Predicting bubble bursts in oil prices using mixed causal-noncausal models

Alain Hecq and Elisa Voisin
Maastricht University

November, 2019

Abstract

This paper investigates oil price series using mixed causal-noncausal autoregressive (MAR) models, namely dynamic processes that depend not only on their lags but also on their leads. MAR models have been successfully implemented on commodity prices as they allow to generate nonlinear features such as speculative bubbles. We estimate the probabilities that bubbles in oil price series burst once the series enter an explosive phase. To do so we first evaluate how to adequately detrend nonstationary oil price series while preserving the bubble patterns observed in the raw data. The impact of different filters on the identification of MAR models as well as on forecasting bubble events is investigated using Monte Carlo simulations. We illustrate our findings on WTI and Brent monthly series.

Keywords: noncausal models, forecasting, predictive densities, bubbles, simulations-based forecasts, detrending, Hodrick-Prescott filter.

JEL. C22, C53

1Corresponding author : Elisa Voisin, Maastricht University, Department of Quantitative Economics, School of Business and Economics, P.O.box 616, 6200 MD, Maastricht, The Netherlands. Email: a.hecq@maastrichtuniversity.nl.
1 Introduction

In a review paper, Frey, Manera, Markandya, and Scarpa (2009) list three categories of econometric approaches for investigating oil prices. These are (i) time series models exploiting the statistical properties of the data, (ii) financial models based on the relationship between spot and future prices and (iii) structural models describing how economic factors and the behavior of economic agents affect the future values of oil prices. We consider the first approach with nonetheless a flavor of the third one. This paper aims at forecasting bubbles in Brent and WTI oil price series using the recent literature on mixed causal-noncausal autoregressive models (hereafter \( MAR \)), that is, time series processes with lags but also leads components and non-Gaussian errors. This new specification can model locally explosive episodes in a strictly stationary setting. It can therefore capture nonlinear features such as bubbles (persistent increase followed by a sudden crash), often observed in commodities prices, while standard linear autoregressive models (e.g. \( ARMA \) models) cannot do so. \( MAR \) models have successfully been implemented on several commodity price series (see inter alia Hecq and Voisin, 2019, Hecq, Issler, and Telg, 2019, Fries and Zakoïan, 2019, Gouriéroux and Zakoïan, 2017, Cubadda, Hecq, and Telg, 2019, Lof and Nyberg, 2017, Karapanagiotidis, 2014). However, oil prices are challenging time series to be forecasted and modeled, and contrarily to many other commodities, they are clearly nonstationary (see Figure 5). Consequently a trending time varying fundamental part must be extracted before estimating \( MAR \) models.

Similarly to Gouriéroux and Zakoïan (2013), our goal when introducing a lead component in prices is not to provide an economic justification for the existence of a rational bubble. However, the link with a present value model between prices and dividends (Campbell and Shiller, 1987) can enrich the discussion and it also explains the difficulties to find economic fundamentals for oil prices. This motivates our choice to use proxies such as technical methods to extract the bubble component. Let us indeed consider a general

\(^2\)An alternative strategy to ours is to consider autoregressive processes with breaks in coefficients. Indeed, autoregressive processes with successively unit roots, explosive and stable stationary episodes are also able to capture locally explosive episodes. See among many others Phillips, Wu, and Yu (2011) and the survey papers by Homm and Breitung (2012) or Bertelsen (2019). Su, Li, Chang, and Loboț (2017) find six bubbles during 1986-2016 in oil price deviations from fundamentals using this framework. We think however that our approach is more suited for forecasting price movements as well as to compute probabilities with which a bubble bursts.
model (see Diba and Grossman, 1988) in which the real current stock price $P_t$ is linked to the present value of next period’s expected stock price $P_{t+1}$, dividend payments $D_{t+1}$ and an unobserved variable $u_{t+1}$,

$$P_t = \frac{1}{1+r} \mathbb{E}_t [P_{t+1} + \alpha D_{t+1} + u_{t+1}],$$  \hspace{1cm} (1)

with $\mathbb{E}_t$ the conditional expectation given the information set known at time $t$. The discount factor is $\frac{1}{1+r}$ with $r$ being a time-invariant interest rate. The general solution of (1) is (e.g. Diba and Grossman, 1988)

$$P_t = \sum_{i=1}^{\infty} \left( \frac{1}{1+r} \right)^i \mathbb{E}_t [\alpha D_{t+i} + u_{t+i}] + B_t,$$ \hspace{1cm} (2)

$$P_t = P_t^F + B_t,$$ \hspace{1cm} (3)

where the actual price deviates from its fundamental value $P_t^F$ by the amount of the rational bubble $B_t$.

However, unlike for equity prices, measuring commodities fundamentals is not as straightforward (Brooks, Prokopczuk, and Wu, 2015). Pindyck (1993) considers the convenience yield, that is, a premium associated with holding an underlying product instead of derivative securities or contracts. It typically increases when costs associated with physical storage are low. Yet, not only is the convenience yield not easy to measure but there also are other factors driving each of the demand and supply side of crude oil: the level of stocks, the temperature, economic activity, geopolitical consideration, etc. Fan and Xu (2011) choose the weekly Baltic Exchange Dry Index (BDI) as a proxy for the supply and demand fundamentals and Miller and Ratti (2009) consider stock market prices. Additional recent contributions in this vein are, among many others, Degiannakis and Filis (2018), Funk (2018), Miao, Ramchander, Wang, and Yang (2017), Wang, Liu, and Wu (2017). This emphasizes the difficulty to assess the fundamental value of crude oil prices.

Obviously, defining a fundamental value series and then presuming that any deviation from it represents speculation (or a rational bubble if one believes in models analogous to (2)) can give misleading results if the fundamental is misspecified. Consequently, investigating the impact of different detrending filters on the identification of bubbles is the first contribution of this paper. We do not construct new indicators for the fundamentals. We treat them
as unobservable components whose behaviors will be captured by a flexible trend function that, and this is the important point to notice, leaves the bubble pattern in the series while making it stationary. Similarly to what Canova (1998) does for business cycles, we investigate the extent to which the identification of MAR models and consequently the identification of bubbles are sensitive to different filters. We then study the consequences on the probability that oil price bubbles burst after applying different detrending methods.

This being said, the presence of a bubble in (2) or (3) introduces an explosive component into $P_t$. The key point, if one wants to link economic models with MAR processes, is to notice that for (2) to satisfy (1) there must exist a law of motion of the bubble (e.g. Cuthbertson and Nitzsche, 2005, Diba and Grossman, 1988) such that

$$ B_t = \frac{1}{1+r} \mathbb{E}_t [B_{t+1}] . $$

(4)

Defining the rational forecasts errors $\xi_{t+1} = \mathbb{E}_t [B_{t+1}] - B_{t+1}$, Engsted (2006) and Engsted and Nielsen (2012) focus on the existence of an explosive root in the autoregressive process $B_{t+1} = (1+r)B_t + \xi_{t+1}$, with $r > 0$. Interestingly, (2) does not contradict Campbell and Shiller (1987) results on cointegration. Indeed, Engsted and Nielsen (2012) and Nielsen (2006) show that $P_t$ and $D_t$ can both be I(1) and cointegrated with potentially bubbles in the deviation from the cointegrating vector. For the purpose of our paper, (4) also means that, in reverse time, $B_t$ has a stationary noncausal representation with

$$ B_t = \frac{1}{1+r} B_{t+1} + \varepsilon_t , $$

(5)

with $\varepsilon_t = \frac{\xi_{t+1}}{1+r}$. All these results give rise to several implications for MAR models among which: (i) commodity I(1) prices should cointegrate with their fundamental variables and (ii) prices $P_t$, after being deflated by the fundamental component $P^F_t$, result in a noncausal bubble in (5). This paper focuses on the latter point and leaves the first point for further research.

The rest of this paper is as follows. Section 2 describes mixed causal-noncausal models and explains the different structures considered for the fundamental trend, leaving the locally explosive components in the cycle. The filtering employed are data-driven as we do not intend to construct fundamental series based on theory. Section 3 describes how to forecast
MAR processes. In Section 4, the impact of the different detrending filters on model identifications is investigated using a Monte Carlo exercise, based on trends estimated in oil prices series. We study the identification of the models but also the magnitude of the coefficients estimated as they are the main drivers of the predictions. Section 5 analyses the impact of these filters on the WTI and the Brent oil price series and shows how this affects the probabilities that oil price bubbles burst. The notion of common bubbles is introduced, namely a situation in which bubbles are present in individual series but not in a combination of them. The presence of common bubble features would simplify the forecasting procedures as several series sharing such bubble co-movements could be predicted with the information obtained from one of those variables. Section 6 concludes.

