in all industrialized and emerging countries to govern the liberalization of previously monopolistic markets. When competitive electricity markets started to be created in several countries around the world, many scientific papers concluded or assumed that electricity prices were mean-reverting (for instance Huisman and Mahieu, 2003; Weron et al., 2004; Ricky Rambharat et al., 2005; Geman and Roncoroni, 2006; Escribano et al., 2011; Keles et al., 2012; Fernandes et al., 2021). However, in many markets, the most frequent (marginal) generation plants determining the wholesale price are fuelled by hydrocarbons (carbon, gas, oil), whose log-price dynamics are known to be well approximated by integrated (thus, non-mean-reverting) processes. Therefore, the resulting electricity price is a function of non-mean-reverting processes and should be non-mean-reverting as well.

In the long run, electricity prices are determined mainly by demand, available installed generating capacity and fuel prices. In the short run, prices are influenced by many micro-structural factors such as line congestions, bidding strategies, and plant (programmed and unexpected) maintenance. More recently, price levels are influenced by import/export flows due to a larger degree of integration of EU electricity markets while price variability is positively affected by the intermittent generation from Renewable Energy Sources (RES). For these reasons, electricity prices are determined not only by their market fundamentals but are also buried in high-variance noise. The influence of short-run components is exacerbated during uncertain times characterized by wars, such as the one in Ukraine, which is heavily influencing the international dynamics of natural gas (and all other hydrocarbons') prices. Extreme movements in global fuel prices can be easily observed at the time of writing, and these are immediately reflected in the bills to final customers. This particular noise is responsible for the unreliable results that least-squares-based techniques, such as ADF and Johansen's tests, tend to produce. Bosco et al. (2010) and Pelagatti and Sen (2013) develop robust tests for stationarity and cointegration, whereas in some other applied papers, the same problem was addressed by using simple filtering techniques such as weekly means or medians (Bosco et al., 2010; Gianfreda et al., 2016a) or extracting the long-run component using unobserved component models (UCM) with Kalman filtering and smoothing (Gianfreda et al., 2016a, 2019).

While electricity prices were the main reason for developing the filtering techniques discussed in this paper, there are other interesting applications for financial data. Indeed, also financial quantities can often be seen as having a long-run component buried in some noise. An example is represented by markets organized with order books and continuous trades. Scholes and Williams (1977) consider the staleness of prices or probability of a trade to describe the functioning of these markets. In this regard, the applied research has stressed that observed returns have different properties with respect to the true return series (Ahn et al., 2015). Observed returns may show higher variance, heteroskedasticity, negative autocorrelation, heavy tails and positive cross-correlation among stocks when dealing with multiple securities and portfolios. The bid/ask spread may also influence the price variance and autocorrelation in markets where dealers post offers and outsiders can trade accepting current prices (Glosten and Milgrom,

1985). Kyle (1985) analyses a model of strategic trade in which the information signal is buried into noise induced by the operation of uninformed traders. When there are many competitive players in the market, the price adjustment to order flow is low. Hence, Kyle's (1985) model is an example where the market structure and agents' behaviour may influence the price dynamics.

Filtering techniques may also be useful in the analysis of electronic markets characterized by an increased number of transactions whose average size is lower than those observed in traditional markets and where the average holding period has shrunk. Information influences prices which show rapid movements, called *flash crashes*, meaning that they are short and relatively deep events with price movements in excess with respect to the amplitude implied by fundamentals.

Another area of interest for the long-run analysis of financial price series is related to the so-called *statistical arbitrage* and, in particular, *pairs trading strategies*. This trading strategy applies cointegration analysis to identify pairs of stocks characterized by an equilibrium price ratio. When price deviations are mean-reverting with a certain degree of persistence, the consequent predictability allows for profitable trades. Then, the ability to correctly determine the real degree of cointegration is essential to avoid false trading signals.[2] In Section 4.2 we will consider a pairs strategy case showing the usefulness of filtering techniques to avoid false signals.

Finally, other examples of very noisy time series can be found in the context of climate change and the environment. Time series for air quality and weather conditions have deeply different factors governing the long- and short-run behaviour (Mudelsee, 2019). As stated by Horowitz and Barakat (1979), air pollutant concentrations are not independently and identically distributed, and are generated by non-stationary autocorrelated stochastic processes.[3] Also, environmental and climate data can be affected by outliers and extreme values (Mudelsee, 2020), which act as confounders in the models.

All these fields of applications necessarily lead to considering the use of appropriate and robust analytical tools with respect to the various characteristics of the data. Therefore, focusing on the integration/cointegration issue, the identification of the long-run components of the data-generating process through pre-filtering procedures, such as those proposed in this paper, should make the analyses more reliable and lead to better conclusions.

The paper is structured as follows: Section 2 illustrates the theory underlying the problem using a simple random walk plus noise model and proves how three filtering techniques can solve it. In Section 3, we present the results of Monte Carlo experiments that compare the performance of ADF and Johansen's tests applied to raw and filtered data. In Section 4, the proposed methodologies are then applied to real-time series of electricity and financial prices, in the second case in the context of pairs trading. Section 5 draws some conclusions.

[2] Examples taken from financial and commodity markets include: Caldeira and Moura (2013), Girma and Paulson (1999), Simon (1999) and Wahab et al. (1994). Another application is presented by Miao (2014) who consider the pair strategy for the U.S. equity market. The trading system was relatively market neutral and unrelated to the S&P index; as such, results of out-of-sample testing showed a good performance of the trading strategy in particular in periods of low market performance. The same approach has been applied to "crush spread", between soybeans and its derived products, to gold–silver spread and in many other cases where an input–output relationship or a market arbitrage linkage is in place. For a survey on statistical arbitrage and pairs trading refer to Krauss (2017)

[3] In general, air quality series are characterized by seasonality, strong persistence, long-memory (Chelani, 2013; Varotsos et al., 2005) or non-stationarity (Ng and Yan, 2004), right skewness (Windsor and Toumi, 2001), and fractal behaviour (Lee, 2002).

Table 1
Actual ADF test size for different values of the signal-to-noise ratio for a nominal size of 5%.

λ	θ	Size
0.001	0.97	1.00
0.010	0.90	0.89
0.100	0.73	0.34
1.000	0.38	0.10
10.000	0.08	0.05

2. Methodology

2.1. Why the ADF and related tests fail when integrated time series are observed with strong noise

Let us consider a simple random walk plus white noise model: for $t = 1, 2, \dots, n$,

$$\begin{aligned} y_t &= x_t + \varepsilon_t, \qquad \varepsilon_t \sim \mathrm{WN}(\sigma_\varepsilon^2) \\ x_t &= x_{t-1} + \eta_t, \quad \eta_t \sim \mathrm{WN}(\sigma_\eta^2), \end{aligned} \tag{1}$$

where the notation $\varepsilon_t \sim \mathrm{WN}(\sigma_\varepsilon^2)$ is to be read as "ε_t is a white noise sequence with variance σ_ε^2". Moreover, let

$$\lambda = \frac{\sigma_\eta^2}{\sigma_\varepsilon^2} \geq 0$$

be the *signal-to-noise ratio*. In other words, y_t is the observed noise-corrupted process and x_t is the true signal that evolves as a random walk and the noise corruption is represented by ε_t.

It is straightforward to prove (see in the Appendix) that the process y_t has the reduced ARIMA$(0, 1, 1)$ form

$$\Delta y_t = \eta_t + \varepsilon_t - \varepsilon_{t-1} = \zeta_t - \theta\zeta_{t-1}, \qquad \zeta_t \sim \mathrm{WN}(\sigma^2) \tag{2}$$

with

$$\theta = 1 + \frac{\lambda - \sqrt{\lambda^2 + 4\lambda}}{2}, \qquad \sigma^2 = \frac{\sigma_\varepsilon^2}{\theta}, \tag{3}$$

where Δ is the difference operator (i.e., $\Delta x_t = x_t - x_{t-1}$). When the signal-to-noise ratio is zero the MA coefficient is $\theta = 1$, the unit root operator cancels out with the MA operator and y_t turns out to be just a white noise sequence:

$$\Delta y_t = \Delta\zeta_t \quad \Leftrightarrow \quad y_t = \zeta_t.$$

When λ is close to zero, the MA coefficient approaches 1 from below. In this case, the exact cancellation does not take place, however in small samples the process y_t is almost indistinguishable from white noise. Such a MA process still has the purely AR representation

$$\Delta y_t = \zeta_t + \theta\Delta y_{t-1} + \theta^2\Delta y_{t-2} + \theta^3\Delta y_{t-3} + \cdots,$$

however, this representation cannot be well approximated by an AR(p) process with small p because θ^j approaches zero very slowly. Now, most unit root tests deriving from the Dickey–Fuller test such as ADF (Said and Dickey, 1984), ADF-GLS (Elliott et al., 1996), Johansen (Johansen, 1991) are based on autoregressive approximations and, if the y_t is generated as above with λ close to zero, then they are severely oversized (see Galbraith and Zinde-Walsh, 1999, for example).

Table 1 gives an idea of the amount of size distortion as a function of the signal-to-noise ratio and of the moving average coefficient. The reported values are the results from a Monte Carlo experiment with the following characteristics: 10,000 replications of time series of length 200, the auxiliary AR(p) model order selected according to the Akaike Information Criterion (AIC) with p ranging from 0 to 20, the ADF test applied with a drift coefficient.

Model (1) is extremely simple and could be generalized by assuming that ε_t and η_t are ARMA processes, for example. However, this generalization would come with a much more cumbersome algebra and

lengthy proofs, without adding much insight to the problem. Indeed, the feature of the process (1) that harms the ADF and Johansen's test is the close-to-unity coefficient θ in Eq. (2) that almost cancels out with the unit root in the difference operator. Thus, if ε_t and η_t are defined as ARMA processes with roots that are bounded away from the unit circle, then y_t becomes an ARIMA$(p,1,q)$ process where the only close-to-unity root of the MA characteristic equation (which cannot be well approximated by an AR model) is due to the differenced noise component as in Eq. (2). Low-pass filtering would have a similar effect also in this case because it annihilates the peak in high-frequency that dominates the spectrum and AR models are not able to approximate.

2.2. Filtering

We propose three simple time series filters that can improve the performance of unit root and cointegration tests:

1. reducing the frequency of the time series by taking averages, for instance by working on weakly means of daily or hourly prices;
2. extracting the level component using the Kalman filter in an unobserved component model (UCM) containing trend, noise and, possibly, seasonal components;
3. extracting the level component using the smoother in a UCM containing trend, noise and, possibly, seasonal components.

In this section, we analyse the effects of these filters on the reduced form of the model on time series generated by Eq. (1). The next section will illustrate the effect of these filters on the size and power of the ADF and Johansen tests by Monte Carlo experiments.

2.2.1. Frequency reduction by averaging

The first filter reduces the sampling frequency of the time series by averaging every m non-overlapping consecutive observations. The following results show that this filter applied to a random walk plus noise is able to annihilate the MA(1) component in the reduced form

Theorem 1. *For $t = 1, 2, \ldots$, let y_t be defined as in (1) and let*

$$\bar{y}_t = \frac{1}{m}\sum_{i=0}^{m-1} y_{t-i}, \qquad \bar{\eta}_t = \frac{1}{m}\sum_{i=0}^{m-1} \eta_{t-i}, \qquad \bar{\varepsilon}_t = \frac{1}{m}\sum_{i=0}^{m-1} \varepsilon_{t-i}, \tag{4}$$

where $\bar{y}_t$ is sampled over the set of time points $t \in \mathcal{T} := \{m, 2m, 3m, \ldots\}$. Then, over the time set $\mathcal{T}$, we have

$$\bar{y}_t - \bar{y}_{t-m} = \bar{\eta}_t + \bar{\varepsilon}_t - \bar{\varepsilon}_{t-m},$$

with $\bar{\varepsilon}_t$ white noise sequence with variance σ_ε^2/m, and $\bar{\eta}_t$ MA(1) process with the following moments

$$\mathrm{VAR}\left(\bar{\eta}_t\right) = \sigma_\eta^2\left[\frac{(m-1)(2\,m-1)}{3m} + 1\right],$$
$$\mathrm{COV}\left(\bar{\eta}_t, \bar{\eta}_{t-m}\right) = \sigma_\eta^2\frac{(m-1)(m+1)}{6m},$$
$$\mathrm{COR}\left(\bar{\eta}_t, \bar{\eta}_{t-m}\right) = \frac{(m-1)(m+1)}{2(m-1)(2m-1)+3m}.$$

Corollary 1.1. *Under the hypotheses of Theorem 1, the process $\bar{y}_t$ is ARIMA(0, 1, 1) for $t \in \mathcal{T}$ with first-order autocorrelation given by*

$$\rho = \mathrm{COR}\left((\bar{y}_t - \bar{y}_{t-m}), (\bar{y}_{t-m} - \bar{y}_{t-2m})\right) = \frac{\lambda\frac{(m-1)(m+1)}{6m} - \frac{1}{m}}{\lambda\left(\frac{(m-1)(2m-1)}{3m} + 1\right) + \frac{2}{m}} \tag{5}$$

and moving average coefficient given by

$$\theta = \frac{-1 + \sqrt{1-4\rho^2}}{2\rho}.$$

The left panel of Fig. 1 depicts the value of the MA coefficient θ as a function of the window size m, for various values of the noise-to-signal ratio λ^{-1}. The right panel of the same figure depicts the locus of points (λ^{-1}, m) for which $\theta = 0$. In principle, for any signal-to-noise ratio λ, there is a value m that annihilates the MA component. Of course, since m can only take integer values, θ will be in general close to zero and not exactly equal to zero. If the noise is very strong, the original time series must be rather long as the length of $\bar{y}_t$ is given by $\lfloor n/m \rfloor$, where n is the length of the time series y_t.

