Oil Price Volatility, Endogenous Regime Switching and Macroeconomic Factors

Yoosoon Chang † Indiana University
Ana María Herrera ‡ University of Kentucky
Elena Pesavento § Emory University
November 8, 2021

Abstract

We study the evolution WTI oil returns using a novel approach to model regime switching with dynamic feedback and dynamic interactions. That is, instead of modeling the process that governs the switching between mean volatility regimes as exogenous, switching depends on whether the underlying latent factors exceed or not a threshold. Moreover, innovations in the latent factors are assumed to be correlated with the innovations of the WTI returns in the previous period; hence, future transitions, say from a low mean-high volatility to a low mean-high volatility regime, depend on past states. We then investigate what macroeconomic variables explain the fluctuations in the mean and volatility factors using two regularization methods (adaptive LASSO and adaptive Elastic-net). We find that two indicators of economic activity and monetary policy play an important role in explaining the variation in the full sample.

JEL Classification: C13, C32, E32, Q35.
Key words: oil price volatility, endogenous regime switching, LASSO.

*We are grateful to participants in the 2019 Workshop on Energy Economics and the 2019 Midwest Econometrics Group meetings for very useful comments. Our special thanks go to Sangmyung Ha and Diego Rojas Baez for their superb research assistance and to Joon Y. Park and Shi Qiu for many helpful discussions.
†Indiana University, Email: yoosoon@indiana.edu.
‡University of Kentucky, Email: amherrera@uky.edu.
§Emory University, Email: epesave@emory.edu.
1 Introduction

Historically, crude oil price returns have experienced periods of increased volatility and periods of relative calm. Spikes in oil prices have been observed in the heels of political unrest in the Middle East, terrorist attacks on Saudi Arabia’s oil facilities, and have often coincided with economic downturns in U.S. economic activity such as the global shutdown that ensued the onset of the Covid-19 pandemic. Over the years, concerns regarding heightened volatility in oil prices and the burdens they may pose on households, businesses and investors have been reflected in policy statements by different chairmen of the Federal Reserve System, demands from senators to U.S. Commodity Futures Trading Commission, and an extensive literature on the economic impact of oil price shocks.

How do we best model the transition from periods of calm to periods of turmoil? What lies beneath the fluctuations in crude oil price returns? Which variables can capture the latent level of stress in oil markets? What information can we gather from a large set of macroeconomic data regarding the observed switches between periods of high and low volatility or high and low mean returns?

The aim of this paper is threefold. First, we use a Markov switching model with dynamic feedback and interactions to inquire into the process that drives transitions from periods of calm to periods of turmoil in oil markets. Our inquiry into these transitions is prompted by the historical behavior of the West Texas Intermediate oil prices and returns depicted in Figure 1. Crude oil prices, as other financial assets, often experience abrupt fluctuations. For instance, as Figure 1 illustrates, the price of the West Texas Intermediate (WTI) rose steadily from a $19.46 per barrel in November 2001 to a peak of $139.96 in June 2008, it then plunged to $41.96 in January 2009. As for the volatility of the WTI returns, it was moderate during the mid and late 1990s, increased considerably in the early 2000s and skyrocketed during the Covid-19 crisis. For this purpose, we estimate an endogenous regime switching model where we allow the mean and volatility of WTI returns to transition between two states (high/low mean and high/low volatility) in an unsynchronized manner. Hereafter, we will refer to this model as the mean-volatility switching model. This modeling strategy allows the mean and volatility regimes to be determined by two correlated latent factors, one for the mean and the other for the volatility, while it also accounts for possible feedback from past innovations of the WTI returns to both mean and volatility regime factors.

Second, we explore whether the extracted volatility latent factor could constitute a measure of oil market “risk” and, thus, provide analysts and investors with a tool to gauge volatility in the crude oil market. For instance, could the extracted volatility latent factor
be used to gauge risk in cases or time periods where alternative measures of oil market volatility are not readily available? After all, the Chicago Board Options Exchange did not start publishing the implied oil volatility index, OVX, until 2007 and its computation is based on the United States Oil Fund’s options.

Finally, we inquire which macroeconomic variables are more likely to explain the evolution of the latent mean and volatility regime factors. To do so we employ two regularization methods, adaptive Least Absolute Shrinkage and Selection Operator (LASSO) and adaptive Elastic-net, to select the relevant variables among a large group of macroeconomic and financial variables commonly used by policy makers and analysts. Thus, instead of taking the path of numerous studies that estimate a structural model of the oil market and inquire into the response of oil prices to structural shocks (see e.g., Kilian 2009; Kilian and Murphy 2012, 2014; Baumeister and Hamilton 2020, Herrera and Rangaragu 2020), we explore the link connecting the transition between high/low volatility (or mean) states and a large set of variables in a reduced-from setup. In other words, our objective is not to identify the effect of particular structural shocks on the evolution of crude oil prices. Instead, our goal is to provide some guidance regarding which macroeconomic variables, selected from a large database, contain information for understanding the transition between states.

We derive several important insights regarding the behavior of WTI returns from our analysis. Our estimation results indicate that the endogenous unsynchronized mean-volatility switching model does a better job at capturing the evolution of oil price returns than an exogenous switching model. Indeed, there is a great degree of variability in the transition probabilities over time, especially around the Persian Gulf War, the Great Recession, and the onset of the Covid-19 pandemic.

Moreover, a result that should be of interest for policy makers, analysts, and investors is that, when the underlying regime is known, allowing the regime switching to evolve in an endogenous manner is very informative regarding the likelihood that the returns will stay in the same regime. Indeed, when we examine four periods of low returns and high volatility, we find that the probability of staying in this regime declined quickly for all episodes, but the Great Recession. Thus, as the evolution of the transition probabilities illustrate, not all episodes should convey the same degree of “fear” to economic agents. Clearly, the Covid-19 pandemic resulted in risk levels not seen in the previous half century, as reflected in the evolution of the extracted volatility latent factor.

Can the fluctuations of the latent factors (and the proposed measure of oil market risk) be linked to particular macroeconomic variables? Using adaptive LASSO and adaptive Elastic
net regressions, we find two variables are key in explaining the evolution of the mean and volatility factors over the full sample: the total business inventories to sales ratio and the St. Louis Adjusted Monetary Base. In addition, indicators of labor market conditions, stock prices, consumer sentiments and inflation are useful in explaining fluctuations in the volatility’s latent factor. These results are robust to expanding the sample to include large databases for Canada, the UK and the Euro area.