2 Mixed causal-noncausal models and trend fundamentals

2.1 The model

$MAR(r,s)$ denotes dynamic processes that depend on their $r$ lags as for usual autoregressive processes but also on their $s$ leads in the following multiplicative form

$$
\Phi(L)\Psi(L^{-1})y_t = \varepsilon_t,
$$

with $L$ the backward operator, i.e., $Ly_t = y_{t-1}$ gives lags and $L^{-1}y_t = y_{t+1}$ produces leads. When $\Psi(L^{-1}) = (1 - \psi_1L^{-1} - ... - \psi_sL^{-s}) = 1$, namely when $\psi_1 = ... = \psi_s = 0$, the univariate process $y_t$ is a purely causal autoregressive process, denoted $MAR(r,0)$ or simply $AR(r)$ model, $\Phi(L)y_t = \varepsilon_t$. Reciprocally, the process is a purely noncausal $MAR(0,s)$ model $\Psi(L^{-1})y_t = \varepsilon_t$, when $\phi_1 = ... = \phi_r = 0$ in $\Phi(L) = (1 - \phi_1L - ... - \phi_rL^r)$. The roots of both the causal and noncausal polynomials are assumed to lie outside the unit circle, that is $\Phi(z) = 0$ and $\Psi(z) = 0$ for $|z| > 1$ respectively. These conditions imply that the series $y_t$ admits a two-sided moving average representation $y_t = \sum_{j=-\infty}^{\infty} \gamma_j \varepsilon_{t-j}$, such that $\gamma_j = 0$ for all $j < 0$ implies a purely causal process $y_t$ (with respect to $\varepsilon_t$) and a purely noncausal model when $\gamma_j = 0$ for all $j > 0$ (Lanne and Saikkonen, 2011). Error terms $\varepsilon_t$ are assumed iid (and not only weak white noise) non-Gaussian to ensure the identifiability of the causal and the noncausal parts (Breid, Davis, Li, and Rosenblatt, 1991).
Figure 1 shows a purely causal (top) and a purely noncausal (bottom) trajectories induced by identical Student’s $t(2)$-distributed errors, both with coefficient 0.8 and 200 observations. For the purely causal process, a shock is unforeseeable and affects the series only once it happened, inducing a large jump in the series. On the other hand, for purely noncausal processes, a shock impacts the process ahead of time, mirroring the purely causal trajectory. Indeed, we see that the series already reacts to a positive shock by increasing until a sudden crash, creating bubble patterns. This anticipative aspect is widely observed in financial and economics time series. The detrended Brent crude oil prices as shown in Figure 6 noticeably exhibit such features, the most apparent episode being the 2008 financial crisis. A combination of causal and noncausal dynamics consequently creates some asymmetry around a shock, varying with the magnitude of the respective coefficients.

$MAR(1,0)$ with $\phi = 0.8$

$MAR(0,1)$ with $\psi = 0.8$

Figure 1: Purely causal (top) and noncausal (bottom) trajectories with Student’s $t(2)$-distributed errors

The advantage with oil prices is that they already underwent bubbles, and those previous locally explosive episodes will help identifying $MAR$ models.
In the case where series are for the first time following a long and abnormal increase, an explosive process is difficultly distinguished from a stationary locally explosive one.

### 2.2 Filtering the trend in the data

The requirement of \( y_t \) being stationarity for both lag and lead polynomials gave rise to different strategies to transform nonstationary series to stationary ones. Hecq et al. (2019) and Cubadda et al. (2019) assume\(^3\) that their commodity price series are \( I(1) \) and work with the returns \( \Delta y_t \). However, this operation eliminates most of the locally explosive behaviors and the transformed series consist of many spikes instead.

In this paper, we capture the trending behavior of the observed series denoted \( \tilde{y}_t \) in different ways using the general form

\[
\tilde{y}_t = f_t + y_t,
\]

where

\[
\Phi(L)\Psi(L^{-1})y_t = \epsilon_t.
\]

In this framework, \( \tilde{y}_t \) is the (potentially nonstationary) observed series and \( f_t \) a generic trend function. The deviation of \( \tilde{y}_t \) from its trend is an \( MAR(r, s) \) process. Several authors, although sometimes not explicitly, use this decomposition. Cavaliere, Nielsen, and Rahbek (2018) opt for the choice of a particular time period with no trend and hence use only an intercept \( f_t = \mu \). Hencic and Gouriéroux (2015) detrend \( \tilde{y}_t \) using a polynomial trend function of order three. Note (see Section 5) that a polynomial trend of order four or six seems to best capture the trending pattern of the monthly oil prices series considered in this analysis. Hecq and Voisin (2019) use the Hodrick-Prescott filter (HP) before detecting bubbles in Nickel monthly prices. In summary we could for instance consider several choices among the following deterministic trends,

\[
\begin{align*}
  f_t^{(1)} &= \mu, \\
  f_t^{(2)} &= \mu + \beta D_t, & \text{with } D_t = 1 \text{ when } t \geq t_{\text{break}} \text{ and } 0 \text{ otherwise,} \\
  f_t^{(3)} &= \alpha_0 + \alpha_1 t + \ldots + \alpha_k t^k, & \text{with } k \text{ some positive integer and } t = 1, 2, \ldots, T.
\end{align*}
\]

More complex trends, constructed as a combination of the aforementioned examples could also be considered, such as (multiple) breaks in trends for

---

\(^3\)The non linear features of the data make unit root tests doubtful.
instance.

The HP filter on the other hand extracts the trend process $f_t$ via the following minimization problem,

$$
\min_{\{f_t\}_{t=1}^T} \left\{ \sum_{t=1}^T \hat{y}_t^2 + \lambda \sum_{t=3}^T \left[ (f_t - f_{t-1}) - (f_{t-1} - f_{t-2}) \right]^2 \right\}.
$$

The parameter $\lambda$ penalizes the variability in the filtered trend and therefore the larger its value, the smoother is the trend component. The limit results of this minimization problem as $\lambda$ approaches infinity is hence a linear trend. As De Jong and Sakarya (2016) indicate, based on the first order condition, the solution of this minimization problem can be expressed in closed-form as such,

$$
f_t = \left( \lambda L^{-2} - 4\lambda L^{-1} + (1 + 6\lambda) - 4\lambda L + \lambda L^2 \right)^{-1} \hat{y}_t.
$$

This equation however only holds for $t = 3, \ldots, T - 2$. They show that overall the trend component is a weighted average of the initial series, with both lags and leads, with a number of adjustments terms at the endpoints of the sample. It is now commonly accepted to use $\lambda = 1600$ for quarterly data (Hodrick and Prescott, 1997). For other frequencies, rules of thumbs consist in adjusting the parameter to the frequency relative to quarterly data,

$$
\lambda = \left( \frac{\text{number of observations per year}}{4} \right)^i \times 1600,
$$

with either $i = 2$ (Backus and Kehoe, 1992) or $i = 4$ (Ravn and Uhlig, 2002), yielding respectively a penalizing parameter of 14400 and 129600 for monthly series. Most criticisms of the HP filter concern its application on series with complex stochastic and deterministic trends. Phillips and Shi (2019) propose an adaptation of the filter improving its accuracy for such series. In our case, the proposed boosting algorithm absorbs too much dynamics and captures the bubble in the trend component.

Note that while we may lack interpretation when the assumed fundamentals are approximated by a polynomial trend of order 6 or extracted using an HP filter, this does not significantly affect the interpretation of our results. Indeed, we investigate probabilities of events, and mostly the probabilities of turning points during explosive episodes. Hence, we are not interested in...
the exact value of a forecast but rather in the direction with respect to the deviation from the fundamental trend (e.g. probabilities of a further increase or of a crash). This stems from the difficulty to construct and evaluate fundamental measures and we expect that if the trend is not accurately identified, the impact will not be as important as for point predictions of prices for instance.

3 Predictions

The focus of this paper is on predicting probabilities of turning points, for example the probabilities of a crash or of entering a bubble. For such inquiries, density forecasts are therefore more adequate than point forecasts. However, the anticipative aspect of MAR models complicates their use for predictions. An MAR\((r,s)\) model can also be expressed as a causal AR model where \(y_t\) depends on its own past and on \(u_t\),

\[
\Phi(L)y_t = u_t, \quad (7)
\]

where \(u_t\) is the purely noncausal component of the errors, depending on its own future and on the contemporaneous error term

\[
\Psi(L^{-1})u_t = \epsilon_t. \quad (8)
\]

If the model is correctly identified and the parameters consistently estimated, it is therefore sufficient to forecast the purely noncausal process \(u_t\) to forecast the variable of interest \(y_t\). However, only a few specifications admit a closed-form conditional density.