When data show seasonal patterns, the value of m can be set equal to the seasonal period serving the scopes of annihilating the seasonal component and shrinking the MA(1) coefficient. However, we can base our choice of m plugging the estimated coefficient of an IMA$(1,1)$ model in Eq. (5) of Corollary 1.1. In fact, by solving Eq. (3) for λ, we can estimate the signal-to-noise ratio from the estimate of θ as $\lambda = (\theta-1)^2/\theta$. An estimate of the optimal value of m can be obtained by setting $\rho = 0$ in Eq. (5) and solving for m:

$$m = \sqrt{\frac{6+\lambda}{\lambda}} = \sqrt{1 + \frac{\theta}{(\theta-1)^2}}. \tag{6}$$

Notice, that the solution is real and, so, the actual m can be fixed at the nearest integer.

2.2.2. Signal extraction in an unobserved component model

An alternative way to reduce the noise in the process y_t defined in Eq. (1) is by estimating the random walk component x_t by projecting it on the linear span of $y_1, y_2, \ldots, y_s$ where s is either equal to t or to n. This operation can be easily carried out by stating the model in state-space form and running the Kalman filter (for $s = t$) and smoother (for $s = n$) on the level component x_t (the volumes by Harvey, 1989; Durbin and Koopman, 2001; Pelagatti, 2015, are detailed references on the subject).

Model (1) is already in state-space form: using the notation of Durbin and Koopman (2001), the transition matrix is $\boldsymbol{T} = 1$, the covariance matrix of the disturbance is $\boldsymbol{Q} = \sigma_\eta^2$, the observation matrix is $\boldsymbol{Z} = 1$ and the covariance matrix of the measurement error is $\boldsymbol{H} = \sigma_\varepsilon^2$. Since the state variable x_t is nonstationary this component is generally initialized with arbitrary mean and infinite variance. If the time series presents seasonal patterns, a seasonal component can be added to the model. For example, using a seasonal dummy approach, the time series could be modelled as

$$y_t = \left[\begin{array}{c|ccc} 1 & 1 & 0 & \ldots & 0 \end{array}\right]\boldsymbol{\alpha}_t + \varepsilon_t$$
$$\boldsymbol{\alpha}_t = \left[\begin{array}{c|cc} 1 & \mathbf{0}^\top & 0 \\ \hline 0 & -\mathbf{1}^\top & -1 \\ \mathbf{0} & \boldsymbol{I}_{s-2} & \mathbf{0} \end{array}\right]\boldsymbol{\alpha}_{t-1} + \left[\begin{array}{c} \eta_t \\ \hline \zeta_t \\ \mathbf{0} \end{array}\right],$$

where s is the seasonal period, $\mathbf{0}$ is a column vector of zeros, and $\mathbf{1}$ is a columns vectors of ones. The first element of the state vector $\boldsymbol{\alpha}_t$ is the random walk, the second element is the seasonal component and the remaining $s-2$ variables are lags of the seasonal component. If the variance of ζ_t is positive, then the seasonal component evolves over time, while if it is zero the seasonal pattern is deterministic. In order to keep the reasoning simple, we derive results for model (1), however, we can reasonably expect that things would not change much if a (deterministic or slowly changing) seasonal component were present.

The Kalman filter projects the unobservable level x_t on the observations $y_1, y_2, \ldots, y_t$ up to time t; while the smoother projects x_t on all the observations $y_1, y_2, \ldots, y_n$. The Kalman filter and smoother for x_t are linear filters whose weights are different for every t: the former is only backwards looking,

$$x_{t|t} = \sum_{i=-t+1}^{0} w_{ti}\, y_{t+i}, \qquad t = 1, 2, \ldots, n, \tag{7}$$

while the latter is two-sided,

$$x_{t|n} = \sum_{i=-t+1}^{n-t} \omega_{ti}\, y_{t+i}, \qquad t = 1, 2, \ldots, n. \tag{8}$$

The reader interested in actual computation of the weights w_{ti} and ω_{ti} should refer to the work of Koopman and Harvey (2003). Nonetheless,

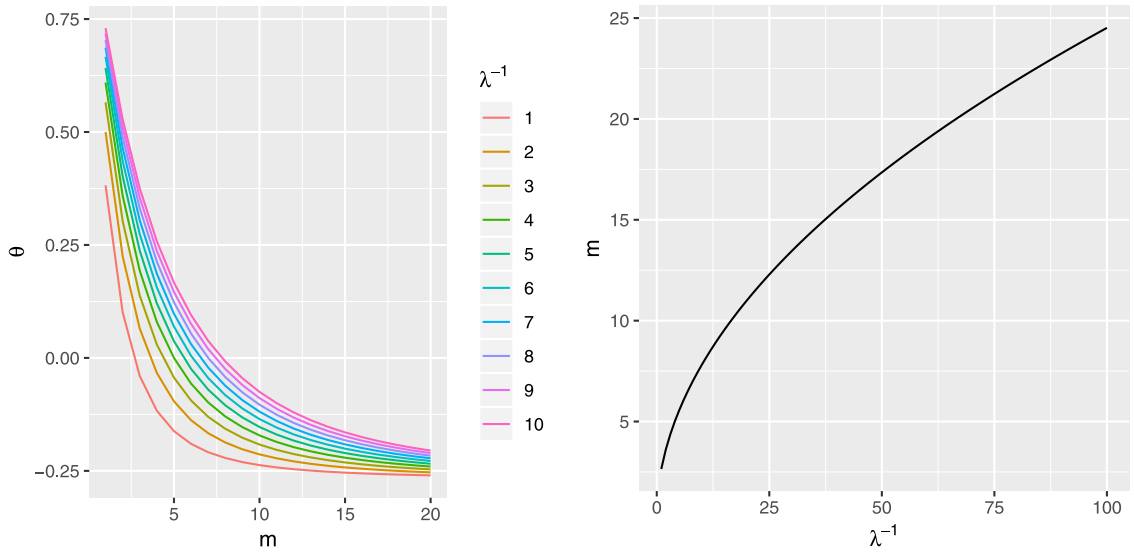


Fig. 1. Left: MA coefficient of mean-filtered time series as a function of the window for different values of the noise-to-signal ratio. Right: the values of the window size m that achieve an MA coefficient equal to zero as a function of the noise-to-signal ratio.

when t is not too close to 1, the Kalman filter for the random walk plus noise model can be well approximated by its steady state version, say $\tilde{x}_t$, given by the recursive filter[4]

$$\tilde{x}_t = \gamma y_t + (1-\gamma)\tilde{x}_{t-1}, \tag{9}$$

where

$$\gamma = \frac{\sqrt{\lambda^2+4\lambda}-\lambda}{2}. \tag{10}$$

Similarly, when t is not too close to 1 or n, the smoother can be well approximated by its steady-state version, say $\hat{x}_t$, which is given by the backward recursion on $\tilde{x}$

$$\hat{x}_t = \gamma \tilde{x}_t + (1-\gamma)\hat{x}_{t+1}. \tag{11}$$

The following result characterizes the processes followed by the steady-state filter and smoother when applied to the data-generating process (1).

Theorem 2. *Assume that the variances σ_ε^2 and σ_η^2 of the process defined in Eq. (1) are known and $\lambda = \sigma_\eta^2/\sigma_\varepsilon^2$ is the signal-to-noise ratio. Then, $\tilde{x}_t$ is a random walk and $\hat{x}_t$ is an ARIMA$(1,1,0)$ process with autoregressive coefficient $\phi = 1-\gamma$.*

In applications, the unknown variances are replaced with their Gaussian (quasi) maximum likelihood estimates $\hat{\sigma}_\varepsilon^2$ and $\hat{\sigma}_\eta^2$ and the population signal-to-noise ratio with its estimate $\hat{\lambda} = \hat{\sigma}_\eta^2/\hat{\sigma}_\varepsilon^2$. At this point, we could try to provide a result that characterizes the processes followed by the filter and smoother when the real parameters are substituted by consistent estimates. However, that result would not be very informative as, in practice, the practitioner uses the same observations to estimate the parameters and extract the signal. We are afraid that characterizing the filter/smoother based on parameters estimated on the same data is an almost impossible task. Fortunately, we can rely on Monte Carlo simulation to assess the quality of the procedure we are proposing.

[4] A simple derivation of this result can be found in Example 5.12 of Pelagatti (2015) and this particular form of the coefficient γ in Harvey (2006, Eq.13).

3. Results from Monte Carlo simulations

To verify the empirical effects of the frequency-reduction and UCM filters on the ADF and Johansen tests, we performed a set of Monte Carlo simulations. Each experiment evaluates the performances of ADF and Johansen tests both under the null and alternative hypothesis, in order to evaluate the size and power of the tests.

In our simulations, we consider values for the noise-to-signal ratio in the range of 0–10. Indeed, by estimating UCMs on the 24 daily time series of Italian day-ahead electricity prices for each hour of the day, we obtain noise-to-signal ratios as high as 12.6 (at the 15th hour) and as low as 0 (midnight and neighbouring hours).

3.1. ADF test

The first set of experiments aims at assessing the performance of the ADF test under the random walk plus noise model. We simulate time series from a random walk buried into leptokurtic noise and each simulation experiment is characterized by a different combination of noise-to-signal ratio and kurtosis. The data generating process (DGP) for the observation y_t is

$$y_t = x_t + \sqrt{c}\,\varepsilon_t, \qquad \varepsilon_t \sim \text{i.i.d. } t_\nu,$$
$$x_t = x_{t-1} + \eta_t \qquad \eta_t \sim \text{i.i.d. } N(0,1),$$

where ε_t is the leptokurtic noise generated by a standardized Student's t with ν degrees of freedom (DF) and $c = \lambda^{-1}$ is the fixed parameter identifying the noise-to-signal ratio. The number of DF governs the thickness of the tails of the noise component: the lower the DF, the larger the kurtosis.

For each experiment, we simulate 10,000 paths of length 1,095, corresponding to 3 years of daily observations, for all of the possible pairs of noise-to-signal ratio c in $\{0,1,2,\ldots,10\}$ and degrees of freedom ν in $\{3,6,9,12\}$. On each of these time series, we apply the mean filter (*mean*), the Kalman filter (*ucm_flt*) and the smoother (*ucm_smo*).

The ADF statistic (with drift and the number of lags selected by AIC) was computed on every simulated time series. The empirical rejection rates for a nominal size of 5% are represented in Fig. 2.

Looking at the four graphs we can conclude that

1. the size of the test applied to raw data quickly drifts away from its nominal size as the noise-to-signal ratio increases;
2. the ADF applied to any pre-filtered time series keeps acceptable size;

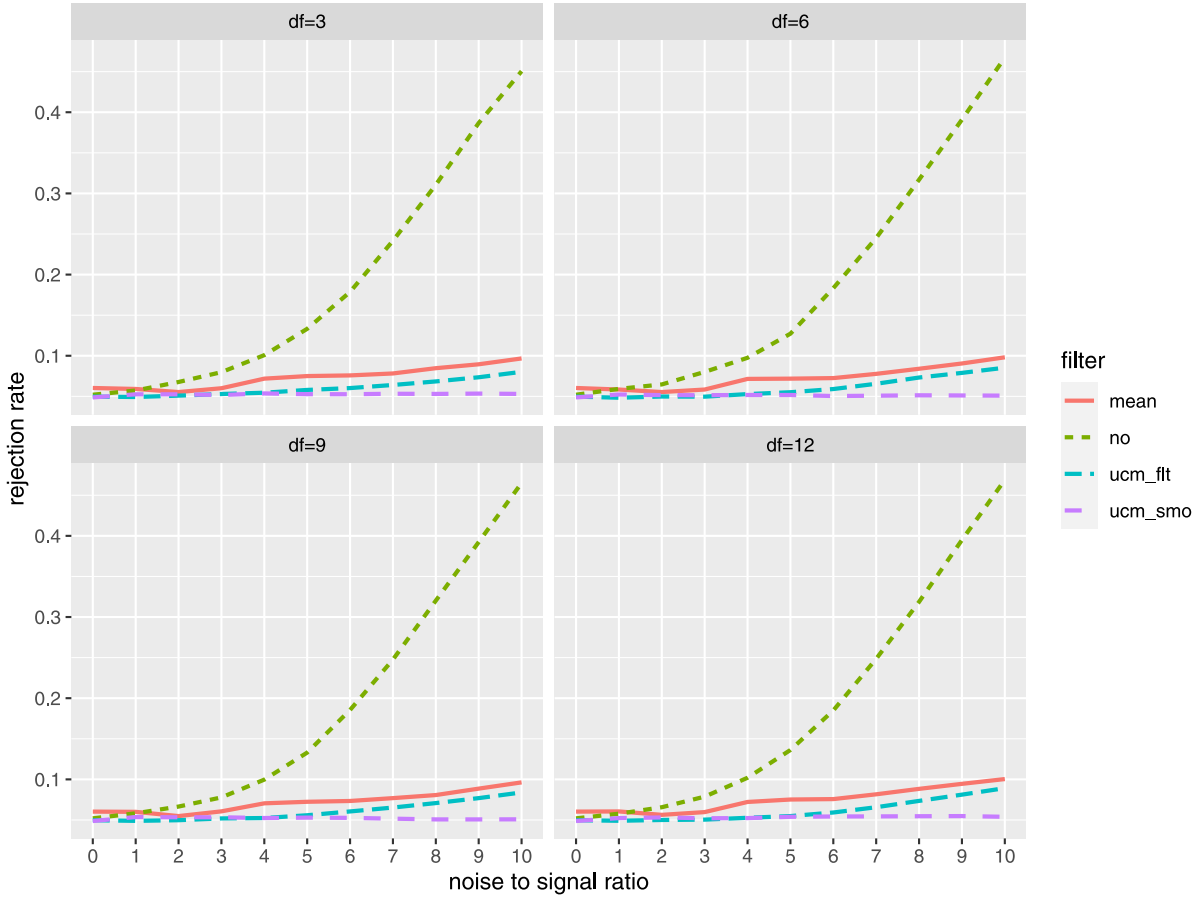


Fig. 2. Actual size of the ADF test for a nominal size of 5%. Note that *no* represents the raw series, *mean* and *ucm_fit* stand for the *mean* and *Kalman* filters, whereas *ucm_smo* represents the *smoother*.