Finally, when we estimate the Adaptive LASSO (or Adaptive Elastic-net) over a series of ten-year rolling samples, we uncover significant variations in the set of variables that account for the evolution of the latent factors. In particular, while total employment was highly relevant up to the 2000s and after the Great Recession, the S&P500 is key in explaining the volatility factor during the 2000s. Employment in mining is selected in most sub-samples that include the fracking revolution and gains importance after the 2014 collapse in oil prices. Since the Great Recession, variables reflecting changes in aggregate demand have become more relevant in explaining the volatility factor, while monetary policy uncertainty is selected around the time the U.S. economy hit the zero lower bound. Variables in the output and income, labor market, and prices groups are selected in the sample that contains the Covid-19 pandemic. In contrast with the volatility factor, we find that price variables also have explanatory power for the mean factor in the sub-samples that span the years before the Great Recession.

In brief, we provide policy makers, market analysts and investors with a new measure to gauge risk in oil markets: the extracted volatility latent factor from the unsynchronized mean-volatility switching model. For periods of time or countries where a measure of implied volatility based on relevant crude oil options cannot be computed, this factor may be used to signal changes in oil market risk.\footnote{Estimation results available from the authors upon request illustrate how the volatility latent factor extracted from a model for Brent oil returns captures changes in oil market risk.} We also show that, in the case of the WTI, this “risk factor” is correlated with a few U.S. macroeconomic variables that can be easily obtained from publicly available sources and could be used to alert agents about transitions from periods of calm to periods of turmoil in oil markets.

Clearly, this is not the first study to use a Markov switching (MS) framework to analyze the evolution of crude oil prices. Our paper is closely related to two strands of literature that employ Markov switching models in the analysis of oil price fluctuations. Researchers have relied on Markov switching models to forecast volatility of crude oil price returns and compare the predictive ability of MS-GARCH and MS long memory models to selected GARCH competitors (see e.g. Di Sanzo 2018, Herrera, Hu and Pastor 2018 and references}
This strand of literature has found evidence that Markov switching models tend to have superior predictive ability, especially during periods of turmoil. Our objective differs from these studies as we do not seek to forecast the volatility of oil price returns; instead, we restrict our investigation to an in-sample analysis. Furthermore, the forecasting literature mainly relies on Markov switching models where the transition probabilities evolve in an exogenous manner. Time-variation in such a setup stems from the forecasting design allowing the estimation sample, and hence the parameter estimates, to vary over recursive or rolling windows used for out-of-sample evaluation. In contrast, in the endogenous Markov switching model employed in this paper, time-variation stems from the feedback mechanism and the dynamic interactions between the mean and volatility latent factors.

In addition, few studies have employed endogenous Markov switching models to investigate the relationship between oil price fluctuations and the macroeconomy. For instance, Bjørnland et al. (2018) develop a Markov switching rational expectations New Keynesian model to study the role of oil price shocks in accounting for variability in the U.S. economy, as well as the contribution of a more hawkish monetary policy regime to the decline in macroeconomic volatility. Along similar lines and using data for the euro area, Holm-Hadulla and Hubrich (2017) identify two different regimes in the response of economic activity and inflation to oil price shocks: an adverse regime, where oil price shocks result in significant and sustained changes in economic activity and inflation, and a normal regime where the response of these variables is smaller and shorter-lived.

Our paper differs from the above strand of literature in several aspects. First, we do not build a structural model of the economy and, hence, do not define a-priori the possible sources of changes in regime (e.g., monetary policy states or past inflation rates), nor do we explicitly model the interactions between oil prices and the macroeconomy. Instead, we estimate a reduced form model where we endogenize the transition probabilities by allowing them to depend only on the past behavior of the crude oil returns. Then, we use regularization methods to select what variables, from a big database, have explanatory power for the extracted mean and volatility latent factors. A disadvantage of our approach is that we cannot use the usual time series tools such as impulse response functions or historical decompositions to conduct inference nor can we provide a structural interpretation of the source of oil price fluctuations. Yet, our approach has two advantages. It allows us to be agnostic regarding the variables that might drive the transition between regimes and to consider a

---

2 Endogenous Markov switching models have also been used to study the transmission of financial, monetary and fiscal policy shocks. See, for instance, Davig and Leeper (2006), Benigno, et al. (2020), Hubrich and Waggoner (2021), and Chang, Kwak and Qui (2021).
larger set of variables than it would be possible to handle in a structural VAR setup. Second, estimation is relatively straight-forward – we use a modified version of the Kalman filter and rely on maximum likelihood estimation instead of Bayesian methods – and it allows us to derive a measure of oil market “risk”.

This paper is organized as follows. Section 2 describes the endogenous regime switching models and the data. Section 3 describes the estimation results for a one factor (volatility switching) and a two-factor (volatility-mean switching) models, discusses how these estimates can be used to evaluate the likelihood to remain in a high volatility or a low mean-high volatility regime, and illustrates how the extracted latent volatility factor may constitute a measure of oil market risk. Section 4 inquiries into the links between macroeconomic variables and the latent factors. Section 5 concludes.

2 The Behavior of Crude Oil Returns: Endogenous versus Exogenous Regime Switching

Several econometric models are available to a researcher interested in studying the evolution of crude oil price returns. Yet, as mentioned in the introduction, an advantage of using a Markov switching model, is that it is well-suited to capture the evolution of time series processes that are characterized by periods of calm and periods of turmoil. Indeed, several studies suggest such models do a good job at tracking and forecasting the evolution of crude oil returns.

Now, the decision of whether to estimate a conventional or an endogenous regime switching model might seem less straight forward. Conventional regime switching models are well-understood and econometric toolboxes are readily available for their estimation. Nevertheless, as we illustrate in the following sections, a conventional regime switching model does not allow feedback from past oil returns into the transition probabilities. As such, it does not provide researchers with an effective tool to infer the likelihood of transitioning to a different state when the current state is known. Thus, in this and the following section we examine the performance of endogenous versus exogenous Markov-switching models in modeling the behavior of WTI oil price returns and illustrate the advantages of allowing the transition probabilities to evolve in an endogenous manner by focusing on well-known periods of turmoil in the oil market.