One example is the purely noncausal process, \(u_t = \psi u_{t+1} + \epsilon_t\), with standard Cauchy-distributed errors for which the conditional density function admits the following closed-form (Gouriéroux and Zakoian, 2013),

\[
l(u_{T+1}^*, \ldots, u_{T+h}^*|u_T) = \frac{1}{\pi^h} \left( \frac{1}{1 + (u_T - \psi u_{T+1}^*)^2} \cdots \frac{1}{1 + (u_{T+h-1}^* - \psi u_{T+h}^*)^2} \right) \\
\times \frac{1 + (1 - \psi)^2 u_T^2}{1 + (1 - \psi)^2 (u_{T+h}^*)^2};
\]

where \(h\) is the forecast horizon and \(l\) denotes the density function related to the process \(u_t\). The asterisk denotes the future points to be predicted.
Figure 2 depicts the evolution of the one-step ahead predictive density for a purely noncausal process with a lead coefficient of 0.8 and standard Cauchy-distributed errors. For levels close to the median of the series ($u_T = 0.79$), the one-step ahead predictive density is unimodal, showing that probabilities to stay around the last value is high (first graph of Figure 2). As the series departs from central values (we consider quantiles 0.85 and 0.975 corresponding to $u_T = 9.81$ and $u_T = 63.53$ respectively), the predictive density starts to split and becomes bimodal. The two resulting modes are 0 and the natural rate of increase $0.8^{-1}u_T$, corroborating the results of Fries and Zakoïan (2019). The corresponding probabilities for each of these events are respectively $(1 - 0.8)$ and 0.8, implying that as the variable follows an explosive episode and tends to infinity, one-step ahead probabilities of a crash will remain constant. With this respect, the intuition is close to the switching regimes models of Hamilton (1989), in which the transition probabilities from an explosive regime to a stable one are constant.

Furthermore, the assumption of other fat-tail distributions, such as Student’s $t$, can lead to the absence of closed-form expressions for the conditional moments and densities. Two approximations methods have been developed to estimate these predictive densities, for any distribution, also allowing for a larger lead order. The first method, based on simulations, was developed by Lanne, Luoto, and Saikkonen (2012). The second approach uses the information carried by the sample and was developed by Gourieroux and Jasiak (2016). For a detailed description and guidance in using those approximations methods, see Hecq and Voisin (2019). We focus on processes with a unique lead as this is what we estimate in the WTI and Brent series in Section 5.
3.1 Simulations-based approach

The purely noncausal component of the errors, \( u \), assumed with one lead, can be expressed as an infinite sum of future error terms. Lanne et al. (2012) base their methodology on the fact that there exists an integer \( M \) large enough so that any future point of the noncausal component can be approximated by the following finite sum,

\[
u^*_T + h \approx \sum_{i=0}^{M-h} \psi^i \epsilon^*_{T+h+i}, \tag{9}\]

for any forecast horizon \( h \geq 1 \).

Let \( \epsilon^*_{T+1} = (\epsilon^*_{T+1}, \ldots, \epsilon^*_{T+M}) \), with \( 1 \leq j \leq N \), be the \( j \)-th simulated series of \( M \) independent errors, randomly drawn from the chosen distribution of the process with estimated parameters (whose pdf is denoted by \( g \)). We are interested in the conditional cumulative probabilities,

\[
\mathbb{P}(y^*_T + h \leq x | \mathcal{F}_T) = \mathbb{E}_T \left[ 1(y^*_T + h \leq x) \right] \\
\approx \mathbb{E}_T \left[ 1(y^*_T + h - \Phi y + \sum_{i=0}^{h-1} \psi^i \epsilon^*_{T+M-i-j} \leq x) \right], \tag{10}\]

with \( M \) the truncation parameter and where \( y^*_T + h \) is replaced by its approximation using recursive substitution of its companion form with

\[
y_T = \begin{bmatrix} y_T \\ y_{T-1} \\ \vdots \\ y_{T-r+1} \end{bmatrix}, \quad \Phi = \begin{bmatrix} \phi_1 & \phi_2 & \ldots & \ldots & \phi_r \\ 1 & 0 & \ldots & \ldots & 0 \\ 0 & 1 & 0 & \ldots & 0 \\ \vdots & \vdots & \ddots & \ddots & \vdots \\ 0 & \ldots & 0 & 1 & 0 \end{bmatrix} (r \times r) \quad \text{and} \quad \iota = \begin{bmatrix} 1 \\ 0 \\ \vdots \\ 0 \end{bmatrix} (r \times 1).
\]

Given the information set known at time \( T \), the indicator function in (10) is only a function of the \( M \) future errors, \( \epsilon^*_T \). Let us denote this indicator function by \( q(\epsilon^*_T) \). Assuming that the number of simulations \( N \) and the truncation parameter \( M \) are large enough, the conditional cumulative probabilities of \( MAR(r,1) \) processes can be approximated as follows (Lanne et al., 2012).
\[ P\left(y_{T+h}^* \leq x \mid \mathcal{F}_T\right) \approx E_T \left[q\left(\varepsilon_+^*\right)\right] \approx \frac{N^{-1} \sum_{j=1}^{N} q\left(\varepsilon_+^*\right) g\left(u_T - \sum_{i=1}^{M} \psi_i \varepsilon_{T+i}^*\right)}{N^{-1} \sum_{j=1}^{N} g\left(u_T - \sum_{i=1}^{M} \psi_i \varepsilon_{T+i}^*\right)}, \] (11)

By computing its value for all possible \( x \) we can obtain the whole conditional cdf of \( y_{T+h}^* \).

Hecq and Voisin (2019) results show that with Cauchy-distributed errors, this approach is a good estimator of theoretical probabilities but are significantly sensitive to the number of simulations \( N \) chosen. For Student’s \( t \) distributions however, results cannot be compared to theoretical ones, but as the number of simulations goes to infinity, the derived densities converge to a unique function. Moreover, analogously to theoretical probabilities, once the series has significantly departed from its central values and diverges, the probabilities of a crash at a given horizon tend to a constant.

3.2 Sample-based approach

As an alternative to using simulations, Gourieroux and Jasiak (2016) employ all past observed values of the noncausal process. The predictive density function of a purely noncausal process with one lead is approximated as follows,

\[ l(u_{T+1}^*, \ldots, u_{T+h}^* \mid u_T) \approx g(u_T - \psi u_{T+1}^*) \ldots g(u_{T+h-1}^* - \psi u_{T+h}^*) \frac{\sum_{i=1}^{T} g(u_{T+h}^* - \psi w_i)}{\sum_{i=1}^{T} g(u_T - \psi w_i)}, \] (12)

where \( g \) is the pdf of the assumed errors distribution.

With this method, the predicted probabilities are a combination of theoretical probabilities and probabilities induced by past events. Results are therefore case-specific and are based on a sort of learning mechanism (Hecq and Voisin, 2019). If this method is used when errors are Cauchy distributed, results can be compared to the theoretical predictive distribution to evaluate the influence of past behaviors on the obtained probabilities. However, if the errors follow a Student’s \( t \) distribution for instance, results cannot be compared to theoretical probabilities as no closed-form expressions exists.
In such cases an approximation of theoretical results can be derived using the simulations-based approach presented above to gauge how much of the probabilities are induced by the underlying process and by past behaviors.

For values around the median of the series, both methods yield identical results. Discrepancies widen as the level of the series increases. Additionally, the larger the lead coefficient, the more the sample-based method tend to overestimate probabilities of a crash (Hecq and Voisin, 2019). That is, for low lead coefficients, they on average yield very similar results, even for explosive episodes, while for large lead coefficients probabilities induced by the two methods can be considerably different. Overall, both methods depend on the whole sample since they both depend on the estimated coefficients. Hence a wrong detrending would affect both methods. Overestimating the lead coefficient for instance would imply lower probabilities of a crash. For Cauchy distributed processes, one-step ahead probabilities of a crash tend to \((1 - \psi)\) during explosive episodes. Thus, identifying a model with a lead coefficient of 0.9 instead of 0.7 for instance would induce a 20% difference in the theoretical probabilities. The sample-based probabilities could be even more distorted based on past behaviors, or on the contrary past behaviors could potentially alleviate the impact on the wrong detrending, but this is case-specific. This is why it is important to investigate the effects of various detrending methods on model identification and on the estimation of the dynamics. Note that formulas for higher lead orders can be found in the respective articles of Lanne et al. (2012) and Gourieroux and Jasiak (2016).

4 Monte Carlo analysis - Effects of detrending

The aim of this section is to analyze the effect of potentially wrongly detrending a series, both on the identification of the \(MAR\) model and on the subsequent predictions performed with the resulting model. We base this analysis on stylized facts observed in oil prices series.