3. the size of the ADF test applied to the smoothed signal (*ucm_smo*) time series is equal to the nominal size regardless of the noise-to-signal ratio;
4. the thickness of the tails of the noise distribution has virtually no effect on the size of the ADF.

The latter remark may sound surprising, however, the asymptotic distribution of the ADF test is derived under the hypothesis of finite second moments and in our simulations, the variance of the noise ϵ_t is always finite and the sample size large ($n = 1095$).

The size-adjusted power of the ADF test for a size of 5% and when the data generating process is an AR(1) with Student's t noise and autoregressive coefficient $\phi = 0.98$ is shown in Fig. 3.

As noted by Haug (1996), there is some trade-off between power and size distortion. Robustness comes at some costs in terms of power. In particular,

1. the decrease in power observed for the frequency-reduction method (*mean*) is due to the consequent smaller sample size and it does not depend on the noise-to-signal ratio;
2. the power loss observed in tests applied to UCM-based filters increases with the amount of noise;
3. tests run on UCM-based filtered signals show no power loss with respect to the ADF applied on raw data, when no or moderate noise is present;
4. the thickness of the tails of the noise distribution has virtually no effect on the power.

The nice feature observed at point 3. is a natural consequence of the fact that in UCM-based filters the "amount" of filtering is data-dependent and, when no (or moderate) noise is present, the resulting signal is identical (or very close) to the raw data.

Let us draw some conclusions from this first set of simulation experiments. Despite the good power of the original ADF test even in the presence of measurement errors, in this situation, the ADF on raw data is unusable because of its potentially unlimited size distortion. All tests based on filtered time series are much more reliable in terms of size, with the one based on UCM smoothing maintaining a perfect size. This size robustness has a cost in terms of power: the power is independent of the quantity of noise, if one uses frequency reduction (*mean*), while the power deficiency grows with the amount of noise for the UCM-based filter and smoother. The latter filters show no power deficiency when the noise is absent or moderate, making their use safe even if the practitioner is not certain about the presence of observational noise in the data. As long as the second moment's requirement of the ADF test is met by the data, the tail thickness does not seem to harm its size or power.

3.2. Johansen test

Similarly to the univariate case, we developed simulation schemes for multivariate time series, which include both integrated and cointegrated processes, to evaluate the statistical properties of Johansen's test in the presence of noise.

Data are simulated according to a vector error correction model (VECM) with r cointegrating relations and $k = 4$ underlying times series augmented by a leptokurtic noise term. The noise is randomly generated by a standardized Student's t random variable with ν degrees of freedom and affect directly the VECM through the noise-to-signal ratio. We performed the simulation analysis on the number of cointegrating relations detected by the test considering the case of $r = 1$ and $r = 2$.

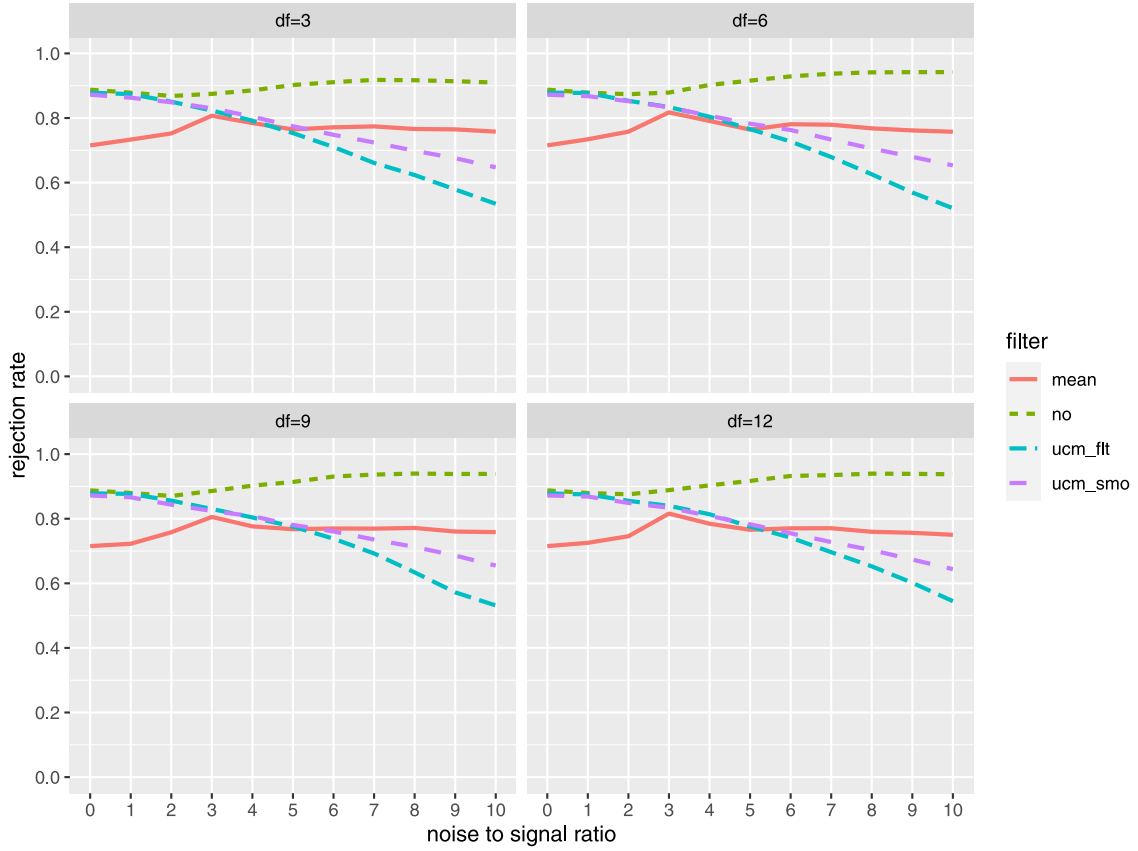


Fig. 3. Size-adjusted power of the ADF test for a size of 5% and data generating process being AR(1) plus Student's t noise with autoregressive coefficient $\phi = 0.98$. Note that *no* represents the raw series, *mean* and *ucm_fit* stand for the *mean* and *Kalman* filters, whereas *ucm_smo* represents the *smoother*.

For the simulation experiments involving Johansen's test, we used the following two DGP: the first one with $r = 1$ cointegrating relations has representation

$$r = 1: \quad \Delta x_t = \begin{bmatrix} 0.0 \\ 0.1 \\ -0.1 \\ 0.1 \end{bmatrix} \begin{bmatrix} 1 & -1 & 1 & -1 \end{bmatrix} x_{t-1} + \varepsilon_t, \tag{12}$$

and the second one with $r = 2$ cointegrating relations has representation

$$r = 2: \quad \Delta x_t = \begin{bmatrix} 0.0 & 0.0 \\ 0.1 & 0.0 \\ 0.2 & 0.0 \\ 0.0 & 0.2 \end{bmatrix} \begin{bmatrix} 1.0 & -0.5 & -0.5 & 0.0 \\ 0.0 & 1.0 & 0.0 & -1.0 \end{bmatrix} x_{t-1} + \varepsilon_t. \tag{13}$$

and the noisy series are

$$y_{it} = x_{it} + \gamma_i z_{it} \quad \text{with} \quad \gamma_i^2 = c \frac{VAR(\Delta y_{it})}{VAR(x_{it})} \tag{14}$$

where z_{it} is a leptokurtic noise and c is a fixed parameter representing the noise-to-signal ratio. Transforming the previous models from VECM to VAR(1) form, it is easy to assess that the characteristics roots of the processes are (1, 1, 1, 0.7) for Eq. (12) and (1, 1, 0.85, 0.80) for Eq. (13).

As in the previous experiment, we simulated 10,000 time series paths of length 1,095 (= 365 × 3) for all the paired combinations of noise-to-signal ratio $c = 0, 1, 2, \ldots, 10$ and degrees of freedom $\nu = 3, 6, 9, 12$. The three linear filters are then applied to the series, rising three further series sharing a common underlying process. The number of cointegrating vectors is finally tested on each simulated quartet using Johansen's trace test.

Johansen's trace test is a sequential procedure based on the null hypotheses that the true number of cointegrating vectors, r, is equal to a given number, $r^* < k$, i.e. $H_0 : r = r^* < k$, against the alternative hypothesis $H_1 : r = k$. The value $r^* = 0, 1, \ldots, k-1$ is updated sequentially from 1 to $k-1$ producing a sequence of tests, whose first non-rejection of the null hypothesis can be considered an estimate of r. The test rejects H_0 when the test statistic exceeds the tabulated critical values.

For each pair of degrees of freedom and noise-to-signal ratios, and for each sequential value r^*, we computed the test's size (rejection rate) as the proportion of tests rejecting the null hypothesis over the total number of simulations. For values of r^* lower than the real one, the expected rejection rate should be the closest possible to 1; while it should be equal to 0.05 when testing the true value of r^*. Johansen's procedure is generally used as a way to estimate the unknown number of cointegrating relations and, thus, we can compare the performances of the various versions of the test using the accuracy in estimating the right cointegration rank.

Figs. 4 and 5 show the empirical selection rates when the real cointegration rank is $r = 1$ and $r = 2$ respectively. When the noise-to-signal ratio is zero, both filtered and raw data perform similarly. However, as the noise increases its variance, the selection rates change considerably. The Kalman filter is the clear winner with the highest accuracy rate under all considered setups. In particular, the filter remains rather stable with values above 90% for $r = 1$ and over 80% when $r = 2$. The figures show also that for high noise-to-signal ratios, only the UCM-based filter is able to maintain acceptable performances. The procedure applied to raw data tends to select larger cointegration ranks, while frequency reduction (*mean*) tends to underestimate the cointegration rank. A similar, although less severe, tendency to underestimation is present when the procedure is applied to UCM-smoothed time series.

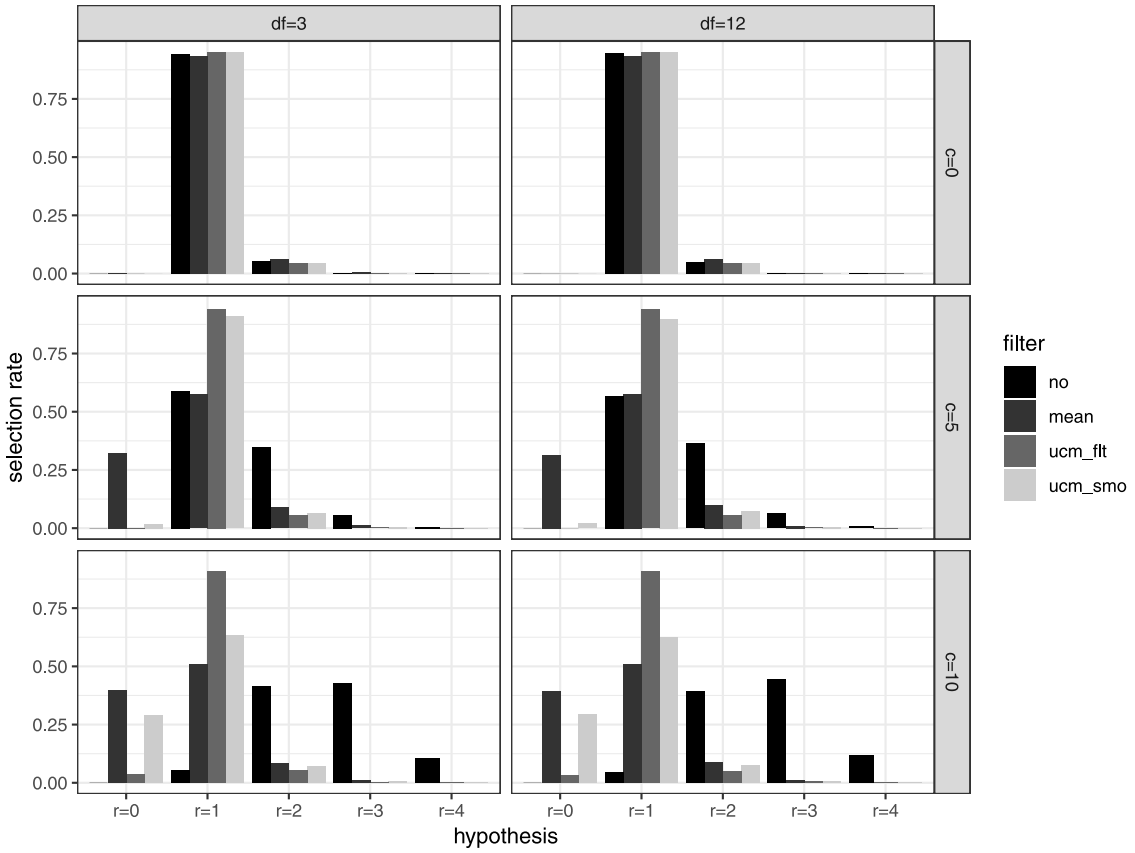


Fig. 4. Selection rate of Johansen's procedure when $r = 1$ and the nominal level of the tests is 5%.