Yet, before we delve into the estimation results in Section 3, we review the model with
endogenous regime switching proposed by Chang, Choi and Park (2017) as it pertains the
dynamics of oil price returns and compare it to the conventional Markov switching model
where the probability of switching between regimes is exogenous. Then, we describe the data
used in the empirical analysis and briefly describe how the model is estimated.

2.1 Endogenous versus Exogenous Regime Switching Models

To gain insight about how the endogenous regime switching model differs from the con-
ventional Markov switching model, and to understand the issues faced by a researcher who
would rely on the latter, let us start by considering a simple case. Suppose we were interested
in modeling switches in the volatility of oil price returns between periods when the economy
is in a state of turmoil, \( s_t = 1 \), or one of calm, \( s_t = 0 \), depending on whether the level of
a latent factor \( w_t \) is above or below a threshold \( \tau \). Let \( y_t \) denote the returns in time \( t \) and
assume volatility is higher when \( s_t = 1 \), so that \( \sigma(1) > \sigma(0) \). We may then model the process
for oil returns as

\[
\begin{align*}
    y_t &= \mu + \sigma(s_t)u_t \\
    s_t &= 1 \{ w_t \geq \tau \} \\
    w_t &= \alpha w_{t-1} + v_t, |\alpha| < 1 \\
    \begin{pmatrix} u_{t-1} \\ v_t \end{pmatrix} &\sim i.i.d. N \left( 0, \begin{pmatrix} 1 & \rho_{v,u} \\ \rho_{v,u} & 1 \end{pmatrix} \right)
\end{align*}
\]

(1)

where the latent factor, \( w_t \), follows an autoregressive process of order one, \( u_{t-1} \) and \( v_t \) are
jointly normal. We will refer to this model as the volatility switching model. Note that
in the conventional Markov switching model \( s_t \) is a Markov chain and is assumed to be
independent of the observed oil price returns, \( y_t \), at all leads and lags. In contrast, here
the autoregressive latent factor \( w_t \) drives the state process \( s_t \).\(^3\) More precisely, the state
switches from calm to turmoil if the latent factor exceeds the threshold \( \tau \) and the unknown
parameter \( \rho_{u,v} \) governs the endogeneity of the regime changes. For instance, a negative \( \rho_{u,v} \)
would indicate that the innovation \( u_t \) of the crude oil returns, \( y_t \), at time \( t \) is negatively
correlated with the volatility of the returns at time \( t + 1, \sigma_{t+1} \). Hence, finding \( \rho_{u,v} < 0 \) would
provide evidence of a leverage effect whereby a negative shock to crude oil returns at time
\( t \) would lead to increased volatility in the following period. When \( \rho_{v,u} = 0 \), the model in
(1) is equivalent to the conventional Markov switching model for volatility; conversely, the

\(^3\)Given the assumption that \( |\alpha| < 1 \), the latent regime factor \( w_t \) is stationary.
endogeneity becomes stronger the larger the magnitude of $|\rho_{v,u}|$. The endogeneity given by $\rho_{u,v}$ is not contemporaneous but consequential and provides dynamic feedback.

As Chang, Choi and Park (2017) – hereafter CCP – show, the model in (1) can be generalized to allow both the volatility $\sigma$ and the mean, $\mu$, to switch across regimes. For instance, if we continue to assume that $\sigma(1) > \sigma(0)$ and allow the mean to switch across regimes, but without imposing any restriction on which mean is higher (i.e., $\mu(0) \neq \mu(1)$), we can model oil price returns as

$$y_t - \mu(s_t) = \sum_{k=1}^{p} \gamma_k (y_{t-k} - \mu(s_{t-k})) + \sigma(s_t) u_t$$  \hspace{1cm} (2)$$

with $s_t$, $w_t$ and $(u_{t-1}, v_t)'$ specified as in (1). In this specification, crude oil returns are modeled as an autoregressive process with both conditional mean and volatility that are state dependent. Furthermore, switches in the mean and volatility take place in a synchronized manner as both are governed by the same latent factor, $w_t$. The reader may recognize this model as a generalized version of the Markov switching model proposed by Hamilton (1989), where the regime switches are endogenized via the covariance between $u_{t-1}$ and $v_t$, $\rho_{u,v}$.

A-priori, there is no reason to believe that switches in the mean and volatility should be synchronized. For instance, after the oil price collapse of 1986, crude oil returns were below the historical mean while volatility was high. Volatility also skyrocketed during the Covid-19 pandemic, whereas returns plummeted. In contrast, at the beginning of the Persian Gulf war both the mean and volatility of oil returns increased. This would suggest considering an encompassing model, where the mean and volatility of crude oil returns are allowed to switch in an unsynchronized manner so that four regimes for oil price returns are possible: (1) low-volatility, low-mean; (2) low-volatility, high-mean; (3) high-volatility, low-mean; and (4) high-volatility, high-mean. Such a model is given by

$$y_t - \mu(s_{m,t}) = \sum_{k=1}^{p} \gamma_k (y_{t-k} - \mu(s_{m,t-k})) + \sigma(s_{v,t}) u_t$$  \hspace{1cm} (3)$$

The parameters $\mu_{s_{m,t}}$ and $\sigma_{s_{v,t}}$ denote the time-varying conditional mean and volatility of the oil returns that depend on two distinct, but correlated state processes $s_{m,t}$ and $s_{v,t}$. The process $s_{m,t}$ specifies the binary state of the mean, with $s_{m,t} = 0$ and 1 respectively representing low and high mean states. Similarly, the binary volatility state process $s_{v,t}$ specifies low and high volatility states with $s_{v,t} = 0$ and 1.\(^4\)