We simulate 5,000 trajectories for 12 distinct data generating processes (hereafter \(dgp\)), composed of a trend and a cycle. All \(dgp\)’s are generated by Student’s \(t\)-distributed errors with 2 degrees of freedom, a value commonly observed in financial time series, and with 400 observations. For the cycles, we analyze purely noncausal processes with a lead coefficient of 0.8, purely causal processes with a lag coefficient of 0.6 and mixed causal-
noncausal processes with a lag coefficient of 0.6 and a lead coefficient of 0.8. The heavy-tailed distribution generates extreme values, inducing bubbles in processes with noncausal components. We are interested in mostly forward looking processes characterized by long lasting bubbles hence the choice of coefficients. We consider three different deterministic trends: a linear trend with breaks (denoted breaks) and two trend polynomials up to orders 4 and 6 (denoted respectively $\tau^4$ and $\tau^6$ for simplicity). The trends were estimated on the monthly WTI crude oil prices series between 1986 and 2019. Figure 3 depicts the three mentioned trends to which purely causal, noncausal and mixed causal-noncausal trajectories are added. This results in 12 sets of 5 000 trajectories ($\tilde{y}_t$’s), among which a third is generated by purely causal processes, a third by purely noncausal processes and third by mixed causal-noncausal processes, with ($\tilde{y}_t = f_t + y_t$) and without ($\tilde{y}_t = y_t$) trends.

![Figure 3: Trends estimated on WTI oil prices series](image)

Four detrending methods are employed for each trajectories, with the general form $\tilde{y}_t = \hat{f}_t + \hat{y}_t$. Estimated polynomial trends of orders 4 and 6 and HP filters with $\lambda = 14 000$ and $\lambda = 129 600$ are applied (respectively denoted $t^4$, $t^6$, $HP_1$ and $HP_2$). Table 1 shows the average mean square errors (MSE) between the true cycle of $\tilde{y}_t$, namely $y_t$ and the one obtained after detrending, $\hat{y}_t$ over the 5 000 replications of each $dgp$’s and for the four detrending approaches,

$$MSE_{k,d} = \frac{1}{5 000} \sum_{i=1}^{5 000} \frac{1}{400} \sum_{t=1}^{400} (\tilde{y}_t^{(k,i)} - \hat{y}_t^{(k,i,d)})^2,$$
where $k$ indicates the $dgp$, $d$ the detrending method used, and $i$ the $i$-th replication with $1 \leq i \leq 5000$.

The MSEs are minimized when the correct polynomial trend is employed or when the lower order is employed (4 in this case) in the absence of trend in the $dgp$. However, underestimating the order of the polynomial trend leads to significantly larger discrepancies. Distortions between the true cycle and the detrended series are greater for mixed causal-noncausal processes than for purely causal or noncausal processes. Furthermore, in the presence of noncausal dynamics the HP filter with $\lambda = 14400$ ($HP_1$) distorts more the series than $HP_2$. Hence, we can expect that a low penalizing parameter in the HP filter mostly captures some of the noncausal dynamics. However, $HP_1$ distorts the less the cycles to which the linear trend with breaks was added. It is the method managing to mimic the best this non smooth trend due to this flexibility induced by its low penalizing parameter.

Table 1: Average Mean Squared Errors between true cycles and detrended series

| $DGP$                      | Detrended with |
|----------------------------|----------------|
|                            | $t^4$ | $t^6$ | $HP_1$ | $HP_2$ |
| $MAR(0,1)$ + no trend      | 5.23  | 7.61  | 11.44  | 7.15   |
| $MAR(0,1)$ + $\tau^4$     | 4.55  | 6.03  | 9.62   | 7.50   |
| $MAR(0,1)$ + $\tau^6$     | 62.42 | 6.38  | 11.35  | 11.26  |
| $MAR(0,1)$ + breaks       | 79.02 | 55.78 | 31.84  | 47.65  |
| $MAR(1,1)$ + no trend      | 22.69 | 31.05 | 48.58  | 30.81  |
| $MAR(1,1)$ + $\tau^4$     | 42.74 | 65.18 | 91.42  | 57.60  |
| $MAR(1,1)$ + $\tau^6$     | 85.91 | 39.57 | 61.21  | 43.02  |
| $MAR(1,1)$ + breaks       | 101.48| 86.93 | 78.36  | 77.18  |
| $MAR(1,0)$ + no trend      | 1.20  | 1.64  | 2.55   | 1.58   |
| $MAR(1,0)$ + $\tau^4$     | 0.96  | 1.34  | 2.14   | 2.70   |
| $MAR(1,0)$ + $\tau^6$     | 59.24 | 2.45  | 4.21   | 6.73   |
| $MAR(1,0)$ + breaks       | 76.42 | 52.19 | 26.30  | 44.10  |

Notes: Are reported the average MSEs over 5000 trajectories with sample size $T = 400$. $HP_1$ corresponds to the HP filter with $\lambda = 14400$ and $HP_2$ to the HP filter with $\lambda = 129600$. 

14
4.1 Effects of detrending on model identification

To investigate the impact of detrending on the dynamics of autoregressive processes, we perform MAR estimations on the raw and detrended series of all dgp’s. The estimation of MAR models first consists in estimating the pseudo causal lag order. Since the autocorrelation structure of mixed or purely causal and noncausal processes are identical, we can estimate the order of autocorrelation \( p \) with information criteria on OLS estimations of standard autoregressive models. Once this order \( p \) is estimated, identification of the lag and lead orders (\( r \) and \( s \) respectively) is performed by maximum likelihood among all \( MAR(r,s) \) models such that \( r + s = p \) (Lanne and Saikkonen, 2011). We do so using the MARX package in R (Hecq, Lieb, and Telg, 2017).

Table 2 presents the proportions (in percentages) of mis-identification of the model for each of the 12 dgp’s, based on the detrending methods, with a maximum pseudo causal lag order of 4.\(^4\) Proportions of a wrongly identified lag order in the first step of the estimation using BIC are reported (\( p \neq 1 \) and \( p \neq 2 \)), as well as the proportions of wrongly identified MAR models (either lag or lead order, or both). We also report the frequency with which no noncausal dynamics is identified (\( s = 0 \)). For the purely causal processes we only report in the last column (\( s > 0 \)), i.e. the frequency with which spurious noncausal dynamics is detected.

Let us first focus on the models with noncausal dynamics (the \( MAR(0,1) \) and \( MAR(1,1) \) dgp’s) for which we report the proportions of wrongly estimated pseudo causal lag order in the first step of the estimation. We can see that \( HP_1 \) under-performs relative to the other approaches. Around twice as many lag orders are wrongly estimated, on average, as for the other methods, with a maximum of 22.84\% for the \( MAR(1,1) \) processes with breaks in the linear trend. However this complex trend significantly affects estimations as can can be seen in the last rows where even on raw data the model is correctly identified (for dgp’s with noncausal dynamics). This can be explained by the construction of the trend, mimicking somehow a bubble pattern, with a long and persistent expansion when the linear trend is present and followed by a sudden crash when the series returns to a stationary process. This might be mistaken for noncausal dynamics, ensuring a non zero lead order identification when the series is not detrended. This claim is

\(^4\)Results when the pseudo lag order is fixed to the correct one (\( p = 1 \) or \( p = 2 \) for mixed models) are available upon request.
supported by the results in the last column, indicating large proportion of wrongly detected noncausal dynamics for each detrending approaches, with 7.54% for $HP_1$ and more than 28% for the others. For the $dgp$’s with other trends (or no trend) $HP_1$ has a maximum of rate of wrongly estimating the pseudo causal lag order of 10.78%. For the three other detrending methods the pseudo causal lag order is wrongly identified in less than 7.3% of cases. Note that when the lag order of wrongly identified, it is almost always due to over-identification. The discrepancy between the two HP filters is explained by the low penalizing parameter in $HP_1$ allowing the trend to mimic the series too much. By that, some of the dynamics of the $MAR$ process are absorbed by the trend. Hence, the detrending methods employed here for those $dgp$’s do not significantly affect the correlation structure of the data. Moreover, the frequencies of obtaining the correct lag lengths increase with the sample size (results available upon request).

It is almost always much more harmful not to detrend when necessary than the contrary. As can be seen on the upper rows of Table 2, applying polynomial trends or $HP_2$ do not increase the proportions of wrongly identified models by more than 1.6%, compared to estimation on the raw series. However, when the existing trend is ignored, the pseudo lag order is wrongly estimated twice as much on the raw series as for the detrended series, and the $MAR$ models are wrongly identified up to 6 times more than the best performing detrending methods. Furthermore, the wrong identification of the pseudo lag order $p$ accounts for most of the proportion of wrongly identified $MAR$ models. If $p$ is correctly estimated, the model is also correctly identified in more than 99% of the cases. Note that the pseudo causal lag order identified is never zero, meaning that no detrending completely absorbs all dynamics. Besides, in no more than 0.62% the detrending methods killed the noncausal dynamics, as is indicated by the columns $s = 0$. Let us now consider the last column, displaying the results for purely causal processes. We here investigate whether detrending can create spurious noncausal dynamics ($s > 0$). We find that (ignoring the $dgp$’s with the trend with breaks) as long as the polynomial trend order is not underestimated, in less than 3.46% of the cases noncausal dynamics was wrongly identified. For the processes with a polynomial trend of order 6, detrending with a polynomial trend of order 4 creates spurious noncausal dynamics in 60.02% of the cases.