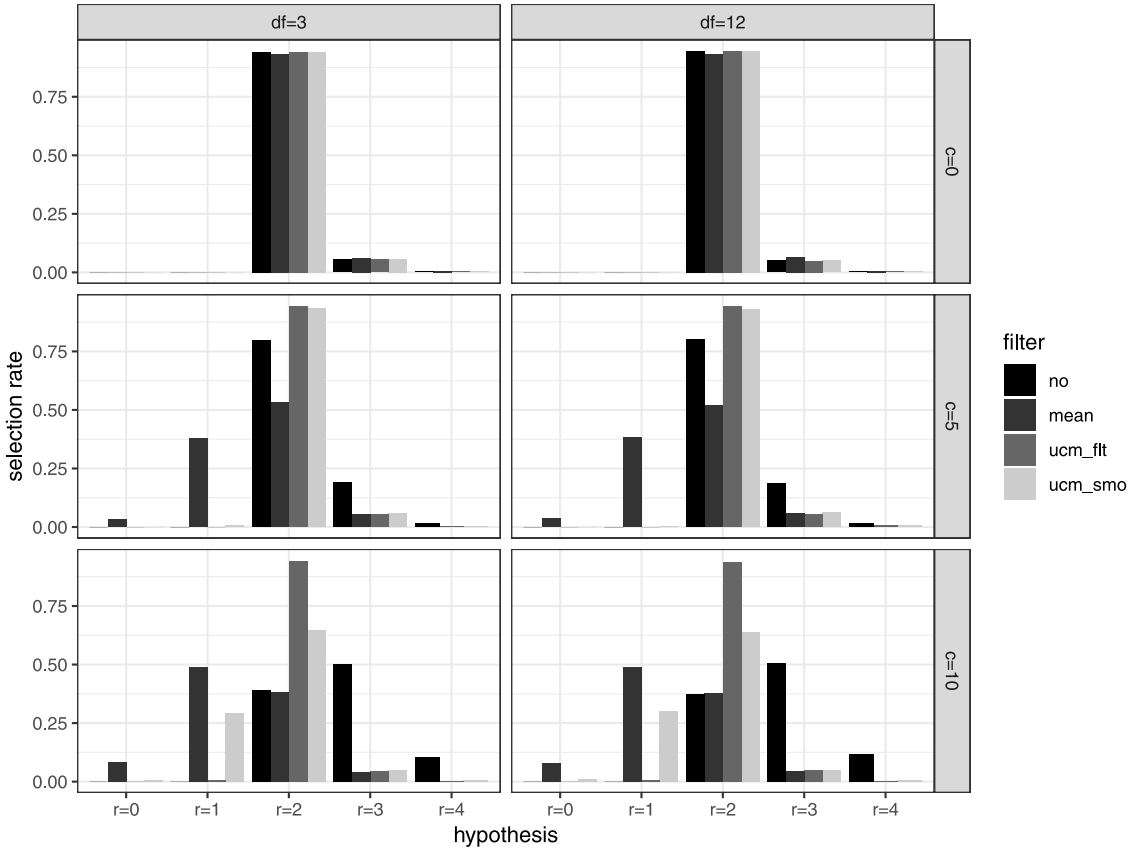


Fig. 5. Selection rate of Johansen's procedure when $r = 2$ and the nominal level of the tests is 5%.

In an additional online appendix, we provide Monte Carlo rejection rates as a function of the noise-to-signal ratio for all the tests building Johansen's sequential procedure. The figures in that appendix show the rejection rates for each of the alternative hypotheses of the test when the true number of cointegration relationships are $r = 1$ and $r = 2$ respectively. When the null hypothesis coincides with the truth, i.e. testing $r = 1$ when $r = 1$ and testing $r = 2$ when $r = 2$, the tests applied to unfiltered series (*no*) tend to over-reject the hypothesis even in presence of weak noise, while the three filters tend to preserve the nominal size of 5%, being the Kalman filter (*ucm_flt*) and the smoother (*ucm_smo*) the most successful. As for the power of the test, when the tested value is smaller than the true value of r, the test applied to raw data has the highest power,[5] while frequency reduction (*mean*) performs the worst. The test applied to the UCM-filtered time series starts losing power only for very high noise-to-signal ratios, with a net improvement with respect to the test applied to the UCM-smoothed time series. We also report the tests when the tested rank (null hypothesis) is larger than the true cointegration rank. In this case, both the null and the alternative are wrong, but within Johansen's procedure, we would be more satisfied with a rejection rate close to zero as this implies a smaller bias in the estimation of the cointegration rank. While all filters keep these rejection rates very close to zero, the test applied to raw data sees rejection rates increase with the level of noise.

4. Applications to real data

As mentioned in the introductory section, two main applications motivated the development of the filtering-before-testing methods proposed in this paper. The first one involves day-ahead electricity prices observed in various European markets. The second one uses the prices of the constituents of the S&P100 stock index to build statistical arbitrage strategies based on cointegration and pairs trading.

4.1. Electricity prices

Electricity prices are a prominent example of time series generated in a bid/ask context organized as a centralized auction market. To better understand their properties, we recall briefly the mechanisms and facts governing electricity markets.

Non-storability and several types of seasonality affecting the demand are perhaps the most relevant characteristics of electricity and these require the coordination of several actors, like producers, buyers, regulators and transmission system operators. Electricity is exchanged in market platforms based on least cost dispatch and different time schedules, starting many days ahead and ending in real-time delivery. Market System Operators (MSOs) handle the power exchange and collect demand and supply offers whereas Transmission System Operators (TSOs) manage transmission, congestion and the balancing of demand and supply close to real-time. Extreme prices generated by sudden events (like grid congestion or blackouts), technical constraints and potential market manipulation are only some of the issues that must be addressed for a viable and welfare-maximizing operation.

Each market session determines a price varying according to the hour of the day, locations and generating technologies. The most important session is called *day-ahead* market since it closes the day before delivery, where the largest share of the electricity consumed is traded in a series of 24 hourly auctions. For each hour bids and offers are submitted and aggregated according to the merit order criterion (the cheapest and, in general, more clean energy plants enter the supply curve before more costly and polluting technologies thanks to the *priority of dispatch*[6]). Market equilibrium is determined at the intersection of hourly aggregated demand and supply curves and a unique system marginal price (SMP) is set for all units that obtain a positive allocation.[7]

Any imbalances between demand and supply and local congestion are subsequently adjusted in intra-day sessions and in real-time balancing sessions. Additional details on properties and dynamics of intra-day and balancing prices can be found in Gianfreda et al. (2016a, 2018, 2019).

In the first phase of liberalization, particular attention was devoted to promoting the integration of EU national markets and the assessment of convergence between electricity and/or fuel prices, as in Zachmann (2008), Fezzi and Bunn (2009), Bunn and Gianfreda (2010), Bosco et al. (2010), Aatola et al. (2013), Huisman and Kiliç (2013), de Menezes and Houllier (2016) and de Menezes et al. (2016).

In Bosco et al. (2010) the main findings obtained through a robust multivariate long-run dynamic analysis reveal the presence of four highly integrated central European markets (France, Germany, the Netherlands and Austria) with prices sharing a common trend also with gas prices. de Menezes et al. (2016) analyse three markets (British, French and Nordpool) with price series showing stationary and non-stationary periods. The results highlight that British electricity spot prices are associated with fuel prices and not with price developments in connected markets, while the opposite is observed in the French and Nordpool day-ahead markets. de Menezes and Houllier (2016) analyse nine EU markets and observe that unit root tests for market integration are inadequate for assessing electricity spot market convergence because spot prices are found to be fractionally integrated and mean-reverting time series. Gugler et al. (2018) considered 25 EU markets obtaining mixed results as they prove that market integration increased from 2010 to 2012 and then decreased from 2012 to 2015, despite new investments in interconnection and market coupling. Since all the above empirical results are based on cointegration analysis, we believe that the filtering techniques proposed in this paper can help assess the correct degree of EU market integration avoiding spurious results.

The second phase of market restructuring and regulation starting in the last decade put the focus on the integration of renewable energy sources (RES) in the system and the best mix of market rules and economic incentives[8] able to foster RES deployment (Gianfreda et al., 2016a; Argentiero et al., 2017).

From an empirical point of view, it has been observed that RES reduce the level of day-ahead prices but increase their variability, and modify their relationship with fossil fuel prices (Gianfreda et al., 2016a,b). In particular, uncertainties and variability associated with RES production challenge the functioning of day-ahead and real-time sessions and introduce additional short-term noise in the price series, switching and reducing their dependencies on fuel prices in the long run. This has resulted in some EU countries becoming less integrated as electricity generated by RES increases (Gianfreda et al., 2016b).

From the above considerations, it is clear that electricity prices are determined by market outcomes but are subject to the influence

[5] The exposed power is not size-adjusted.

[6] Priority dispatch implies that quantities produced from RES units are supplied at a zero price. Thus the supply curve shifts to the right changing the equilibrium allocation. This is known as the Merit Order Effect of RES Cló et al., 2015; Hirth, 2018.

[7] Infra-marginal units are those submitting bids below the SMP, while the last dispatched unit is called the marginal operator/technology since its bid sets the market price.

[8] The Renewable Energy Directive 2009/28/EC, revised in 2018, first established a European framework for the promotion of renewable energy. The Commission presented Europe's new 2030 climate targets, including a proposal for amending the Renewable Energy Directive, on 14 July 2021. It seeks to increase the current target to at least 40% renewable energy sources in the EU's overall energy mix by 2030.

of many endogenous and exogenous factors. In this regard, it is important to recognize that energy price series show many peculiarities and require specific statistical tools for estimation, filtering and testing (Pelagatti and Sen, 2013).

Electricity market conditions tend to appear with regularity within the day (nighttime vs light-time), across days of the week (weekdays vs weekends for business activities) and seasons (summers vs winters), determining typical seasonalities. Demand is indeed the main driver of these prices, and it is largely predictable unless unexpected weather conditions occur. Moreover, in the last fifteen years, non-programmable RES introduced significant short-run variability into market equilibrium and prices.

To take into account the known within-day periodicity of electricity markets, in our analyses we have considered both peak and off-peak day-ahead prices, together with the 24-hour daily average. Peak prices are those determined from 8 AM to 8 PM and off-peak prices cover the remaining hours. As for the countries, following previous investigations, our sample includes Germany, Belgium, France, The Netherlands, Austria, Switzerland, Italy and Spain. The former six markets take part in the EPEX platform, whereas Italy and Spain are included to test the overall EU integration beyond the EPEX platform.[9] Daily electricity prices together with natural gas and oil prices were collected from Datastream from 01/01/2007 to 31/12/2021. Note that our dataset contains 5-day per week observations. The dynamic evolution for a sample of countries is presented in Fig. 6, where we can observe some of the stylized facts discussed in Escribano et al. (2011) among others. Given the frequent spikes, the strong local noise, some volatility clustering and seasonality, it is very hard to discern if a non-stationary long-run component is present in the data. Common sense would suggest that given the non-stationary nature of hydrocarbon fuel prices, also electricity prices should manifest non-stationarity.

The considered countries show quite different generation mixes and evolution during our sample period. Germany increased the share of RES, solar PV and especially wind, lowering production from nuclear and coal, with a stable presence of gas. Electricity generation in France is largely based on nuclear technology, with a small share of wind, hydro and gas, whereas Spain is characterized by an almost equal share of gas, nuclear and wind. Austria mainly relies on hydro and gas, with a low share of RES (wind and biomass), and Belgium shows a mix of nuclear and gas, with a relatively significant introduction of wind and solar PV in the last five years, accounting for almost 20% of the generation. Natural gas and hydro are the most relevant source of electricity production in Italy, with a notable increase in RES in the last ten years at the expense of oil and coal. In the Netherlands, the mix is dominated by gas, while coal has lost most of its share since 2015, thanks to the introduction of RES. Finally, Switzerland has a steady share of nuclear and hydro for the whole sample period. However, even in countries where nuclear is the dominant generation technology, like France, the equilibrium price is often determined by a hydrocarbon-based generator.

To test these EU prices for integration and cointegration, we have adopted a rolling window approach (with an estimation window of $260 \times 3 = 780$ observations) to depict the evolution of the ADF and Johansen statistics for an empirical assessment over the full considered sample. We chose a three-year window as it is not uncommon to carry out statistical analysis of electricity prices on time series of this length. Logarithmic average prices have been seasonally adjusted by using sinusoidal regressors, and then four time series have been compared: the (seasonally-adjusted) raw series, the time series of weekly means, and the Kalman filtered and smoothed series. The ADF regressions contained a constant term and the number of lagged differences was automatically selected using the Akaike Information Criterion. Results were compared with the critical values at 1, 5 and 10% levels, that is, −3.435, −2.863, −2.568 (Cheung and Lai, 1995).