\(^4\)For regime identification, we assume $\mu(0) < \mu(1)$ and $\sigma(0) < \sigma(1)$.
We specify each of the state processes $s_{i,t}$ for $i = m, v$ as

$$s_{i,t} = 1\{w_{i,t} \geq \tau_i\}$$

with a latent regime factor $w_{i,t}$ and a threshold $\tau_i$. The mean (volatility) regime factor $w_{m,t}$ ($w_{v,t}$) determines the switch between states of low and high oil return mean (volatility) according to whether it is below or above the threshold $\tau_m$ ($\tau_v$). We let $w_t = (w_{m,t}, w_{v,t}^{'})$ and jointly consider the dynamics of the two latent factors by assuming they follow a first-order stationary bivariate autoregressive process

$$w_t = Aw_{t-1} + v_t$$

where

$$A = \begin{pmatrix} a_{mm} & a_{mv} \\ a_{vm} & a_{vv} \end{pmatrix},$$

the modulus of all eigenvalues of $A$ is less than unity, the innovations $v_t = (v_{m,t}, v_{v,t}^{'})$ are independent and identically distributed over time and correlated with the previous oil return innovation $u_{t-1}$. Specifically, we assume $(u_{t-1}, v_t^{'}) \sim i.i.d. \mathcal{N}(0, P)$ with a correlation matrix

$$P = \begin{pmatrix} 1 & \rho_{vu} \\ \rho_{vu} & P_{vv} \end{pmatrix} = \begin{pmatrix} 1 & 1 \\ \rho_{v_m,u} & \rho_{v_m,v} \end{pmatrix}$$

where variances are normalized for identification.

To get a better grasp of what this generalization entails for the dynamics of crude oil returns, let us examine the role of the new parameters. The parameters in the AR coefficient matrix $A$ and the correlation matrix $P$ jointly determine dynamics of the regime factors $w_{m,t}$ and $w_{v,t}$ which in turn determine mean and volatility regimes and their interactions. As in the simpler volatility model, here the latent factors drive the transition between regimes and the endogeneity of the regimes hinges on the correlation parameters $\rho_{u,v_m}$ and $\rho_{u,v_v}$. The two parameters $\rho_{u,v_m}$ and $\rho_{u,v_v}$ represent the aforementioned dynamic feedback. The main difference is that in this more general case we have two different, but not mutually independent, latent factors driving the regimes in the mean and the volatility.

Note that the evolution of the bivariate latent regime factor $w_t = (w_{m,t}, w_{v,t}^{'})$ is driven by the innovations $v_{m,t}$ and $v_{v,t}$, which we collect in the vector $v_t = (v_{m,t}, v_{v,t}^{'})$ as well as by the dynamic interaction between the two factors, which is determined by the autoregressive coefficient matrix $A$. In particular, if $a_{mv} \neq 0$, then the volatility regime factor helps to predict the mean regime factor. Conversely, if $a_{vm} \neq 0$, then the mean regime factor helps to
predict the volatility regime factor. In addition, larger values of \(a_{mm}\) and \(a_{vv}\), respectively, would indicate higher persistence in the mean and volatility regime factors.

We allow the volatility and mean regime factors to be correlated via \(\rho_{v_m,v_v}\), and model feedback from past return innovations, \(u_{t-1}\), to the mean and volatility factors through \(\rho_{u,v_m}\) and \(\rho_{u,v_v}\). We can also measure the contemporaneous correlation between the factor innovations \(v_{m,t}\) and \(v_{v,t}\) net of the contribution of \(u_{t-1}\) as \(\rho_{v_m,v_v} = \rho_{v_m,v_v} - \rho_{v_m,u}\rho_{v_v,u}\). To examine a null hypothesis of no contemporaneous or dynamic interaction between the mean and volatility regime factors, we can then test whether \(\rho_{v_m,v_v} = 0\) or \(a_{vm} = a_{mv} = 0\).

Finally, in this generalized setup, endogenous feedback may occur through two different channels. Shocks to crude oil returns at time \(t\) would affect the regime switching in the mean at time \(t+1\) if \(\rho_{u,v_m} \neq 0\). Furthermore, shocks to crude oil returns at time \(t\) could also affect the regime switching in the volatility at \(t+1\) if \(\rho_{u,v_v} \neq 0\). Clearly, these feedback channels are absent in the conventional Markov switching model. The null hypothesis of no-feedback can be tested using the standard likelihood ratio test.

### 2.2 Data and Empirical Methodology

To study the behavior of crude oil price returns we use the monthly spot price of the West Texas Intermediate (WTI) spanning the period between January 1986 and January 2021. We compute returns as the first difference of the natural log of the spot price. We focus on the WTI price as it is commonly used as the reference price for buyers and sellers of crude oil in the U.S., it is produced in the U.S. and – until late 2015 when the export ban on U.S. crude oil was lifted – was only sold in the U.S. Nevertheless, we note that estimation results for the endogenous Markov switching models are robust to using Brent crude oil returns.\(^5\)

Crude oil prices, as other financial assets, often experience abrupt changes in behavior. As Figure 1 illustrates, the price of the WTI, denoted by the blue line, exhibited an upward trend during the 1990s and most of the 2000s, reaching a peak of $139.96 per barrel in June 2008. Then, it plunged during the Great Recession to a trough of $41.96 per barrel in January 2009, slowly recovering during the first half of the 2010s until it collapsed in July of 2014. A striking decline was also observed in the wake of the global shutdown induced by the Covid-19 pandemic.\(^6\)

\(^5\)For the sake of brevity, we relegate the results for Brent oil returns to the online appendix.

\(^6\)Figure A.1 of the online appendix reveals a similar behavior for Brent oil returns.
In addition, the series of WTI returns depicted in Figure 1 suggest periods of political and economic turmoil might be associated with higher volatility. See, for instance, the period surrounding the Persian Gulf War (August 1990 to February 1991), the global financial crisis in the late 2000s, and at the onset of the Covid-19 pandemic. Yet, periods of high volatility do not always coincide with periods of high returns. For instance, both oil returns and their volatility were high during the period following the outbreak of the Persian Gulf war; however, the 2008 financial crisis was characterized by low returns and high volatility.\footnote{The same periods of high volatility are observable for Brent returns. See Figure A.1 of the online appendix.}

The endogenous regime switching models in (1), (2), and (3) can be estimated via maximum likelihood. We refer the reader to CCP for a detailed description of the estimation procedure for regime switching models with a single latent factor such as our simple volatility switching model (1) and the synchronized mean-volatility switching model (2). For the estimation of our general unsynchronized mean and volatility switching model (3) with joint dynamics of the two latent factors, we use the algorithm and filter developed for regime-switching models with multiple latent regime factors by Chang, Park and Qiu (2020).