Overall, for $dgp$’s with noncausal dynamics, the impact of ignoring a trend is quite significant while detrending when not necessary has negligible effects.
Table 2: Identification of MAR models

| Detrending method | p ≠ 1 wrong MAR s = 0 | p ≠ 2 wrong MAR s = 0 | s > 0 |
|-------------------|------------------------|------------------------|-------|
| **raw**           | 5.50                   | 5.50                   | 0.00  |
| t^4               | 5.46                   | 5.58                   | 0.14  |
| t^6               | 5.70                   | 5.86                   | 0.22  |
| HP_1              | 10.78                  | 11.24                  | 0.52  |
| HP_2              | 6.84                   | 7.10                   | 0.28  |

| **raw**           | 12.10                  | 13.84                  | 35.04 |
| t^4               | 6.44                   | 6.72                   | 0.28  |
| t^6               | 6.76                   | 6.96                   | 0.20  |
| HP_1              | 10.28                  | 10.88                  | 0.60  |
| HP_2              | 6.24                   | 5.56                   | 0.32  |

| **raw**           | 13.18                  | 36.14                  | 26.04 |
| t^4               | 7.30                   | 7.36                   | 0.08  |
| t^6               | 6.54                   | 6.68                   | 0.14  |
| HP_1              | 9.40                   | 9.84                   | 0.44  |
| HP_2              | 5.94                   | 6.12                   | 0.18  |

| **raw**           | 4.54                   | 4.92                   | 0.68  |
| t^4               | 3.44                   | 4.00                   | 0.60  |
| t^6               | 3.40                   | 3.86                   | 0.58  |
| HP_1              | 4.00                   | 4.58                   | 0.62  |
| HP_2              | 3.38                   | 3.70                   | 0.40  |

Notes: During the first stage of the model identification, the maximum number of lags in the pseudo lag model is set to 4. Results are in percentages of the 5,000 trajectories. T = 400. HP_1 corresponds to the HP filter with \( \lambda = 14,400 \) and HP_2 to the HP filter with \( \lambda = 129,600 \).

on estimation. Both the polynomial trends and the HP filter with \( \lambda = 129,600 \) (HP_2) perform equally likely with respect to identifying the correct orders of the model. Choosing a penalizing parameter \( \lambda \) too low alters the dynamics of the process as shown by the results from HP_1. All of the approaches almost always retain the noncausal dynamics, but rarely create spurious noncausal dynamics when nonexistent in the \( dgp \) (except when the polynomial trend order is underestimated). The lead order is not always the correct one but in less than 0.62% for all cases no noncausal dynamics is found. The presented results only report identification of the model lag and lead orders. To have a better understanding of the impact of the detrending methods on the dynamics, focus needs to be put on the distribution of the estimated coefficients and parameters of the models identified.
4.2 Effects of detrending on estimated coefficients

We now investigate the persistence of the dynamics from the magnitude of the estimated coefficients. For instance, a lower lead coefficient will indicate shorter lived bubbles compared to the true generated process and thus increases the probabilities of a crash during an explosive episode. The same goes for larger degrees of freedom when the errors follow a Student’s $t$ distribution: larger degrees of freedom correspond to thinner tails, and thus rarer extreme values and thus makes less probable long lasting explosive episodes.

We investigate the distribution of the estimated coefficients given a correctly identified model. Proportions of wrongly identified models per dgp and detrending method are shown in the columns ‘wrong MAR’ of Table 2. Hence, proportions of correctly identified models range between 76.76% and 96.3% of the 5000 replications, but are almost always above 90%. Figure 4 reports the box plots of estimated coefficients for the purely noncausal (left column) and mixed causal-noncausal (center and right columns, for the lag and lead coefficients respectively) processes after each of the four detrending approaches is applied. We indicate the true coefficients, 0.6 and 0.8 for the lag and lead respectively, by the vertical dotted line. The box plots indicate the minimum, maximum, the interquartile range and the median. The $HP_1$-filtered series (with $\lambda = 14000$) are on average characterized by lower estimated lead and lag coefficients than the other detrended series. This is due to the low penalization of the filter, leading to a trend mimicking too much the initial series and thus capturing too much of the dynamics, reducing the persistence of the true noncausal process. Furthermore, we can see that using polynomial trends does not affect estimations of the coefficients, on average, as long as the order of the trend estimated is at least that of the true trend. That is, underestimating the order of the trend leads to an alteration of the dynamics and in our case, to more persistent noncausal dynamics. The $HP_2$ filter performs similarly to $t^6$, but we can expect that if the true trend was a higher order, $HP_2$ would perform better. The constructed linear trend with breaks leads to much larger noncausal coefficients for all detrending methods. The second break in the trend mimics the crash of a bubble and the long expansion preceding it leads to the identification of the model with a larger lead coefficient, which corroborates the earlier findings. Importantly, lag coefficients are on average correctly identified (the distributions of the estimated degrees of freedom, available upon request, show that they are not significantly affected by the detrendings either). A wrong detrending therefore mostly affects the noncausal dynamics of the processes.
Figure 4: Distribution of estimated MAR coefficients
Overall, $HP_1$, due to low penalization, absorbs too much of the dynamics (mostly the noncausal ones) in the resulting trend. Hence, for monthly data, we advise to use the HP filter with penalization parameter $129\,600$. It is also rather harmful to underestimate the order of the polynomial trend, which results in a significantly larger lead coefficient. When the fundamental trend itself mimics bubbles (comprises long phases of increase followed by a crash), detrending methods usually do not succeed in capturing the trend and this translates in much more persistent noncausal dynamics. We also investigated the effect of detrending white noise series; while for the raw series, $6.82\%$ of the models were identified with dynamics, only $7.34\%$ were identified with dynamics for the HP-filtered series with penalizing parameters $129\,600$. Hence we find no significant creation of dynamics when applying the HP filter to a white noise.

5 Predicting oil price bubbles

This section analyzes the impact of detrending on the model estimated and the resulting predictions. We employ WTI and Brent crude oil monthly prices series, respectively from January 1986 and from May 1987, to September 2019. We consider monthly series since we focus on one-step ahead forecasts. The series are both characterized by what seem to be bubble episodes, that is, rapidly increasing episodes followed by a sudden crash. As shown in Figure 5, the series are noticeably nonstationary but considering their growth rate would eliminate the locally explosive episodes that we want to exploit. Bubbles are assumed to be deviations from the fundamental value, which, as mentioned in the introduction, is not easily appraised for commodities. The two series appear almost identical until the 2008 financial crisis, period from which we can observe more apparent discrepancies. The last part of the samples is rather noisy and volatile, and estimating a trend on such a part is not straightforward. Without any information about the fundamentals, we cannot know whether the end of the sample corresponds to the outset of an explosive episode or whether it is stationary around its fundamental value. We have seen in Section 4 that underestimating the order of the trend has a significant impact on estimation. Yet, if the last part of the sample is indeed an explosive episode, then a polynomial trend of order 4 would capture it better than a polynomial trend of order 6 and overestimating the trend could here absorb some significant dynamics.

The questions about which transformation to consider as well as what fil-
tration to apply to obtain a stationary time series hence arise. Taking the logarithm or adjusting the series with the consumer price index alter the magnitudes of the bubbles and trends, but do not render the series stationary. We will therefore consider the mentioned transformations as well as three detrending methods (polynomial trends of orders 4 and 6 and the HP filter with $\lambda = 129\,600$, denoted $t^4$, $t^6$ and HP respectively) and investigate their implications for predictions. Furthermore, we are interested in forecasting locally explosive episodes. The first one, representing the 2008 financial crisis, serves for the identification of the model. If we cut the sample before the crash in October 2008 there is not enough information to differentiate an autoregressive model with an explosive root from an non-causal process which has not returned to its central values yet. Hence we want to focus on another episode which could have been the outset of a bubble but turned out not to be. In May 2018 (as indicated by the dashed vertical line in Figure 5) we can see that the two series were both entering an increasing phase. We focus on this point in time and investigate probabilities of entering in a bubble or of crashing, which they both eventually did in November 2018. As an illustration, Figure 6 depicts the detrended Brent series. We see that the magnitude of the last episode in the end of the sample depends significantly on the detrending method applied. While the resulting series are not substantially different, the HP-detrended series (dashed line) is almost always located between the two other series and follows the $t^6$-detrended series (light gray line) at the end of the sample. With a trend of order 4 (dark gray line), the last episode in the sample reaches the level of the bubble of the 2008 financial crises. This suggests a much
more explosive episode than the two other approaches, for which May 2018 resembles the period of 2011-2015 instead. Note that for logarithmic series, the polynomial trend of order 4 even induces a larger bubble in May 2018 that the one of the 2008 financial crisis, for both log(WTI) and log(Brent) series.