Given the expected dependence of electricity prices on fossil fuels, at least in some countries, and given the three-year window we used in our tests, we thought it could be a good idea to examine the dynamics of gas and oil prices during the same period we consider electricity prices. Fig. 7 depicts gas and oil prices (first row), their moving standard deviation (second row) and inter-quartile range (third row). It is possible to observe that there are two different regimes: one for high and one for low volatility, with the sample of years 2011–2014 characterized by some kind of local stationarity for both gas and oil. We expect this feature to be reflected in the outcomes of the ADF applied to electricity prices at least in those countries whose generation technology is based on these fuels. Therefore, in what follows we undertake our analysis on the whole sample 2010–2021 as well as on the stable years 2011–2014. In the stable subsample, we expect electricity prices to "show more stationary"at least at off-peak hours.

ADF test results for all considered markets are reported in Tables 2–4. In those tables, we report the rejection rates of the ADF tests carried out on each three-year window rolling from the first to the last observation of the considered time samples. The windows were updated daily.

We can first observe that during the steady-price period of gas and oil (years 2011–2014) the rejection rates for the null of integration are higher compared to those computed over the whole sample, consistently for daily average, peak and offpeak periods. We observe different degrees of rejections across countries because of their substantially different composition of the generation mix. In particular, we notice lower rejection rates for countries characterized by important levels of gas generation, like Italy, The Netherlands, Spain and Belgium. Over the full sample, the rejection rates decrease substantially and consistently across all three delivery periods (apart from Italy and Spain).

Considering the performance of filters, both the UCM filter and smoother perform almost equally for mean prices (which are averaged over 24 h, hence with an ex-ante reduced noise). Exceptions are Italy and Spain. When peak and offpeak prices are considered, we generally observe that the smoother has lower rejection rates than the filter with a few exceptions. This holds for peak prices except for France and Switzerland (two countries with high shares of nuclear generation), whereas for off-peak prices exceptions are Belgium and Spain.

In peak hours stationarity is more often detected on raw series (with No filter = 65% vs UCM-filtered = 40% vs UCM-smoothed = 36% in Belgium, and with No filter = 66% vs UCM-filtered = 63% vs UCM-smoothed = 56% in Germany). Similar comments apply to results in the off-peak period, when solar production, demand and electricity prices are at their lower levels. The tests applied to the weekly means reject non-stationarity much less often than all other tests.

Interestingly, we observe that all considered EU prices switch from stationary to non-stationary according to the sample period used in the rolling window and this confirms the contradictory results found in the literature. These empirical findings confirm that results can differ substantially according to the periods and samples considered, especially because changes in the local generation mixes were observed through the years. As investigated by Gianfreda et al. (2016b), the advent of RES has dramatically changed the stochastic nature of electricity prices in the short run, whereas they are still exposed to the international dynamics of fossil fuel prices in the long run. Therefore, these prices need to be carefully handled with appropriate filters according to the target of the analysis and with respect to periods under investigations[10]

[9] Note that we excluded the NordPool market (Denmark, Finland, Norway and Sweden) since its mix is substantially different and mainly based on hydro generation. Moreover, Germany, France and The Netherlands have been largely investigated as examples of a trilateral market.

[10] It is worth noting that solar PV capacity increased in Germany from 11 to 49 GW in the last ten years, while in Spain it was equal to 10 GW in 2019. Moving to wind power capacity, Germany, Spain and France were among the top 10 world countries for additional capacity in 2019 (REN21, 2020).

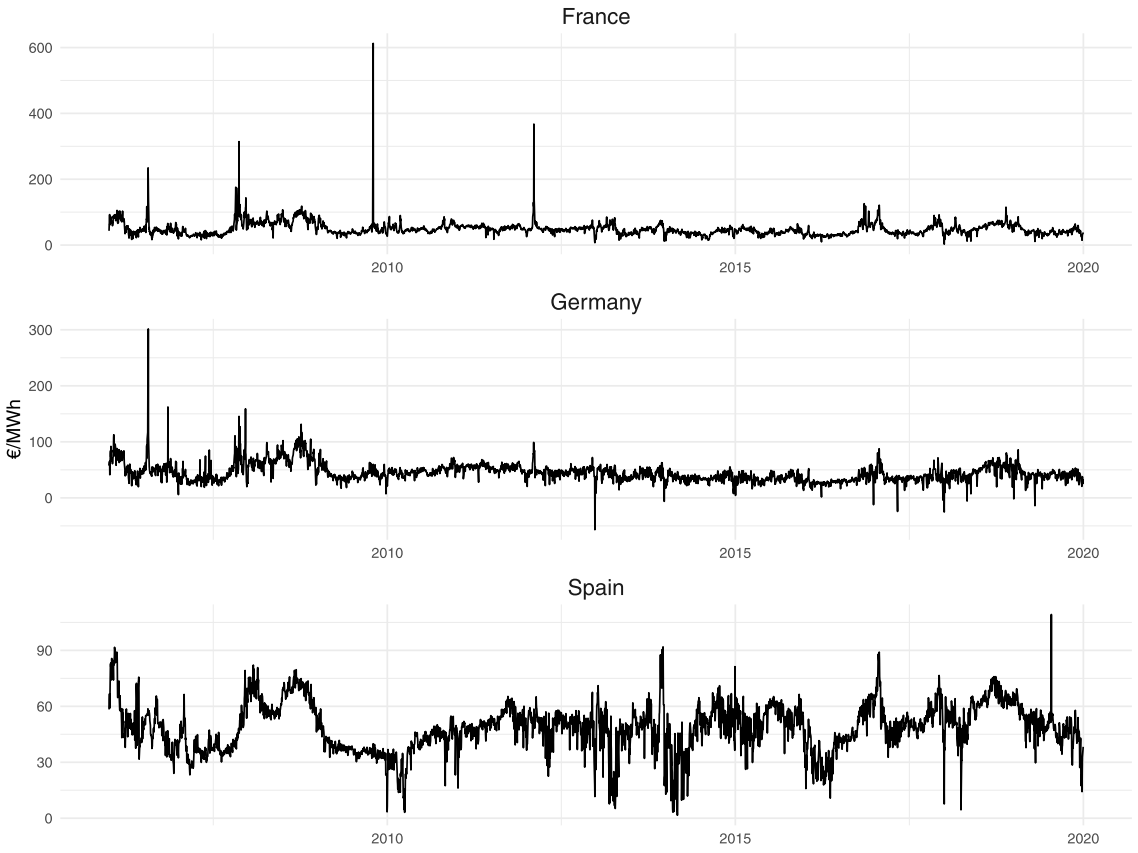


Fig. 6. Day-Ahead Electricity Prices for a Sample of Markets from 2006:01 to 2019:12 (daily averages). Please note that Germany allows for negative prices.

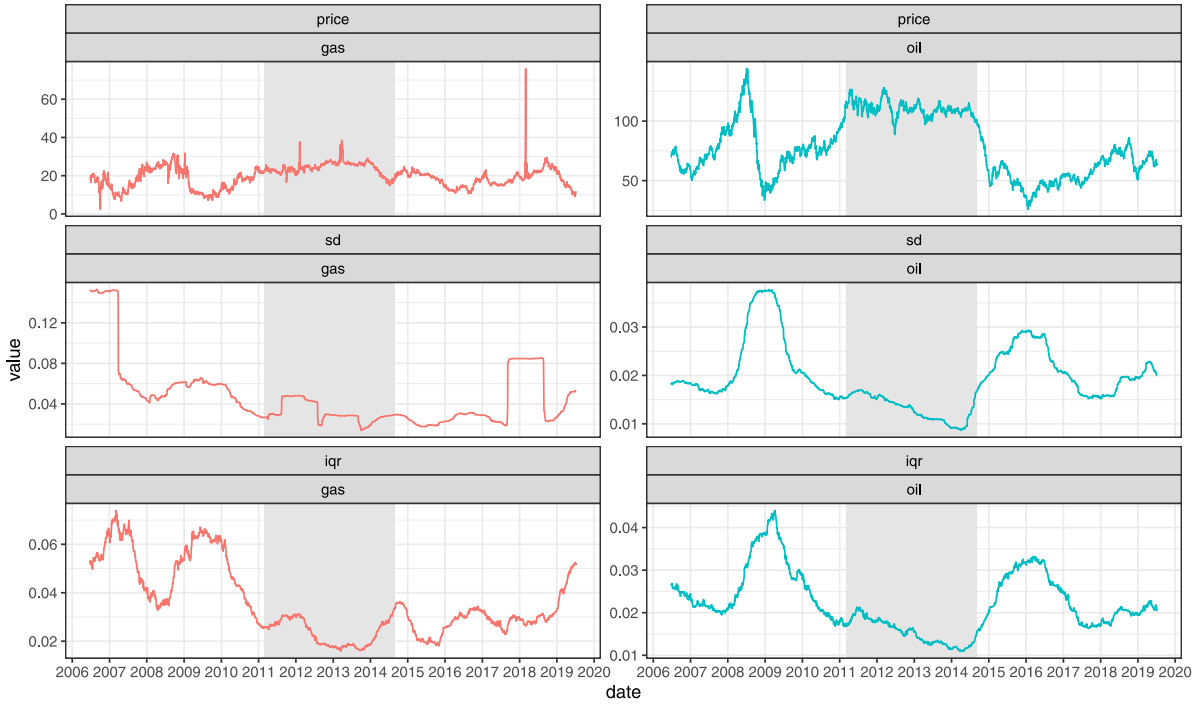


Fig. 7. Oil and gas price and their local volatility. The second and third rows report the centred moving standard deviation (SD) and inter-quartile range (IQR) based on moving samples of one year. The shaded area represents a period of steady prices both for gas and oil ranging from 1st March 2011 to 1st September 2014.

Table 2
Rejection rates at 5% for the ADF tests, for daily average Electricity Prices.

Filter	Austria	Belgium	France	Germany	Italy	Netherlands	Spain	Switzerland
	2010–2021	(N = 3131)						
UCM-filtered	46.89	40.21	55.76	55.61	25.23	24.69	39.83	55.25
UCM-smoothed	45.19	40.59	57.46	51.61	23.73	23.16	37.98	56.72
Weekly means	18.65	34.81	52.60	39.25	21.02	17.47	34.17	43.34
No filter	49.09	55.70	55.16	54.62	30.41	27.72	43.21	49.95
	2011–2014	(N = 914)						
UCM-filtered	57.55	74.29	81.84	68.49	2.08	70.79	43.98	70.46
UCM-smoothed	56.67	73.85	84.90	69.15	1.09	68.27	44.31	70.90
Weekly means	5.69	71.12	79.32	26.59	0.44	57.77	43.22	33.70
No filter	57.22	79.32	86.87	72.98	3.06	66.19	61.49	71.01

Table 3
Rejection rates at 5% for the ADF tests for Peak Electricity Prices.

Filter	Austria	Belgium	France	Germany	Italy	Netherlands	Spain	Switzerland
	2010–2021	(N = 3131)						
UCM-filtered	44.27	39.96	57.59	63.24	34.05	38.52	37.91	61.45
UCM-smoothed	41.84	36.25	59.63	56.40	21.46	25.68	36.31	62.22
Weekly means	16.35	35.23	50.65	37.69	19.39	21.24	33.47	45.32
No filter	51.93	64.52	56.85	66.24	36.03	34.78	37.59	55.80
	2011–2014	(N = 914)						
UCM-filtered	68.27	76.48	84.90	83.04	29.32	80.42	43.76	76.81
UCM-smoothed	64.22	72.98	86.21	80.42	3.83	76.15	44.31	78.99
Weekly means	3.94	72.10	79.21	33.15	3.72	69.58	41.14	44.64
No filter	71.44	86.76	87.09	81.95	21.55	74.40	41.47	76.15

Table 4
Rejection rates at 5% for the ADF tests, for Off-Peak Electricity Prices.