Recall that in the conventional model, switches between regimes are completely determined by the constant transition probabilities of the single state process \((s_t)\), and the switching evolves according to a Markov chain. When the state is correlated with the observed process, namely crude oil returns, a modified filter is needed to explicitly account for this endogenous feedback channel that results in time-varying transition probabilities. Hence, we use the Chang-Park-Qiu filter, which extends the CCP filter to a multivariate setup. As in the standard Kalman filter, the modified filter involves the usual prediction and updating steps, but they are carried out with time-varying transition probabilities.

To illustrate how the endogenous feedback gives rise to time-varying transition probabilities, first net out the effect of the past return innovation \(u_{t-1}\) from the regime factor innovations \(v_t\) to obtain the orthogonal regime factor shock \(\varepsilon_t = v_t - \rho_{vu}u_{t-1} \sim N(0, P_{vv\cdot})\). Then, we may use \(\varepsilon_t\) to compute the transition probability, for example, from the low mean-high volatility regime to the low mean-high volatility regime (LH-to-LH), or staying in low mean-high volatility regime, as

\[
P\{s_t = (0, 1)'|s_{t-1} = (0, 1)', \mathcal{F}_{t-1}\} = \Phi(\tau)^{-1} \int_{-\infty}^{\tau} \Phi_{v|u}(\tau - \rho_{vu}u_{t-1} - Aw_{t-1}) \phi(w_{t-1}) dw_{t-1}
\]
where \( s_t = (s_{m,t}, s_{v,t})' \), \( \Phi_{v,u} \) is the distribution function of the regime factor shock \( \varepsilon_t \), \( \tau = (\tau_m, \tau_v)' \) and \( \mathcal{F}_{t-1} \) is the information set available at time \( t - 1 \) given by the past oil returns, \( y_{t-1}, y_{t-2}, \ldots, y_1 \). The vector \( \rho_{vu} u_{t-1} = (\rho_{vm,u} u_{t-1}, \rho_{vv,u} u_{t-1})' \) contains the feedback effects from \( u_{t-1} \) to the mean regime factor \( w_{m,t} (\rho_{vm,u} u_{t-1}) \), and to the volatility regime factor \( w_{v,t} (\rho_{vv,u} u_{t-1}) \). These feedback channels make the otherwise constant transition probabilities time-varying. This shows how the correlation parameters contribute to the final impact of the past return innovation \( u_{t-1} \) on the transition probabilities. The final impact of a one-unit realized return innovation \( u_{t-1} \) therefore depends on the sign and magnitude of its correlation with the mean and volatility regime factor innovations \( v_{m,t} \) and \( v_{v,t} \). Together they influence the regime determination process by effectively lowering or raising the threshold. When the feedback channels are shut down, \( \rho_{vu} = 0 \), then the model reduces to the conventional Markov switching model with constant transition probabilities.

3 Regime Switching in the Volatility and Mean of Crude Oil Returns

As we will see later, our estimates clearly identify two separate latent factors for the mean and volatility of crude oil returns. Nevertheless, to build some intuition on why the behavior of crude oil returns may be better captured by an endogenous regime switching model, we first discuss the results when only the volatility is allowed to switch across regimes (model (1)). Parameter estimates with 68% confidence intervals\(^8\) from the simple model with volatility switching only are reported in the first and second columns of Table 1. They show that two regimes, low and high volatility, are clearly identified with the volatility during periods of turmoil (\( \sigma_1 = 0.222 \)) being more than three times as large as the volatility in periods of calm (\( \sigma_0 = 0.069 \)) and with a latent factor that is very persistent (\( \alpha = 0.971 \)). In addition, the negative and significant coefficient on \( \rho_{v,u} \) is evidence of a strong leverage effect: the estimate of \( \rho_{v,u} < 0 \) (\( -0.974 \)) indicates that a negative shock to crude oil returns in period \( t \) implies an increase in volatility in period \( t + 1 \).\(^9\)

\(^8\)Confidence intervals are obtained using the stationary block bootstrap procedure by Politis and Romano (1994). We obtain percentile bootstrap confidence intervals by estimating 500 block bootstrapped samples of length 420 (404 for the Brent oil returns). The average block size is 17 for the stationary block bootstrap, which is selected by averaging the optimal block size for each vector of time series.

\(^9\)Estimation results for Brent returns can be found in Table A.1 of the online appendix.
As Figure 2 illustrates, several periods of high volatility (shaded in gray) are identified during the period under analysis. These periods correspond to the times when the extracted latent factor exceeds the threshold, $\tau$ (red line). Three features stand out. First, regimes with high volatility are recurrent, but short-lived. We observe 29 months (out of 419 observations) in the high volatility regime, which appear to be concentrated in five different periods. Second, not surprisingly, periods of political unrest in key oil producer countries (e.g., the invasion of Kuwait) and economic contractions (e.g., the 2008 Great Recession and the 2020 Great Shutdown) constitute periods of heightened risk in the oil market. Finally, of special interest are two other episodes related to recent developments in the oil market: the increased financialization of the oil market in the early 2000s\textsuperscript{10} and the 2014 collapse. The latent factor increases only slightly and for a brief period of time around the financialization of the oil market (but does not approach the threshold for a high volatility regime); yet it raises significantly and for a prolonged period of time after the 2014 oil price collapse. The latter clearly reflects a period of heightened volatility and increased risk.\textsuperscript{11}

FIGURE 2 HERE

Figure 3 puts in evidence the difference between the transition probabilities estimated from the endogenous and exogenous regime switching models. The left panel reports the probability of staying in the high volatility state and the right panel illustrates the transition probability from the low to the high volatility state. The black solid line represents the time-varying transition probability estimated from the endogenous switching model, and the red dashed line corresponds to the constant transition probability from the exogenous regime switching model. Note that, in the exogenous model, the probability to stay in the high regime remains constant at 0.86. In contrast, in the endogenous model the probability varies over time with the realized values of the crude oil returns and differs significantly across various periods of high volatility. For instance, on the one hand, the probability of remaining in the high volatility regime drops to 0.39 and 0.27 during the Persian Gulf War and the Covid-19 Pandemic, respectively, suggesting both events would lead only to a temporary increase in volatility. On the other hand, the endogenous time-varying transition probabilities remained high during the Great Recession and the 2014 oil price collapse.\textsuperscript{10} The presence of financial investors has increased considerably since the early 2000s. Financial players without an interest in holding physical crude oil (e.g., pension funds, hedge funds, insurance companies) have since held larger positions in derivatives and futures markets. These developments have led to a heated debate regarding the role of financial speculation in driving oil price volatility.\textsuperscript{11} Estimation results lead us to identify almost identical periods of high volatility for Brent returns. See Figure A.2 of the online appendix.
Perhaps more striking is the difference between the transition probability from the low to the high-volatility regime in the exogenous and endogenous model. Whereas in the exogenous switching model the low to high transition probability would have remained constant at 0.021, the endogenous model reveals low to high transition probabilities that vary significantly over time and exceed 0.15 around the invasion of Kuwait, during the Great Recession, and when oil prices plunged after the second half of 2014.