![Figure 6: Monthly detrended Brent crude oil prices](image)

5.1 Estimation results: models identifications

We use the series in levels, their logarithmic transformation and the series adjusted for inflation using the Consumer Price Index. We detrend them all using polynomial trends up to orders 4 and 6 and the HP filter with a penalization parameter $\lambda = 129,600$ (denoted $t^4$, $t^5$ and HP respectively) on the samples up to May 2018. We then estimate $MAR$ models with Student’s $t$-distributed errors and set the maximum pseudo lag length in the first stage on the estimation to 4. All resulting models are $MAR(1,1)$ and are reported in Table 3. We report the lag and lead coefficients as well as the degrees of freedom of the distribution and their respective standard errors in parentheses.

Models estimated on series that were detrended with a polynomial trend of order 6 and with the HP filter are the most similar. Models estimated after detrending with the polynomial trend of order 4 slightly deviate from...
Table 3: Estimated MAR models

| Series       | t⁴       | Results per detrending method | t⁵     | HP       | Results per detrending method |
|--------------|----------|-------------------------------|--------|----------|-------------------------------|
|              | φ        | ψ                             | t(γ)   | φ        | ψ                             | t(γ)   |
| WTI          | 0.37     | 0.86                          | 1.96   | 0.46     | 0.80                          | 1.99   | 0.46 | 0.78 | 1.92 |
|              | (0.03)   | (0.01)                        | (0.29) | (0.03)   | (0.02)                        | (0.30) | (0.03) | (0.02) | (0.29) |
| log(WTI)     | 0.34     | 0.85                          | 5.36   | 0.38     | 0.81                          | 5.96   | 0.38 | 0.80 | 5.50 |
|              | (0.04)   | (0.02)                        | (1.52) | (0.04)   | (0.03)                        | (1.87) | (0.04) | (0.03) | (1.58) |
| WTI_real     | 0.35     | 0.85                          | 2.94   | 0.38     | 0.83                          | 2.98   | 0.39 | 0.81 | 2.93 |
|              | (0.04)   | (0.02)                        | (0.55) | (0.04)   | (0.02)                        | (0.56) | (0.04) | (0.02) | (0.55) |
| Brent        | 0.86     | 0.47                          | 1.77   | 0.87     | 0.41                          | 1.78   | 0.84 | 0.43 | 1.77 |
|              | (0.02)   | (0.03)                        | (0.23) | (0.02)   | (0.03)                        | (0.25) | (0.02) | (0.03) | (0.24) |
| log(Brent)   | 0.87     | 0.34                          | 5.53   | 0.86     | 0.30                          | 5.56   | 0.85 | 0.32 | 5.52 |
|              | (0.02)   | (0.04)                        | (1.55) | (0.02)   | (0.04)                        | (1.57) | (0.03) | (0.04) | (1.56) |
| Brent_real   | 0.84     | 0.47                          | 2.46   | 0.86     | 0.39                          | 2.49   | 0.83 | 0.43 | 2.46 |
|              | (0.02)   | (0.03)                        | (0.41) | (0.02)   | (0.03)                        | (0.42) | (0.02) | (0.03) | (0.41) |

Notes: The models are obtained with a maximum pseudo lag order of 4 and for each series the model identified was an MAR(1,1). φ is the lag coefficient, ψ is the lead coefficient and γ the degrees of freedom of the Student’s t distribution. The polynomial trend are trends up to the order indicated and the HP filtering is performed with a penalization parameter λ = 129 600. In parentheses are reported the standard error of the coefficients estimated obtained with the MARX package (Hecq et al., 2017).

the two others and always have a larger lead coefficient. This is due to the assumption that if the series follows a trend of order 4, the last part of the sample is already more deeply in a bubble. This is hence captured by a more persistent noncausal coefficient, as what Section 4 suggests for underestimating the trend order in mixed causal-noncausal models. What is striking is that while WTI and Brent prices series follow similar patterns, as shown by Figure 5, WTI is mostly forward looking and Brent prices are mostly backward looking. This can be seen in the magnitude of the lag and lead coefficients. The bimodality of the coefficient distribution in the estimation can lead, in the optimization of the likelihood function, to a local maximum (Bec, Bohn Nielsen, and Saïdi, 2019). This phenomenon is subject to initial values and can induce a switch between the lag and lead coefficients. This is however not the case in this analysis, multiple initial values were employed, and when two distinct models were identified, the one
with the largest likelihood was chosen. Let us now look at the estimated degrees of freedom of the Student’s $t$ distributions. Adjusting the series for inflation increases (roughly by one) the estimates of the degrees of freedom and will consequently imply lower probabilities of bubbles. Taking the logarithm of the series, which mostly affects the magnitude of the locally explosive episodes, yields degrees of freedom always superior to 5, implying significantly lower probabilities of bubbles, as investigated in the subsequent part.

5.2 Predictions

Since the errors follow Student’s $t$ distributions with degrees of freedom different from 1, the predictive densities do not admit closed-form expressions. We use the two data-driven approaches presented in Section 3. The simulations-based approach of Lanne et al. (2012) only depends on the model estimated and the last observed point while the sample-based approach of Gourieroux and Jasiak (2016) uses again all past observed values at the forecasting step. Table 4 shows the one-month ahead probabilities that the series is going to keep on increasing and thus potentially follow a bubble (columns 4 to 6) and probabilities that the series will crash (we define a crash as a drop of at least 25% of the last observed detrended value, shown in the last three columns). Results from the two prediction methods are reported for each of the series, their transformations, and the three detrending methods employed in the analysis. The last column of each block of results correspond to the difference between the results of the two methods. In the third column are reported the empirical quantiles corresponding to the detrended values in May 2018 based on the estimated model for each series.

Hecq and Voisin (2019) show that the discrepancy between the sample- and simulations-based approaches widens during explosive episodes. They also show that the larger the lead coefficient, the more the sample-results tend to yield larger probabilities of a crash than the ones computed with simulations (or the theoretical probabilities). Those phenomena can be observed here (presented in columns ‘diff.’), and especially for the WTI and WTI_{real} series detrended with $t^4$. Remember that the sample-based approach is characterized by its learning mechanism. In the series observed here, and for the ones for which we assumed the last phase is an explosive episode (detrended with $t^4$), the sample-based result indicates 11.5% larger probabilities that WTI
Table 4: One-step ahead probabilities performed in May 2018

| Series | Detrended with | Emp. quant. | Proba. increase samp. | Proba. decrease > 25% samp. | diff. | diff. |
|--------|----------------|-------------|-----------------------|-----------------------------|-------|-------|
| WTI    | $t^4$          | 0.989       | 0.790 0.675 0.115     | 0.149 0.231 -0.082         |       |       |
|        | $t^6$          | 0.953       | 0.581 0.604 -0.023    | 0.230 0.204 0.026          |       |       |
|        | $HP$           | 0.971       | 0.551 0.596 -0.045    | 0.283 0.253 0.030          |       |       |
| log(WTI)| $t^4$          | 0.999       | 0.373 0.372 0.001     | 0.186 0.317 -0.131         |       |       |
|        | $t^6$          | 0.911       | 0.457 0.447 0.010     | 0.250 0.249 0.001          |       |       |
|        | $HP$           | 0.981       | 0.419 0.399 0.020     | 0.235 0.242 -0.007         |       |       |
| WTI$_{real}$ | $t^4$          | 0.997       | 0.802 0.547 0.255     | 0.106 0.293 -0.187         |       |       |
|        | $t^6$          | 0.943       | 0.535 0.520 0.014     | 0.209 0.200 0.008          |       |       |
|        | $HP$           | 0.977       | 0.508 0.505 0.003     | 0.237 0.242 -0.005         |       |       |
| Brent  | $t^4$          | 0.984       | 0.302 0.363 -0.061    | 0.028 0.028 0.000          |       |       |
|        | $t^6$          | 0.946       | 0.403 0.436 -0.032    | 0.057 0.058 -0.001         |       |       |
|        | $HP$           | 0.966       | 0.347 0.391 -0.044    | 0.072 0.075 -0.003         |       |       |
| log(Brent)| $t^4$          | 0.999       | 0.270 0.266 0.004     | 0.032 0.032 0.000          |       |       |
|        | $t^6$          | 0.920       | 0.430 0.343 -0.004    | 0.181 0.183 -0.002         |       |       |
|        | $HP$           | 0.980       | 0.361 0.362 -0.001    | 0.134 0.133 0.001          |       |       |
| Brent$_{real}$ | $t^4$          | 0.993       | 0.279 0.306 -0.027    | 0.046 0.048 -0.002         |       |       |
|        | $t^6$          | 0.951       | 0.429 0.440 -0.011    | 0.090 0.091 -0.001         |       |       |
|        | $HP$           | 0.975       | 0.365 0.377 -0.012    | 0.105 0.108 -0.003         |       |       |