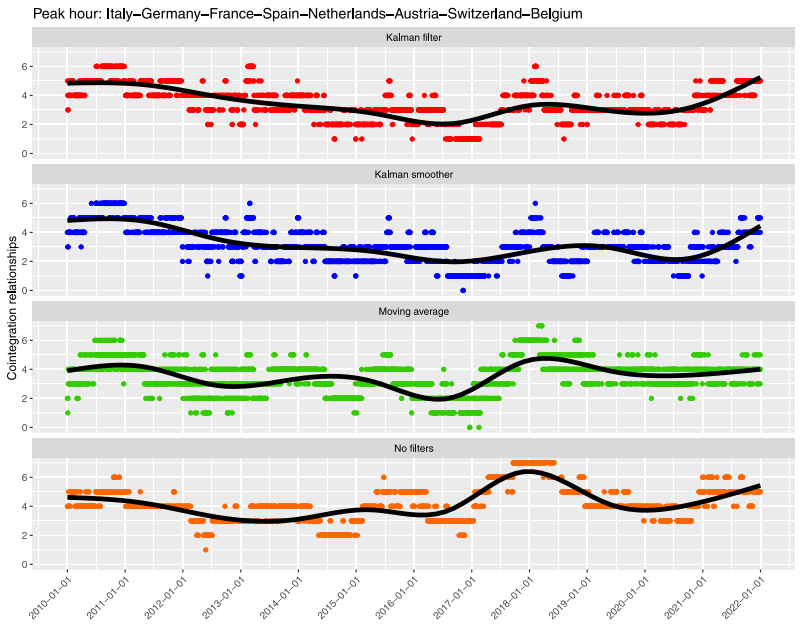
Filter	Austria	Belgium	France	Germany	Italy	Netherlands	Spain	Switzerland
	2010–2021	(N = 3131)						
UCM-filtered	55.16	48.87	66.24	58.67	38.26	43.34	53.82	67.26
UCM-smoothed	48.07	49.47	62.60	32.99	25.46	29.89	55.25	64.01
Weekly means	23.28	45.16	55.89	39.60	27.21	17.37	44.27	50.05
No filter	74.00	55.03	63.65	66.05	44.04	34.88	56.47	47.33
	2011–2014	(N = 914)						
UCM-filtered	70.13	87.20	96.28	75.60	16.41	57.55	53.50	66.19
UCM-smoothed	55.69	87.75	93.54	45.19	4.81	51.09	46.28	64.22
Weekly means	9.85	86.11	88.29	35.89	14.22	49.78	42.56	51.64
No filter	99.67	89.72	95.73	83.59	32.82	64.00	45.08	68.05

Trying to draw some conclusions on the "mean-reversion feature" of electricity prices based on our filtering-before-testing approach, we can affirm that unit roots are dominant at peak hours in Italy, Spain, Netherlands, Belgium and Austria, while France, Switzerland and Germany are found mean-reverting most of the times. At off-peak hours only Italy and Netherlands still show a clear absence of mean-reversion, remaining Spain and Austria more ambiguous among the two hypotheses. Finally, during the time of steady gas and oil prices, all countries excluding Italy and Spain, tend to show mean-reverting electricity prices. An interpretation of this fact is that, while in Austria, Belgium, and the Netherlands the non-stationarity of electricity prices is only due to their dependence on hydrocarbon prices, in Italy and Spain there are other non-stationary factors driving electricity prices. This said, given the substantial number of times ADF tests do not reject the null of integration, the idea that all price time series could be non-stationary seems quite reasonable.

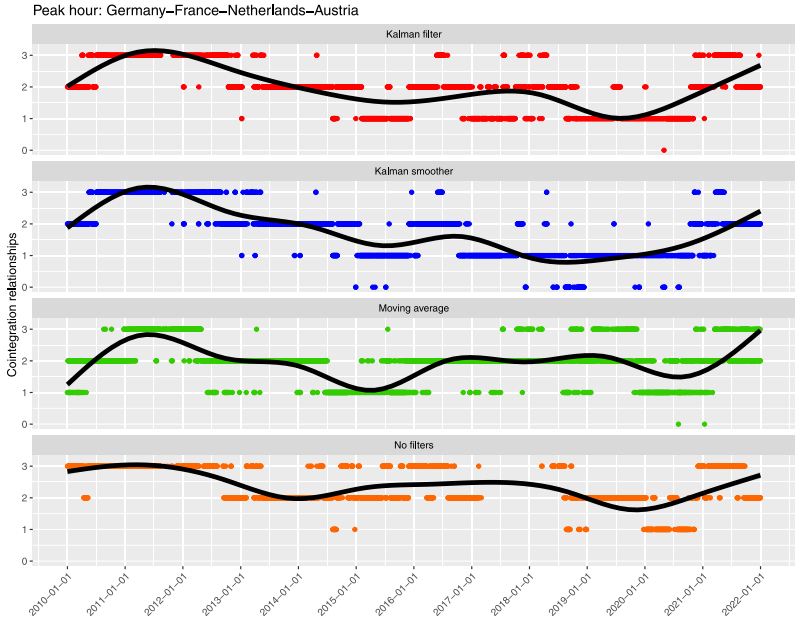
Turning to Johansen's cointegration tests, we expect the uncertainty of the conclusions to be even stronger than for ADF tests. Indeed, if we are uncertain about a set of time series showing a trending behaviour or being mean-reverting, we are even more uncertain about the number of eventual common trends. The number of cointegrating relationships identified by Johansen's procedure ($\alpha = 0.05$) on 3-year rolling windows[11] is depicted in Figs. 8 (peak) 9 (off-peak) for all considered countries and for a smaller group of central European countries, in which only Germany, France, The Netherlands and Austria are included (as in Bosco et al., 2010). Note that the EEX Phelix auction daily prices were determined for delivery of electricity in both Germany and Austria until 30th September 2018. Thereafter, a zone market splitting occurred between these two zones and the EEX Phelix prices were used for the delivery of electricity in Germany and Luxembourg. Hence, it is normal to expect these markets to be strictly connected by their participation in the EPEXspot. Recalling that simulation results show that the UCM-based filter produces the most reliable estimates of the cointegrating rank, we mainly look at these tests for the two groups of markets.

Overall, on peak hours (Fig. 8), the UCM-based filter tends to select 4 cointegrating equations (CEs) for the eight-country sample and 2 CEs for the four-country sample. However, in both samples, at the

[11] The window size is 3 years of 260 daily observations, and the first estimation window considers the years 2007–2009. At every new daily observation a new test is carried out.



(a)



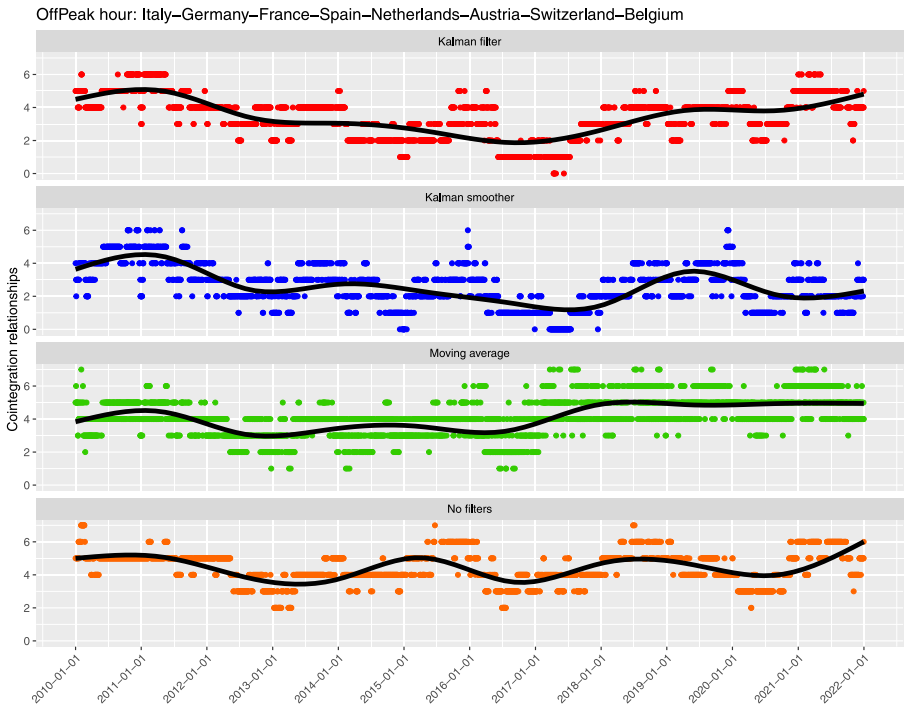
(b)

Fig. 8. Number of cointegrating relationships identified in the rolling sample using average daily prices during *peak* period by the Kalman filter (in the first row), Kalman smoother (in the second row), five-day moving average (in the third row) and unfiltered data (in the last row) for all countries (in the first column) and for a smaller sample (on the right column).

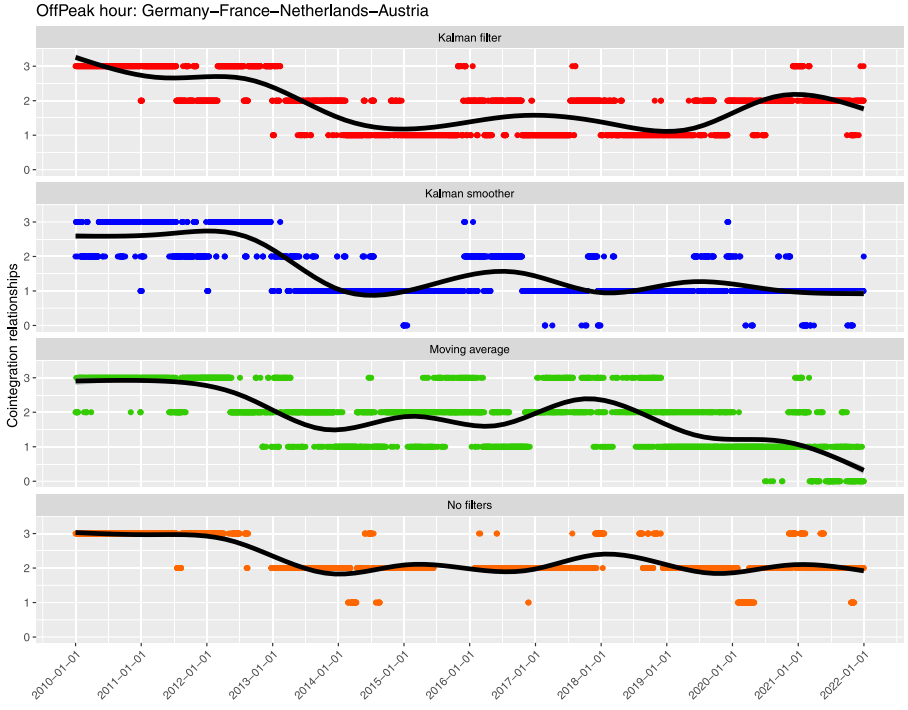
beginning and at the end of the considered time spans the number of cointegrating relations seems to increase to 5 and 3 CEs, respectively. The UCM-based smoother leads to a similar conclusion with a tendency to select a number of CEs slightly smaller (as expected from the simulation exercise). The procedure applied to weekly means produces conclusions that are similar to those of the UCM filter, however, its variability (the dispersion of the green dots about the bold line in the figures) is higher. The test applied to raw data tends to find a higher number of CEs.

If we consider off-peak hours (Fig. 9), in the eight-country sample the number of CEs is similar to that of peak hours, with an overall tendency to select a slightly higher number of CEs more often. Turning to the four-country sample, in the first part of the period (2010–2013) all tests tend to select 3 CEs. After 2013 the number of CEs tend to reduce to 2 when the procedure is applied to raw data and UCM-smoother data. The tests applied to UCM-filtered data and weekly means are more uncertain between 1 and 2 CEs. Considering the lower noise-to-level ratio for off-peak prices, 2 CEs seems to be a reasonable choice. The larger number of CEs at the beginning of our sample could be an effect of the observed low volatility of gas and oil prices over that period.

As a conclusion to this application, we can state that ADF results seem to be more reasonable if one considers testing after filtering, as suggested by our Monte Carlo experiments. However, electricity markets are constantly evolving in many aspects (technologies, interconnections, regulation, incentives) and this makes any joint analysis



(a)



(b)

Fig. 9. Number of cointegrating relationships identified in the rolling sample using average daily prices during *offpeak* period by the Kalman filter (in the first row), Kalman smoother (in the second row), five-day moving average (in the third row) and unfiltered data (in the last row) for all countries (in the first column) and for a smaller sample (on the right column).

intrinsically unstable over time, as our cointegration analyses reveal. Thus, the robustification of ADF and Johansen's test thorough filtering is certainly a good practice when working on electricity prices, however, all changes in technologies, interconnections, regulations and incentives should be monitored and, possibly, modelled, when analysing prices of a pool of markets over a relatively long period.

4.2. Financial prices and pairs trading

As mentioned in the introductory section, an important field of application of integration and cointegration tests in finance is that of statistical arbitrage and, in particular, pairs trading. If a linear combination of the prices of two or more financial assets (stocks, currencies,

Algorithm 1 Statistical arbitrage algorithm for cointegrated price pairs

$V_1 \leftarrow 100$, $p_1 \leftarrow 0$, $t \leftarrow 1$
while (t < n) **do**
  if $(p_t = 0)$ & $(z_t > \theta)$ **then**
    $p_{t+1} \leftarrow -1$
  else if $(p_t == -1)$ & $(z_t \leq 0)$ **then**
    $p_{t+1} \leftarrow 0$
  else if $(p_t == 0)$ & $(z_t < -\theta)$ **then**
    $p_{t+1} \leftarrow 1$
  else if $(p_t == 1)$ & $(z_t \geq 0)$ **then**
    $p_{t+1} \leftarrow 0$
  else
    $p_{t+1} \leftarrow p_t$
  end if
  $V_{t+1} \leftarrow V_t \left[0.5 \left(\frac{y_{t+1}}{y_t}\right)^{p_{t+1}} + 0.5 \left(\frac{x_t}{x_{t-1}}\right)^{p_{t+1}}\right]$
  $t \leftarrow t + 1$
end while

commodities, etc.) is found mean-reverting and persistent and, thus, predictable, then it is possible to apply statistical arbitrage strategies. The same holds if the ratio of the prices of two financial activities is mean-reverting and persistent and, thus, predictable.