Of particular interest are the transition probabilities during the Covid-19 pandemic: the low-to-high transition probability increased to about 70% at the onset of the global shutdown while the high-to-high probability declined below 30%. This behaviour suggests a period of very high uncertainty in oil markets. As we will see later, this impression is corroborated by the behavior of the oil market volatility index produced by the Chicago Board Options Exchange (the CBOE OVX).

**FIGURE 3 HERE**

Furthermore, the first panel of Figure 3 illustrates how, once crude oil returns entered a period of high volatility (denoted by the gray shaded area), the probability of remaining in the high-volatility state quickly declined for all episodes, but the Great Recession. In fact, only for the period corresponding to the Covid-19 pandemic did the endogenous switching model estimate a high-to-high probability that temporarily exceeded the estimate obtained from the exogenous model. This coincides with a period in which the WTI returns remained negative for several months (see Figure 1).\(^{12}\)

We conclude this section by noting that the information contained in the time-varying transition probabilities could be useful for policy analysts and investors. For instance, during periods of turmoil when volatility is high, an endogenous regime switching model could aid in assessing the risk of remaining in the high volatility state. Regardless of whether we use the WTI or Brent to measure crude oil returns, the endogenous switching model provides useful information regarding oil market risks possibly associated with political unrest and economic downturns reflected already in the oil price returns. We will turn back to this issue in Section 4.

\(^{12}\)Similar results are found when using the Brent crude oil price. See Figure A.3 of the online appendix.
3.1 Unsynchronized Switching in the Mean and Volatility of Crude Oil Returns

The estimation results presented in the previous section suggest that an endogenous regime switching model does a good job at capturing changes in the volatility of oil price returns. A natural question, however, is whether the behavior of oil price returns is better captured by model (1) or by a model that also allows the mean to switch as in (2) or (3). Indeed, as Figure 1 suggests, oil price returns exhibit periods of high volatility and low mean as well as high volatility and high mean. To answer this question, we start by estimating the endogenous regime switching model in (3) where the switches in mean and volatility are driven by two different – yet correlated – latent factors, and then proceed to test a series of hypotheses that speak to the fit of the general model.

The third column of Table 1 reports the estimates (with 68% confidence intervals in the fourth column) from the general model where the conditional mean and volatility are allowed to switch in an unsynchronized manner.\(^\text{13}\)

The estimates for \(\sigma_0\) and \(\sigma_1\) are similar to those obtained in the simpler model with volatility switching only and also indicate that the volatility in crude oil returns is more than three times as large in the high-volatility regime \((\sigma_1 = 0.219)\) compared to that in the low-volatility regime \((\sigma_0 = 0.059)\). The fact that the estimates of \(\sigma_0\) and \(\sigma_1\) do not change much from model (1) indicates that the difference in the volatility states is not driven by the mean.

We find evidence of endogeneity both for the volatility and the mean regimes. The estimate of \(\rho_{u,v}\), which measures the degree of endogenous feedback to the volatility regime switching, is negative and significant \((-0.917)\), that of \(\rho_{u,v_m}\), which accounts for endogeneity in the mean regime, is positive and significant \((0.529)\).

\textbf{FIGURE 4 HERE}

Figure 4 reports the extracted latent factors for the mean (left panel) and volatility (right panel). The left panel illustrates how the mean latent factor is not very persistent, which is consistent with the estimate of \(a_{mm} = 0.133\). Yet, it remains below the threshold for a few months after the end of the Persian war and then again during the Great Recession, the 2014 oil price collapse, and the Great Shutdown. About 91% of the observations in the sample belong in the high-mean regime. As for the volatility latent factor, the right panel of

\(^{13}\text{Results for a model where the mean and the volatility switch in a synchronized manner are available from the authors upon request}\)
Figure 4 is roughly consistent with the factor extracted from the simpler volatility switching model (see Figure 2), although the high volatility regime appears to be slightly less prevalent here (5% of the observations are in the high-volatility regime versus 7% in the model with volatility switching only).

Given the similarity between the results for the simpler volatility model and the model with unsynchronized mean-volatility switching, the reader may wonder how the mean and volatility latent regime factors are related. To explore this issue we report the 24-month rolling window correlations among the latent factors as well as the coherence (see Figure 5). Three features are noticeable in this figure. First, the correlation is negative throughout the sample but exhibits a large degree of time variation. Second, although the correlation between the latent factors drops significantly during the Gulf War, it remains high throughout the Great Recession, during the 2014 oil price collapse and the Great Shutdown. Third, the coherence plot indicates that the co-movement among the two factors is accounted for, slightly more, by lower than higher frequencies, especially the frequencies corresponding to periods longer than three months. Very similar patterns emerge when we redo the analysis using Brent returns (see Figure A.6 of the online appendix).