Notes: The empirical quantiles (Emp. quant.) corresponding to the last observed points in May 2018 and are computed using simulations, based on the estimated model of each detrended series. For the sample-based approach (samp.) approximations employs all the past observed values until May 2018. For the simulations-based approach (sims.) the truncation parameter $M = 100$ and 500000 simulations were used. Proba. decrease > 25% represents the probabilities to decrease at least by 25% of the last observed value.

will keep on increasing than the simulations results suggest (79% against 67.5%), and 25.5% larger for WTI$_{real}$ (80.2% against 54.7%). This stems from the fact that the series has attained once before this point and then kept on increasing. It is therefore, based on the learning mechanism, not unlikely that it will happen again. For the two other detrending methods, the series already attained multiple times the point reached in May 2018 and either kept on increasing or crashed, this is why the learning mechanism does not bring much more information to the simulations-based approach. For $t^6$ and $HP$ the difference between the two methods is less than 4.5% for
all other WTI series. For the Brent series, the discrepancies do not exceed 6.1%. This is due to the lower lead coefficient of around 0.4 and probabilities of a further increase range between 26.6% and 44% for the simulations-based results. Hecq and Voisin (2019) also show that for a given quantile, probabilities of a further increase are lower for large degrees of freedom. The larger degrees of freedom estimated for the logarithmic series imply lower probabilities of extreme values and therefore lower probabilities of long-lasting bubbles. This is why we can observe lower probabilities of entering an increasing phase for the logarithmic series.

Comparing results for the increase and for the crash of at least 25%, we can see that the sum of the probabilities of the two events, for a given method, do not sum up to 1. The minimum is for $t^4$-detrended log(Brent), for which probabilities (for both prediction methods) sum up to roughly 0.3. This means that there are nonzero probabilities (in the mentioned example, 0.7) that the variables will decrease, but by no more than 25%. The definition of the crash employed to compute probabilities (here a drop of 25%, but it could have been a drop of 50% or 10% for instance) is arbitrary and can affect results. The larger the last observed value, the less this arbitrary choice impacts results, especially for the simulations-based approach. Indeed, the two modes of the predictive distribution are 0 and $\psi^{-1}u_T$ for purely non-causal processes (shifted by $\phi_T$ for $MAR(1,1)$ processes). Therefore, the larger is $u_T$, the further apart will be the two modes of the distribution, and the two events (crash and increase) will be more easily distinguished from one another. For instance, for the $t^4$-detrended WTI series, May 2018 corresponds to the quantile 0.989 of the underlying distribution. The sum of the probabilities of a further increase and of a crash of at least 25% is more than 0.9 for both methods, as opposed to the $t^6$- and $HP$-detrended WTI series, for which probabilities sum up to around 0.8. For the sample-based approach, which is very sensitive to past behaviors and assigns significant probabilities to events in between the modes, the effect might be lessen with large quantiles but still rather persistent. Brent has a lower lead coefficient and therefore lower probabilities of pursuing a bubble, as can be seen both from the sample- and simulations-based approaches. On the other hand, the probabilities of a crash of at least 25% are also much lower, and this is due to the large lag coefficient, shifting the densities and ensuring around 85% of the last observed value. Hence, probabilities that the series will be less than 75% of the last observed value are close to zero.

Figure 7 illustrates where the differences in probabilities between the series
and the detrending methods come from. Each graph depicts the one-step ahead predictive density functions of the detrended WTI (left) and Brent (right) series. The dashed vertical line represents the last observed detrended value while the dotted one represents 75% of the last value. First, for each transformation (nominal level, logarithm, real series) a given detrending method yields roughly the same quantile corresponding to the last point for WTI and Brent, as shown in the column Emp. quant. of Table 4. The difference between the predictive densities of WTI and Brent (given a detrending) hence stems from the inversion of the lag and lead coefficients. This somehow mirrors the distributions. The larger the lead coefficient, the more probable is a bubble to be long-lasting and thus keep on increasing. Hence, for the WTI series, we can see the larger mode of the distribution is the one corresponding to the further increase while for the Brent series...
it is on the left mode, corresponding to the crash. Secondly, the differences in probabilities between the detrending methods is that for $t^4$, the series is clearly in an explosive episode while for the two others it is at the outset. If the series is only on the outset of a bubble, it may or may not increase, with probabilities close to a half, but if it does not increase, it has significant probabilities to remain close to its last values, and 25% of its last value is still not a significant crash (relative to the level of the series). On the other hand, if the series is in an explosive phase, it will most likely keep on increasing, as we are now certain there is an extreme value triggering it, but if it crashes, it will almost surely crash by a significant amount. Probabilities that are computed for this analysis are the cumulative probabilities on the outside of the two indicated thresholds and what remains is the probability that the series will not increase but will not drop by more than 25% (hence in between the two vertical lines). We can see that for the WTI series, the two thresholds considered ($0.75y_T$ and $y_T$) still remain between the two main modes of the distribution, but for the Brent series, the significant shift (around $0.8y_T$) due to the large persistence of the causal part of the process, makes the threshold for the crash go to the very left tail of the distribution, and the probabilities remaining (a crash not more than 25%) becomes substantial.

Overall, the probabilities computed in this analysis indicate that the WTI series is much more likely to keep on increasing, suggesting that Brent is a riskier investment. However, if the series decrease, it is more likely that WTI will drop by a more substantial amount than Brent.

5.3 Testing for Common Bubbles

Economic and financial time series exhibit many distinctive characteristics among which the presence of serial correlation, seasonality, stochastic trends, time varying volatility or non-linearities. However, in a multivariate analysis, it is frequent to observe that one or more of these features, detected in individual series, are common to several variables and thus disappear with some suitable combination. We then talk about common features. The leading example is probably cointegration, namely the presence of common stochastic trend (Engle and Granger, 1987). Other forms of co-movements have also been studied, giving rise to developments around the notions of common cyclical features (Engle and Kozicki, 1993), common deterministic seasonality (Engle and Hylleberg, 1996), common volatility (Engle and Susmel, 1993), etc. Recognizing these common feature structures presents
numerous advantages from an economic perspective (e.g. the whole literature on the presence of a long-run relationship). There are also several implications for statistical modeling. For instance, imposing some commonalities helps to reduce the number of parameters that must be estimated. That potentially leads to efficiency gains and to improvements of forecasts accuracy (Issler and Vahid, 2001). Using the common factor structure can also be used to forecast a set of time series using only the forecast of the common component and the estimated loadings.

Building on such a common features approach, we investigate whether the noncausal process (the bubbles) that we have identified on the WTI and Brent series is common to both. That is, while two series $y_t$ and $x_t$ individually display a speculative bubble pattern, a combination of those series may not. Given their similar patterns, the empirical investigation of WTI and Brent oil prices series is a good example of the potential existence of such relationships. However, the difference between estimated MAR parameters hints that the dynamics of those series differ. Cubadda et al. (2019) extend the canonical correlation framework of Vahid and Engle (1993) from purely causal VARs to purely noncausal VARs. They show that more commonalities emerge when we also look at VARs in reverse time. The tests statistics they developed do not generally work for mixed models though. Consequently we consider a more heuristic approach in this paper. We investigate the existence of a scalar $\delta$ such that

$$z_t = (y_t - \delta x_t),$$

where, under the null hypothesis, $z_t$ is a white noise or a purely causal process with reduced order of total dynamics; $y_t$ and $x_t$ are stationary MAR processes with a non-zero lead order. Since all transformations and detrended series in this analysis are $MAR(1,1)$, we investigate the existence of a linear combination leading to $MAR(0,0)$ or $MAR(1,0)$, estimated using the same procedure as described in Section 4. Note that it is not possible to estimate $\delta$ using the canonical correlation or a GMM approach. Indeed, due to the double moving average representation of MAR series, we cannot find instruments that are orthogonal for both leads and lags. Consequently we rely on a grid search strategy.

We do not find any combination of the WTI and Brent series that annihilated the noncausal components of the series. This points to the fact that the series do not have a common bubble, and that they indeed have different roots in
their construction (WTI being mostly forward looking and Brent backward looking).