To understand if filtering before (co)integration testing brings benefits also when working with financial data, we consider 96 constituents of the S&P100 index for the 10 years 14th March 2008 to 14th March 2018. Financial theory suggests that there should be no persistent cointegration relationship among stock prices. Indeed, there is no economic reason for the value of two companies to evolve proportionally and the efficiency of financial markets should rule out any arbitrage opportunity. Thus, we can assume that the truth is that no cointegration is present among pairs of stock log prices. However, to keep an agnostic approach, we look at cointegrated pairs of log prices by applying cointegration tests on the first half of the sample and we check if the same result is found on the second half of the sample. Moreover, using the second half of the sample, on those stock pairs that in the first half of the sample have been found cointegrated, we apply a trading strategy that produces positive returns on pairs of persistently cointegrated log prices.

In summary, this is how we proceed.

1. We split the ten-year sample into two 5-year sub-samples.
2. We consider all pairs of the 96 stocks (i.e., all 4560 pairs).
3. In the first experiment, using the first half of the sample, we use the ADF to test on each pair if the difference of log prices is integrated or stationary. In the second experiment, using the first half of the sample, we use Johansen's test to determine if the number of cointegrating relations in the time series pairs of log prices is zero or one. In both experiments, we use the tests on raw and filtered data.
4. We repeat the same tests on the second half of the sample and asses how often the decisions based on the tests are homogeneous in the two samples.
5. On those pairs that were selected as cointegrated in the first half of the sample, we apply the trading strategy on the second half of the sample and compare its returns with the same strategy applied to non-cointegrated pairs. If most of the pairs that were found cointegrated are really cointegrated, then the returns generated on these pairs should be larger than those found on the rest of the pairs.

The statistical arbitrage strategy we set up is very simple. Let y_t and x_t be the prices of the two stocks in the pair, and let $z_t = \log(y_t) - \beta \log(x_t)$ be the cointegrating relations among the two log prices (in the ADF case $\beta = 1$, in the Johansen's case β is estimated). Algorithm 1 describes the trading strategy, for a given positive threshold value θ.

In simple words, we sell y_t and buy x_t if z_t is above a positive threshold, we buy y_t and sell x_t if z_t is below a negative threshold and stay out of the investment if z_t is between the two thresholds. The algorithm has been tested on simulated time series from vector error correction models estimated on pairs that were found cointegrated obtaining annualized returns above 5%.

As mentioned above, there are no reasons for the log prices of liquid stocks such as S&P100 constituents to be cointegrated and, thus, by commenting on the results of our experiment we assume that the truth is *no cointegration among all pairs*. Under this assumption, some 1% of the tests should find cointegrating relations as we fixed the nominal significance level at 1%. However, since we do not want to rule out the possibility of statistical arbitrage at once, we compare the outcome of the trading strategy applied to the cointegrated and not cointegrated pairs according to the ADF and Johansen test.

Table 5 summarizes the results of this experiment. Using the ADF applied to the difference between pairs of log prices, the test on filtered series finds only a slightly smaller number of cointegrated pairs. If we look at Johansen's procedure applied to log-price pairs, the tests on filtered time series find half the number of cointegrating relations with respect to the same test on raw data. If the theory is correct and no statistical arbitrage relations are present, the tests based on filtered time series are preferable even though they find cointegrating relations in more than 1% of the cases (nominal size). Furthermore, coherence between the test outcomes in the two sub-samples is higher when the UCM filter or smoother is used before testing.

The annualized returns of the strategy applied to time series pairs in the second sub-sample are approximately the same for pairs that were found cointegrated and pairs that were found not cointegrated. This confirms that the pairs trading strategy works identically on cointegrated and not cointegrated pairs. Notice that, to assess if the trading strategy worked on cointegrated pairs, we estimated a cointegrated vector error correction model (CVECM) on pairs that were found cointegrated by the tests and by simulating from the estimated CVECM and applying the test strategy on each simulated path, we found much higher average returns (above 5% for the Johansen's case and above 20% for the ADF case).

5. Conclusions

For many years, applied econometricians have known that the popular ADF and Johansen's tests have serious size problems when time series are observed with noise (measurement error) or when the data-generating process has a MA component with a root that almost cancels with the unit root of the difference operator. However, no universal simple-to-use solutions have been proposed. In this paper, we test three different linear filtering strategies that the authors have informally been

Table 5
Results of the application of the ADF and Johansen tests to all the pairs of the S&P100 constituents and of the trading strategy.

	RAW	FLT	SMO	MEAN
ADF test				
Pairs found cointegrated in sample 1 (%)	4.2	4.0	4.0	3.1
Coherence rate between samples (%)	94.8	95.0	95.0	95.5
Mean return of strategy on cointegrated pairs (%)	0.2	0.2	0.2	−0.2
Mean return of strategy on not cointegrated pairs (%)	0.2	0.2	0.2	0.2
p-value for t-test for equality of means (%)	83.8	97.4	86.4	0.2
Johansen test				
Pairs found cointegrated in sample 1 (%)	12.1	6.8	5.1	6.1
Coherence rate between samples (%)	79.2	84.6	86.3	76.1
Mean return of strategy on cointegrated pairs (%)	0.1	0.0	−0.1	0.1
Mean return of strategy on not cointegrated pairs (%)	0.1	0.1	0.1	0.1
p-value for t-test for equality of means (%)	37.0	29.1	2.4	67.4

using on electricity price data for years. Theoretical considerations suggest that reducing the frequency of observation by taking means every m-observations can be successful, as well as applying the test statistics to the filtered or smoothed level of an unobserved component model whose trend is a random walk observed with noise (and, possibly, seasonal components).

Monte Carlo simulations confirm that filtering before testing assures an appropriate size of the ADF test also in the presence of very strong measurement error. In particular, the filter based on UCM-smoothing keeps the size of the test exactly at its nominal level regardless of the amount of noise. Robustness comes at the cost of some power: for frequency reduction, the power deficiency is independent of the quantity of noise, while for UCM-based filtering and smoothing the power loss increases with the noise-to-signal ratio.

Our simulations suggest a clear winner when Johansen's sequential procedure is used to estimate the number of cointegrating relations. UCM-based filtering provides the most accurate estimates for both data-generating processes considered in our experiments. Of course, this does not guarantee that the same result holds for *all* data-generating processes, but there is no way to test the whole universe of DGPs.

A very important feature of UCM-based filtering is that filters are data-driven and when no observational noise is present, the time series are left virtually unchanged and the properties of integration and cointegration tests are identical to those applied to raw data.

We use the proposed filtering techniques in two real applications: one to European electricity prices and one to stock prices. In the first case, the ADF based on filtered data seems to provide more reliable results by revealing unit roots more often than the ADF on raw data. Despite some authors holding electricity prices as mean-reverting, common sense suggests that electricity prices should have a long-run component following the price of hydrocarbon fuels, at least in those countries where hydrocarbon-based generation prevails. At the moment of writing, the world is witnessing a period of extremely high gas and oil prices because of the war in Ukraine and other international matters and the relationship between electricity prices with gas and oil prices is more than evident. Results on the number of cointegrating relations among European electricity prices are less clear to interpret. On one side, this larger uncertainty was expected since uncertainty about the presence of unit roots in a pool of time series implies uncertainty about the number of common trends. Furthermore, the continuous evolution of electricity markets due to technological change, green policies, grid and interconnection expansions, etc. makes the relation among prices in different countries time-varying.

The application of cointegration testing finalized to the selection of pairs of stock prices that should allow for statistical arbitrage strategies (pairs trading) enhances the higher level of reliability of Johansen's test when applied to filtered time series. The test after filtering approaches reveals a halved number of cointegrated pairs and out-of-sample evaluations confirm that pair trading is not possible among S&P100 constituents. Practitioners looking for this kind of trading strategy should benefit from using our filters before cointegration testing.

Data and codes

All results presented in this paper can be reproduced using the R software (R Core Team, 2018). All scripts and the data of the financial application (electricity prices were provided by Datastream and we have no permission to republish them) are available at the following GitHub web page: https://github.com/PaoloMaranzano/AG_PM_MP_LVP_FilterCointStat.

Declaration of competing interest

None

Data availability

We share code and data on GitHub (https://github.com/PaoloMaranzano/AG_PM_MP_LVP_FilterCointStat). We cannot share electricity prices because got from a commercial platform that does not allow data sharing.

Acknowledgments

The authors thank two anonymous referees, the Associate Editor and one of the Editors for their comments and suggestions which helped to significantly improve the paper.

We greatly acknowledge the DEMS Data Science Lab for supporting this work by providing computational resources.

Appendix. Proofs

Derivation of Eq. (2). Even though the result is well-known in the time series literature, for self-completeness we derive Eq. (2). Since there is a one-to-one mapping of a MA process with its autocovariance function (ACF), we just need to show that the ACF of $\eta_t+\varepsilon_t-\varepsilon_{t-1}$ is that of a MA(1) process, say $\zeta_t-\theta\zeta_{t-1}$, and equate their values. The following table displays the quantities of interest (the symbol 2^+ stands for any integer equal to or larger than 2).

Equating the first two autocovariances, we get the system

$$\begin{cases} \sigma^2(1+\theta^2) = \sigma_\eta^2 + 2\sigma_\varepsilon^2 \\ -\theta\sigma^2 = -\sigma^2\varepsilon \end{cases}$$

whose solution is

$$\theta = 1 + \frac{\lambda - \sqrt{\lambda^2+4\lambda}}{2}, \qquad \sigma^2 = \frac{\sigma_\varepsilon^2}{\theta},$$

where $\lambda := \sigma_\eta^2/\sigma_\varepsilon^2$. □

ACF	Generic MA(1)	$\eta_t + \varepsilon_t - \varepsilon_{t-1}$
$\gamma(0)$	$\sigma^2(1+\theta^2)$	$\mathsf{E}\left((\eta_t+\varepsilon_t-\varepsilon_t)^2\right) = \sigma_\eta^2 + 2\sigma_\varepsilon^2$
$\gamma(1)$	$-\theta\sigma^2$	$\mathsf{E}\left((\eta_t+\varepsilon_t-\varepsilon_{t-1})(\eta_{t-1}+\varepsilon_{t-1}-\varepsilon_{t-2})\right) = -\sigma_\varepsilon^2$
$\gamma(2^+)$	0	$\mathsf{E}\left((\eta_t+\varepsilon_t-\varepsilon_{t-1})(\eta_{t-2^+}+\varepsilon_{t-2^+}-\varepsilon_{t-3^+})\right) = 0$

Proof of Theorem 1 and Corollary 1.1. Let y_t be a random walk plus noise with $t \in \{1, 2, 3, \dots, \}$ with given starting value x_0:

$$y_t = x_t + \varepsilon_t, \qquad \varepsilon_t \sim \mathsf{WN}(\sigma_\varepsilon^2)$$
$$x_t = x_{t-1} + \eta_t, \qquad \eta_t \sim \mathsf{WN}(\sigma_\eta^2).$$

We can rewrite the random walk as

$$x_t = x_{t-m} + \sum_{j=0}^{m-1} \eta_{t-j},$$

and taking the average over m subsequent observations, we obtain

$$\frac{1}{m}\sum_{i=0}^{m-1} x_{t-i} = \frac{1}{m}\sum_{i=0}^{m-1} x_{t-m} + \frac{1}{m}\sum_{i=0}^{m-1}\sum_{j=0}^{m-1} \eta_{t-j-i},$$

which we can restate as

$$\bar{x}_t = \bar{x}_{t-m} + \bar{\eta}_t,$$

where $\bar{x}_t = \frac{1}{m}\sum_{i=0}^{n-1} x_{t-i}$ and

$$\bar{\eta}_t = \frac{1}{m}\sum_{i=0}^{n-1}\sum_{j=0}^{m-1} \eta_{t-j-i} = \frac{1}{m}\Big[\, \eta_t + 2\eta_{t-1} + 3\eta_{t-2} + \cdots + m\eta_{t-m+1} + (m-1)\eta_{t-m} + (m-2)\eta_{t-m-1} + \cdots + 2\eta_{t-2m+1} + \eta_{t-2m+2} \,\Big].$$

Now, let us consider the process for $t \in \mathcal{T} := \{m, 2\ m, 3\ m, \dots\}$. Using the above formulae, it is simple to derive the autocovariance function (ACF) of $\bar{\eta}_t$ over the set $\mathcal{T}$:

$$\mathsf{E}\left(\bar{\eta}_t \bar{\eta}_{t-km}\right) = \begin{cases} \left[\frac{(m-1)(2m-1)}{3m} + 1\right]\sigma_\eta^2 & \text{for } k = 0, \\ \frac{(m-1)(m+1)}{6m}\sigma_\eta^2 & \text{for } |k| = 1, \\ 0 & \text{for } |k| = 2, 3, \dots, \end{cases}$$

which is the ACF of a MA(1) process. Consequently, the process

$$\bar{y}_t - \bar{y}_{t-m} = \bar{\eta}_t + \bar{\varepsilon}_t - \bar{\varepsilon}_{t-1},$$

with $\bar{\varepsilon}_t := \frac{1}{m}\sum_{i=0}^{m-1} \varepsilon_{t-i}$, is also MA(1) with ACF

$$\begin{cases} \left[\frac{(m-1)(2m-1)}{3m} + 1\right]\sigma_\eta^2 + \frac{2}{m}\sigma_\varepsilon^2 & \text{for } k = 0, \\ \frac{(m-1)(m+1)}{6m}\sigma_\eta^2 + \frac{1}{m}\sigma_\varepsilon^2 & \text{for } |k| = 1, \\ 0 & \text{for } |k| = 2, 3, \dots. \end{cases}$$

Therefore, its order-1 autocorrelation over $\mathcal{T}$ is

$$\rho := \frac{\frac{(m-1)(m+1)}{6m}\sigma_\eta^2 + \frac{1}{m}\sigma_\varepsilon^2}{\left[\frac{(m-1)(2m-1)}{3m} + 1\right]\sigma_\eta^2 + \frac{2}{m}\sigma_\varepsilon^2} = \frac{\frac{(m-1)(m+1)}{6m}\lambda - \frac{1}{m}}{\left[\frac{(m-1)(2m-1)}{3m} + 1\right]\lambda + \frac{2}{m}},$$

where $\lambda := \sigma_\eta^2/\sigma_\varepsilon^2$ is the signal-to-noise ratio.