3.2 Endogenous Feedback, Granger Causality and Exogenous Regime Switching

To evaluate the fit of the unsynchronized mean-volatility regime switching model, we first test the null hypothesis of no endogenous feedback from the shock to crude oil returns at time \( t \) to the regime switching in mean and volatility latent factors at time \( t + 1 \) (i.e., we test the null \( \rho_{u,v_m} = \rho_{u,v_v} = 0 \)). However, we allow the mean and volatility latent factors to be contemporaneously correlated and interact dynamically through the autoregressive parameters in the matrix \( A \) and the correlation parameter \( \rho_{v_v,v_m} \). The likelihood ratio test for this hypothesis equals 80.46, thus allowing us to reject the null at a 1% significance level and providing strong evidence that endogenous feedback plays an important role in modeling the dynamics of crude returns.

\[ \rho_{m,v}(\lambda) = \frac{|f_{m,v}(\lambda)|^2}{f_{m,m}(\lambda)f_{v,v}(\lambda)}, \]

where \( f_{m,m}(\lambda) \) and \( f_{v,v}(\lambda) \) are the spectral densities of \( w_{m,t} \) and \( w_{v,t} \), and \( f_{m,v}(\lambda) \) the cross-spectral density between \( w_{m,t} \) and \( w_{v,t} \). These spectral and cross-spectral densities are the components of the spectral density matrix of the bivariate regime factor \( w_t = (w_{m,t}, w_{v,t})' \) given by

\[ F_w(\lambda) = A^{-1}(e^{i\lambda})F_v(\lambda)A^{-1}(e^{-i\lambda})^*, \quad \lambda \in [-\pi, \pi] \]

where \( F_v(\lambda) \) is the (2 \( \times \) 2) spectral density matrix of the regime factor innovations \( v_t \), and \( * \) denotes the adjoint operator. Due to the iid assumption on \( v_t = (v_{m,t}, v_{v,t})' \), we have \( F_v(\lambda) = P_{vv} \) for all \( \lambda \in [-\pi, \pi] \).

---

14 Figure A.5 of the online appendix illustrates the results for the Brent returns.

15 The coherence between the mean and volatility regime factors, \( w_{m,t} \) and \( w_{v,t} \), is computed as

\[ \rho_{m,v}(\lambda) = \frac{|f_{m,v}(\lambda)|^2}{f_{m,m}(\lambda)f_{v,v}(\lambda)}, \]

where \( f_{m,m}(\lambda) \) and \( f_{v,v}(\lambda) \) are the spectral densities of \( w_{m,t} \) and \( w_{v,t} \), and \( f_{m,v}(\lambda) \) the cross-spectral density between \( w_{m,t} \) and \( w_{v,t} \). These spectral and cross-spectral densities are the components of the spectral density matrix of the bivariate regime factor \( w_t = (w_{m,t}, w_{v,t})' \) given by

\[ F_w(\lambda) = A^{-1}(e^{i\lambda})F_v(\lambda)A^{-1}(e^{-i\lambda})^*, \quad \lambda \in [-\pi, \pi] \]

where \( F_v(\lambda) \) is the (2 \( \times \) 2) spectral density matrix of the regime factor innovations \( v_t \), and \( * \) denotes the adjoint operator. Due to the iid assumption on \( v_t = (v_{m,t}, v_{v,t})' \), we have \( F_v(\lambda) = P_{vv} \) for all \( \lambda \in [-\pi, \pi] \).
In addition, the estimated autoregressive parameters, more specifically the off-diagonal elements of the matrix $A = \begin{bmatrix} a_{mm} & a_{mv} \\ a_{vm} & a_{vv} \end{bmatrix}$, reveal a significant dynamic interaction between the mean and volatility latent factors. Note that estimation results reported in the second panel of Table 1 lead us to reject the null that the volatility factor does not Granger cause the mean factor ($a_{mv} = 0.133$, and statistically significant) as well as the null of no Granger-causality from the mean factor to the volatility factor ($a_{vm} = -2.177$, and statistically significant). It is also worth noting that our estimate of the contemporaneous correlation between the innovations $v_m$ and $v_v$ net of the contribution of the past return innovation $u_{t-1}$, which can be computed as $\rho_{vv,u} = \rho_{v_v,v_m} - \rho_{v_m,u} \rho_{v_v,u}$, equals $-0.339$. This indicates a negative contemporaneous comovement between the mean and volatility factors, even after netting out the effect of past innovations in crude oil returns. We conclude this section by noting that, estimation results not reported herein, but available from the authors upon request indicate that: (a) the model where mean and volatility switching occurs in a synchronized manner as in (2), is rejected in favor of the unsynchronized model (3); (b) our results are robust to estimating the model after excluding the three largest outliers (March, April and May of 2020); and (c) our conclusions are unchanged when we estimate the model using Brent instead of WTI returns.

3.3 Lessons from Four Episodes of High Volatility and Low Returns

The previous section shows how a model where the mean and volatility switch between regimes in an unsynchronized manner does a good job at capturing the behavior of crude oil returns. But, what lessons can oil investors and policy makers derive from a model that allows the regime to evolve in an endogenous manner? At a first glance, gauging the policy implications or empirical relevance of our estimates might seem cumbersome, especially given the multiplicity of states and transition probabilities. Thus, we focus on four historical episodes that an analyst or policy maker may have considered as particularly risky for oil markets given the WTI returns exhibited a low mean and high volatility: the Persian Gulf war, the 2008 financial crisis, the oil price collapse of 2014 and the Covid-19 pandemic. Doing so allows us to distill a clear message: during periods of turmoil, the transition probabilities provide a more realistic assessment of the likelihood of remaining in a known low mean-high-volatility regime than the constant transition probability derived from an exogenous model.

More specifically, given that monthly WTI returns are rather small, we classify a month as
exhibiting a low mean if the monthly return is negative and lower than ten times the average of the returns over the previous 32 months from the start of the particular low-mean-high-volatility regime. Similarly, we classify a month as exhibiting high volatility if the volatility is more than twice the average volatility computed over the 32-month period ending at the start of the given low-mean-high-volatility regime. Months that fall in both categories are classified as low-return-high-volatility. Note that the source of the disruptions that led to heightened volatility and low returns was distinct and varied across the four episodes, thus providing a good way to illustrate the empirical relevance of our results. Moreover, of particular interest is the sample covering the Covid-19 pandemic, which resulted in the oil market experiencing an “all-time high volatility” according to U.S. Energy Information Administration, 2020.

Figure 6 plots the time-varying transition probabilities from a low mean-high volatility to a low-mean-high volatility (hereafter LH-to-LH) regime and the WTI returns. The top left panel depicts the evolution during the Persian Gulf War. Recall from Figure 1(b) that the WTI returns and its volatility increased dramatically when Iraq invaded Kuwait in August 1990; yet, they later decreased reaching a trough in March 1991. In turn, the volatility increased with the onset of the war and remained high until Iraq accepted the terms of the cease-fire agreement (March 3, 1991). As illustrated by the black solid line, the LH-to-LH transition probability increased with Iraq’s invasion of Kuwait and remained relatively high until the U.S. and allied forces entered Kuwait at the end of February 1991. Contrast the time-varying transition probability with the transition probability obtained from the exogenous switching model denoted by the red dashed line. In the exogenous model, the transition probability remains constant at 0.524. Our time-varying transition probabilities are considerably lower, although they rise throughout the war.