6 Conclusion

This paper aims at shedding light upon how transforming or detrending a series can substantially impact predictions of mixed causal-noncausal models. Assuming a polynomial trend of order 4 for WTI and Brent series probably underestimates the trend component. The HP filter (with $\lambda = 129600$) does not require any further assumptions with respect to the trend and can therefore be an adequate filter in cases where the fluctuations of fundamental values are unknown. Overall, caution is needed when detrending a series and some filtering, such as polynomial trends, may require additional understanding regarding the deviations of the series from its fundamental trend. However, once the series is detrended, resulting in a stationary series, using $MAR$ models is a straightforward approach to model nonlinear time series. They capture the locally explosive episodes observed in oil prices in a strictly stationary setting. While the bi-modality of the predictive density would not be detected with standard Gaussian $ARMA$ models, it could be detected with complex nonlinear models, but such model lacks the parsimonious characteristic of $MAR$ models. The data-driven prediction methods may lack theoretical grounds but provide valuable information based on the estimated model and on past behaviors of the series in a parsimonious way. This paper focuses on one-step ahead predictions of turning points. However, probabilities of longer trajectories (e.g. six months) can also be computed.
Acknowledgments

The authors would like to thank Francesco Giancaterini. All errors are ours.

References

Backus, D. K., and Kehoe, P. J. (1992). International evidence on the historical properties of business cycles. The American Economic Review, 864–888.

Bec, F., Bohn Nielsen, H., and Saïdi, S. (2019). Mixed causal-noncausal autoregressions: Bimodality issues in estimation and unit root testing (Tech. Rep.).

Bertelsen, K. P. (2019). Comparing tests for identification of bubbles (Tech. Rep.). Department of Economics and Business Economics, Aarhus University.

Breid, F. J., Davis, R. A., Li, K.-S., and Rosenblatt, M. (1991). Maximum likelihood estimation for noncausal autoregressive processes. Journal of Multivariate Analysis, 36(2), 175–198.

Brooks, C., Prokopczuk, M., and Wu, Y. (2015). Booms and busts in commodity markets: bubbles or fundamentals? Journal of Futures Markets, 35(10), 916–938.

Campbell, J. Y., and Shiller, R. J. (1987). Cointegration and tests of present value models. Journal of political economy, 95(5), 1062–1088.

Canova, F. (1998). Detrending and business cycle facts. Journal of monetary economics, 41(3), 475–512.

Cavaliere, G., Nielsen, H. B., and Rahbek, A. (2018). Bootstrapping noncausal autoregressions: with applications to explosive bubble modeling. Journal of Business and Economic Statistics, 1–13.

Cubadda, G., Hecq, A., and Telg, S. (2019). Detecting co-movements in non-causal time series. Oxford Bulletin of Economics and Statistics, 81(3), 697–715.

Cuthbertson, K., and Nitzsche, D. (2005). Quantitative financial economics: stocks, bonds and foreign exchange. John Wiley and Sons.

Degiannakis, S., and Filis, G. (2018). Forecasting oil prices: High-frequency financial data are indeed useful. Energy Economics, 76, 388–402.

De Jong, R. M., and Sakarya, N. (2016). The econometrics of the hodrick-prescott filter. Review of Economics and Statistics, 98(2), 310–317.

Diba, B. T., and Grossman, H. I. (1988). Explosive rational bubbles in stock prices? The American Economic Review, 78(3), 520–530.
Engle, R. F., and Granger, C. W. (1987). Co-integration and error correction: representation, estimation, and testing. *Econometrica: Journal of the Econometric Society*, 251–276.

Engle, R. F., and Hylleberg, S. (1996). Common seasonal features: Global unemployment. *Oxford Bulletin of Economics and Statistics*, 58(4), 615–630.

Engle, R. F., and Koiziki, S. (1993). Testing for common features. *Journal of Business and Economic Statistics*, 11(4), 369–380.

Engle, R. F., and Susmel, R. (1993). Common volatility in international equity markets. *Journal of Business and Economic Statistics*, 11(2), 167–176.

Engsted, T. (2006). Explosive bubbles in the cointegrated var model. *Finance Research Letters*, 3(2), 154–162.

Engsted, T., and Nielsen, B. (2012). Testing for rational bubbles in a coexplosive vector autoregression. *The Econometrics Journal*, 15(2), 226–254.

Fan, Y., and Xu, J.-H. (2011). What has driven oil prices since 2000? a structural change perspective. *Energy Economics*, 33(6), 1082–1094.

Frey, G., Manera, M., Markandya, A., and Scarpa, E. (2009). Econometric models for oil price forecasting: A critical survey. In *Cesifo forum* (Vol. 10, pp. 29–44).

Fries, S., and Zakoïan, J.-M. (2019). Mixed causal-noncausal AR processes and the modelling of explosive bubbles. *Econometric Theory*, 1–37.

Funk, C. (2018). Forecasting the real price of oil-time-variation and forecast combination. *Energy Economics*, 76, 288–302.

Gouriéroux, C., and Jasiak, J. (2016). Filtering, prediction and simulation methods for noncausal processes. *Journal of Time Series Analysis*, 37(3), 405–430.

Gouriéroux, C., and Zakoïan, J.-M. (2013). Explosive bubble modelling by noncausal process. *CREST. Paris, France: Centre de Recherche en Economie et Statistique*.

Gouriéroux, C., and Zakoïan, J.-M. (2017). Local explosion modelling by non-causal process. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 79(3), 737–756.

Hamilton, J. D. (1989). A new approach to the economic analysis of non-stationary time series and the business cycle. *Econometrica: Journal of the Econometric Society*, 357–384.

Hecq, A., Issler, J. V., and Telg, S. (2019). *Mixed causal-noncausal autoregressions with exogenous regressors* (Tech. Rep.). Escola de Pós-Graduação em Economia da FGV.
Hecq, A., Lieb, L., and Telg, S. (2017). Simulation, estimation and selection of mixed causal-noncausal autoregressive models: The marx package.

Hecq, A., and Voisin, E. (2019). Forecasting bubbles with mixed causal-noncausal autoregressive models. *MPRA Paper No. 96350 at https://mpra.ub.uni-muenchen.de/96350/.*

Hencic, A., and Gouriéroux, C. (2015). Noncausal autoregressive model in application to bitcoin/usd exchange rates. In *Econometrics of risk* (pp. 17–40). Springer.

Hodrick, R. J., and Prescott, E. C. (1997). Postwar us business cycles: an empirical investigation. *Journal of Money, credit, and Banking*, 1–16.

Homm, U., and Breitung, J. (2012). Testing for speculative bubbles in stock markets: a comparison of alternative methods. *Journal of Financial Econometrics*, 10(1), 198–231.

Issler, J. V., and Vahid, F. (2001). Common cycles and the importance of transitory shocks to macroeconomic aggregates. *Journal of Monetary Economics*, 47(3), 449–475.

Karapanagiotidis, P. (2014). Dynamic modeling of commodity futures prices.

Lanne, M., Luoto, J., and Saikkonen, P. (2012). Optimal forecasting of noncausal autoregressive time series. *International Journal of Forecasting*, 28(3), 623–631.

Lanne, M., and Saikkonen, P. (2011). Noncausal autoregressions for economic time series. *Journal of Time Series Econometrics*, 3(3).

Lof, M., and Nyberg, H. (2017). Noncausality and the commodity currency hypothesis. *Energy Economics*, 65, 424–433.

Miao, H., Ramchander, S., Wang, T., and Yang, D. (2017). Influential factors in crude oil price forecasting. *Energy Economics*, 68, 77–88.

Miller, J. I., and Ratti, R. A. (2009). Crude oil and stock markets: Stability, instability, and bubbles. *Energy Economics*, 31(4), 559–568.

Phillips, P. C., and Shi, Z. (2019). Boosting the hodrick-prescott filter. *arXiv preprint arXiv:1905.00175*.

Phillips, P. C., Wu, Y., and Yu, J. (2011). Explosive behavior in the 1990s nasdaq: When did exuberance escalate asset values? *International economic review*, 52(1), 201–226.

Pindyck, R. S. (1993). The present value model of rational commodity pricing. *The Economic Journal*, 103, 511–530.

Ravn, M. O., and Uhlig, H. (2002). On adjusting the hodrick-prescott filter for the frequency of observations. *Review of economics and statistics*, 84(2), 371–376.
Su, C.-W., Li, Z.-Z., Chang, H.-L., and Lobont¸, O.-R. (2017). When will occur the crude oil bubbles? Energy Policy, 102, 1–6.

Vahid, F., and Engle, R. F. (1993). Common trends and common cycles. Journal of Applied Econometrics, 341–360.

Wang, Y., Liu, L., and Wu, C. (2017). Forecasting the real prices of crude oil using forecast combinations over time-varying parameter models. Energy Economics, 66, 337–348.