We can conclude that over the set $\mathcal{T}$, $\bar{y}_t$ is ARIMA(0, 1, 1) and the MA coefficient can be easily obtained from ρ using the formula in Corollary 1.1. □

Proof of Theorem 2. By assumption y_t is defined by the ARIMA(0, 1, 1) process in Eq. (2). Thus, indicating with L the lag operator (i.e., $Lx_t = x_{t-1}$),

$$\tilde{x}_t = \gamma y_t + (1-\gamma)\tilde{x}_{t-1}$$
$$\big(1 - (1-\gamma)L\big)\tilde{x}_t = \gamma y_t$$
$$\big(1 - (1-\gamma)L\big)(1-L)\tilde{x}_t = \gamma(1-\theta L)\zeta_t.$$

It is straightforward to check that $\theta = 1 - \gamma$, which implies that the operator $\big(1 - (1-\gamma)L\big)$ annihilates $(1 - \theta L)$ and

$$(1-L)\tilde{x}_t = \gamma\zeta_t,$$

defines a random walk.

To obtain the approximate smoother $\hat{x}_t$ one has to pass a backward autoregressive filter on $\tilde{x}_t$. The second-order properties of the resulting process are the same if the filter is run backwards or forward and, thus, the process $\hat{x}_t$ has the same autocorrelation function as

$$\big(1 - (1-\gamma)L\big)(1-L)\tilde{x}_t = \gamma\zeta_t,$$

which defines an $ARIMA(1, 1, 0)$ process with autoregressive coefficient $\phi = 1 - \gamma$. □

References

Aatola, P., Ollikainen, M., Toppinen, A., 2013. Impact of the carbon price on the integrating European electricity market. Energy Policy 61, 1236–1251.

Ahn, D.-H., Boudoukh, J., Richardson, M., Whitelaw, R.F., 2015. Partial adjustment or stale prices? Implications from stock index and futures return autocorrelations. Rev. Financ. Stud. 15 (2), 655–689.

Argentiero, A., Atalla, T., Bigerna, S., Micheli, S., Polinori, P., 2017. Comparing renewable energy policies in EU-15, U.S. and China: A Bayesian DSGE model. Energy J. 38, 1944–9089.

Bosco, B., Parisio, L., Pelagatti, M., Baldi, F., 2010. Long-run relations in European electricity prices. J. Appl. Econometrics 25 (5), 805–832.

Bunn, D.W., Gianfreda, A., 2010. Integration and shock transmissions across European electricity forward markets. Energy Econ. 32 (2), 278–291.

Caldeira, J.F., Moura, G.V., 2013. Selection of a portfolio of pairs based on cointegration: A statistical arbitrage strategy. Braz. Rev. Finance 11 (1), 49–80.

Chelani, A.B., 2013. Statistical characteristics of ambient PM2. 5 concentration at a traffic site in Delhi: source identification using persistence analysis and nonparametric wind regression. Aerosol Air Qual. Res. 13 (6), 1768–1778.

Cheung, Y.-W., Lai, K.S., 1995. Lag order and critical values of the augmented Dickey–Fuller test. J. Bus. Econom. Statist. 13 (3), 277–280.

Christiano, L.J., Vigfusson, R.J., 2003. Maximum likelihood in the frequency domain: The importance of time-to-plan. J. Monetary Econ. 50 (4), 789–815.

Cló, S., Cataldi, A., Zoppoli, P., 2015. The merit-order effect in the Italian power market: The impact of solar and wind generation on national wholesale electricity prices. Energy Policy 77, 79–88.

de Menezes, L.M., Houllier, M.A., 2016. Reassessing the integration of European electricity markets: a fractional cointegration analysis. Energy Econ. 53, 132–150, Energy Markets.

de Menezes, L.M., Houllier, M.A., Tamvakis, M., 2016. Time-varying convergence in European electricity spot markets and their association with carbon and fuel prices. Energy Policy 88, 613–627.

Diebold, F.X., Ohanian, L.E., Berkowitz, J., 1998. Dynamic equilibrium economies: A framework for comparing models and data. Rev. Econom. Stud. 65 (3), 433—451.

Durbin, J., Koopman, S.J., 2001. Time Series Analysis By State Space Methods. Oxford University Press.

Elliott, G., Rothenberg, T.J., Stock, J.H., 1996. Efficient tests for an autoregressive unit root. Econometrica 64 (4), 813–836.

Escribano, A., Peña, J.I., Villaplana, P., 2011. Modelling electricity prices: International evidence. Oxford Bulletin of Economics and Statistics 73 (5), 622–650.

Fernandes, M.C., Dias, J.C., Nunes, J.P.V., 2021. Modeling energy prices under energy transition: A novel stochastic-copula approach. Econ. Model. 105, 105671.

Fezzi, C., Bunn, D., 2009. Structural interactions of European carbon trading and energy prices. J. Energy Mark. 2 (4), 53–69.

Fischer, A.M., 1990. Cointegration and I(0) measurement error bias. Econom. Lett. 34 (3), 255–259.

Galbraith, J.W., Zinde-Walsh, V., 1999. On the distributions of augmented Dickey–Fuller statistics in processes with moving average components. J. Econometrics 93 (1), 25–47.

Geman, H., Roncoroni, A., 2006. Understanding the fine structure of electricity prices. J. Bus. 79 (3), 1225–1261.

Gianfreda, A., Parisio, L., Pelagatti, M., 2016a. The impact of RES in the Italian day-ahead and balancing markets. Energy J. 37 (SI2), 161–184.

Gianfreda, A., Parisio, L., Pelagatti, M., 2016b. Revisiting long-run relations in power markets with high RES penetration. Energy Policy 94, 432–445.

Gianfreda, A., Parisio, L., Pelagatti, M., 2018. A review of balancing costs in Italy before and after RES introduction. Renew. Sustain. Energy Rev. 91, 549–563.

Gianfreda, A., Parisio, L., Pelagatti, M., 2019. The RES-induced switching effect across fossil fuels: An analysis of day-ahead and balancing prices. Energy J. 40, 365–386.

Girma, P.B., Paulson, A.S., 1999. Risk arbitrage opportunities in petroleum futures spreads. J. Futures Mark. 19 (8), 931–955.

Glosten, L.R., Milgrom, P.R., 1985. Bid, ask and transaction prices in a specialist market with heterogeneously informed traders. J. Financ. Econ. 14 (1), 71–100.

Gonzalo, J., Pitarakis, J.-Y., 1998. On the exact moments of asymptotic distributions in an unstable AR(1) with dependent errors. Internat. Econom. Rev. 39 (1), 71–88.

Gugler, K., Haxhimusa, A., Liebensteiner, M., 2018. Integration of European electricity markets: Evidence from spot prices. Energy J. 39 (SI2), 97–116.

Habimana, O., Månsson, K., Sjölander, P., 2021. A wavelet-based approach for Johansen's likelihood ratio test for cointegration in the presence of measurement errors: An application to CO2 emissions and real GDP data. Commun. Stat.: Case Stud. Data Anal. Appl. 7 (2), 128–145.
Hannan, E.J., 1969. The estimation of mixed moving average autoregressive systems. Biometrika 56 (3), 579–593.
Harvey, A.C., 1989. Forecasting Structural Time Series and the Kalman Filter. Cambridge University Press.
Harvey, A., 2006. Chapter 7 forecasting with unobserved components time series models. In: Elliott, G., Granger, C.W.J., Timmermann, A. (Eds.), Handbook of Economic Forecasting, Vol. 1. Elsevier, pp. 327–412.
Hassler, U., Kuzin, V., 2009. Cointegration analysis under measurement errors. In: Binner, J.M., Edgerton, D.L., Elger, T. (Eds.), Measurement Error: Consequences, Applications and Solutions. In: Advances in Econometrics, vol. 24, Emerald Group Publishing Limited, pp. 131–150.
Haug, A.A., 1996. Tests for cointegration a Monte Carlo comparison. J. Econometrics 71 (1), 89–115.
Hirth, L., 2018. What caused the drop in European electricity prices? A factor decomposition analysis. Energy J. 39 (1), 143–157.
Hong, H., Ahn, S.K., Cho, S., 2016. Analysis of cointegrated models with measurement errors. J. Stat. Comput. Simul. 86 (3), 623–639.
Horowitz, J., Barakat, S., 1979. Statistical analysis of the maximum concentration of an air pollutant: Effects of autocorrelation and non-stationarity. Atmos. Environ. (1967) 13 (6), 811–818.
Huisman, R., Kiliç, M., 2013. A history of European electricity day-ahead prices. Appl. Econ. 45 (18), 2683–2693.
Huisman, R., Mahieu, R., 2003. Regime jumps in electricity prices. Energy Econ. 25 (5), 425–434.
Johansen, S., 1991. Estimation and hypothesis testing of cointegration vectors in Gaussian vector autoregressive models. Econometrica 59 (6), 1551–1580.
Keles, D., Genoese, M., Möst, D., Fichtner, W., 2012. Comparison of extended mean-reversion and time series models for electricity spot price simulation considering negative prices. Energy Econ. 34 (4), 1012–1032.
Koopman, S.J., Harvey, A., 2003. Computing observation weights for signal extraction and filtering. J. Econom. Dynam. Control 27 (7), 1317–1333.
Krauss, C., 2017. Statistical arbitrage pairs trading strategies: Review and outlook. J. Econ. Surv. 31 (2), 513–545.
Kyle, A.S., 1985. Continuous auctions and insider trading. Econometrica 53 (6), 1315–1335.
Lee, C.-K., 2002. Multifractal characteristics in air pollutant concentration time series. Water Air Soil Pollut. 135 (1–4), 389–409.
Miao, G.J., 2014. High frequency and dynamic pairs trading based on statistical arbitrage using a two-stage correlation and cointegration approach. Int. J. Econ. Finance 6 (3), 96–110.
Mudelsee, M., 2019. Trend analysis of climate time series: A review of methods. Earth-Sci. Rev. 190, 310–322.
Mudelsee, M., 2020. Statistical Analysis of Climate Extremes. Cambridge University Press.
Ng, C.N., Yan, T.L., 2004. Recursive estimation of model parameters with sharp discontinuity in non-stationary air quality data. Environ. Model. Softw. 19 (1), 19–25.
Pelagatti, M.M., 2015. Time Series Modelling with Unobserved Components. Chapman and Hall / CRC.
Pelagatti, M.M., Sen, P.K., 2013. Rank tests for short memory stationarity. J. Econometrics 172 (1), 90–105.
Phillips, P.C.B., Perron, P., 1988. Testing for a unit root in time series regression. Biometrika 75 (2), 335–346.
Proietti, T., 2008. Band spectral estimation for signal extraction. Econ. Model. 25 (1), 54–69.
R Core Team, 2018. R: A Language and Environment for Statistical Computing. R Foundation for Statistical Computing, Vienna, Austria.
REN21, 2020. Renewables 2020 - Global Status Report. REN21 Secretariat, Paris, ISBN: 978-3-948393-00-7.
Ricky Rambharat, B., Brockwell, A.E., Seppi, D.J., 2005. A threshold autoregressive model for wholesale electricity prices. J. R. Stat. Soc. Ser. C. Appl. Stat. 54 (2), 287–299.
Said, S.E., Dickey, D.A., 1984. Testing for unit roots in autoregressive-moving average models of unknown order. Biometrika 71 (3), 599–607.
Scholes, M., Williams, J., 1977. Estimating betas from nonsynchronous data. J. Financ. Econ. 5 (3), 309–327.
Schwert, G.W., 1989. Tests for unit roots: A Monte Carlo investigation. J. Bus. Econom. Statist. 7 (2), 147–159.
Simon, D.P., 1999. The soybean crush spread: Empirical evidence and trading strategies. J. Futures Mark. 19 (3), 271–289.
Varotsos, C., Ondov, J., Efstathiou, M., 2005. Scaling properties of air pollution in Athens, Greece and Baltimore, Maryland. Atmos. Environ. 39 (22), 4041–4047.
Wahab, M., Cohn, R., Lashgari, M., 1994. The gold-silver spread: Integration, cointegration, predictability, and ex-ante arbitrage. J. Futures Mark. 14 (6), 709–756.
Watson, M.W., 1993. Measures of fit for calibrated models. J. Polit. Econ. 101 (6), 1011–1041.
Weron, R., Bierbrauer, M., Trück, S., 2004. Modeling electricity prices: jump diffusion and regime switching. Physica A 336 (1), 39–48, Proceedings of the XVIII Max Born Symposium "Statistical Physics outside Physics".
Windsor, H., Toumi, R., 2001. Scaling and persistence of UK pollution. Atmos. Environ. 35 (27), 4545–4556.
Zachmann, G., 2008. Electricity wholesale market prices in Europe: Convergence? Energy Econ. 30 (4), 1659–1671.