FIGURE 6 HERE

Three additional episodes of low-return and high-volatility regime are depicted in Figure 6. The top right panel illustrates the financial crisis of 2008, the bottom left panel corresponds to the oil price collapse of 2014 and the right bottom panel depicts the Great Shutdown resulting from the Covid-19 pandemic. In all cases, the time-varying nature of the transition probability estimated from the endogenous model stands in sharp contrast with the constant transition probability obtained from the exogenous model. The LH-to-LH transition probability increases with the decline in crude oil returns and stays high until the monthly returns start to increase. However, whereas the time-varying probability from our endogenous model remained below the constant probability from the exogenous model for most of
the sample, it did not for two episodes. First, during the financial crisis the time-varying transition probability was very close to the probability from the exogenous model. Second, and more notably, in March and April of 2020 when shutdowns and border closures were implemented in the wake of the Covid-19 pandemic, the time-varying transition probability from the endogenous model surpassed the constant transition probability from its exogenous counterpart and then quickly declined in May 2020. The latter coincided with expectations of improved global demand for crude oil and an agreement between OPEC and its allies to cut production. Clearly, transition probabilities derived from an endogenous regime switching model provide a more realistic assessment of the likelihood to remain in the known low return-high volatility regime. In other words, a lesson to be learned from these four episodes is that not all periods of low return and high volatility should convey the same degree of “fear” to economic agents.16

Finally, we note that the extracted latent factors contain valuable information for policymakers and investors. To illustrate this point, Figure 7 plots the extracted latent volatility factor from our mean-volatility switching model and the Chicago Board Options Exchange (CBOE) Crude Oil ETF Volatility Index (OVX), retrieved from FRED at the Federal Reserve Bank of St. Louis. The OVX measures the 30-day implied volatility of crude oil prices and is computed using fluctuations of the prices of financial options for the WTI. The OVX inception dates from May 10, 2007 and thus covers only part of our sample, but is increasingly cited as a measure of expected volatility in oil markets (see e.g., Energy Information Agency 2020). Note that the extracted volatility factor evolves in a manner similar to the OVX, with both series increasing during the financial crisis, the 2014 oil price collapse and the Great Shutdown. Fluctuations in the OVX somewhat leads fluctuations in the extracted volatility factor as one would expect given that the OVX is a measure of near-term price changes in the WTI. We also note that the extracted latent factor dropped faster than the OVX in 2020, suggesting that the increase in oil market volatility was shorter lived than the markets had originally expected. In brief, the extracted latent volatility factor could be employed by economic agents and investors as an alternative measure of the overall risk or stress in the oil market.17

16See Figure A.4 of the online appendix for time-varying probabilities and Brent returns for these four low return-high volatility periods. 17Similar results have been found by Chang et al. (2017) for the VIX and the volatility latent regime factor extracted from the U.S. excess market returns.
4 Understanding the Mean and Volatility Factors

So far, we have found some suggestive evidence that switches in the mean and volatility regimes coincide with the business cycle. Moreover, we have shown that the volatility latent factor can be interpreted as a measure of stress in oil markets. The question to which we now turn is whether the evolution of these latent factors can be formally linked to fluctuations in macroeconomic, financial and oil market variables.

4.1 Latent Factors and “Big Data”

The idea of using a large number of time series to understand fluctuations in crude oil prices is relatively new, despite it having a long history in the macroeconomic literature (see Burns and Mitchell 1946, Friedman 2009, Stock and Watson 2017). While a few papers have used large data sets to forecast crude oil prices (see e.g., Lu et al. 2020, Zhang et al. 2019), to the best of our knowledge, big data sets have not been used to explore what drives the behavior of crude oil prices. Thus, in this section, we empirically investigate how the estimated latent factors are related to macroeconomic and financial market information available to policy makers and economic agents. To do so we consider a large number of time series that span different categories such as U.S. macroeconomic variables, oil market indicators, uncertainty and global economic activity indicators. Our benchmark data set comprises the following groups of variables.

Global Crude Oil Market. To capture conditions in the global oil market we use data from the Energy Information Agency. We employ the log growth in world crude oil production18 to measure oil supply. To proxy for global crude oil inventories, we use data on OECD petroleum inventories, U.S. crude oil inventories, and U.S. petroleum inventories. As in Kilian and Murphy (2014), Baumeister and Hamilton (2019) and Herrera and Rangaraju (2020), we compute the proxy for OECD crude oil as the ratio of OECD to U.S. petroleum inventories times the U.S. crude oil inventories.

World Economic Activity. Several indicators of global economic activity have been used in the oil literature to capture the effect of fluctuations in the global demand for crude oil on its price. We include three indices. Kilian (2009) proposed an index of global real economic activity derived from a panel of global bulk dry cargo rates, which has been extensively used in the literature. Baumeister and Hamilton (2019) use the industrial production index for OECD countries and the six major non-member economies (China, Brazil, India, Indonesia

---

18Production includes lease condensates (i.e., light liquid hydrocarbons).
the Russian Federation and South Africa). Finally, Baumeister, Korobilis and Lee (forthcoming) compute a monthly Global Economic Conditions indicator and show that this indicator is useful for forecasting oil prices and global petroleum consumption. We take an agnostic approach and include all indices in our data set, allowing the regularization procedure to select which index may be linked to the extracted latent factors.

**U.S. Macroeconomic Variables.** We use the large monthly data set FRED-MD developed by McCracken and Ng (2016), constructed and updated by the FRED data desk at the Federal Reserve Bank of St. Louis. This data base contains 134 monthly U.S. variables made available by the Federal Reserve Bank of St. Louis; it comprises macroeconomic aggregates, coincident and leading indicators that can be classified in the following groups: output and income, labor market, consumption and housing, orders and inventories, money and credit, interest rates and exchange rates, prices, and stock market. Information regarding detailed data construction and transformations applied to each time series can be found in McCracken and Ng (2016).¹⁹

**U.S. Economic Policy Uncertainty Indices.** To capture U.S. economic policy uncertainty we employ the categorical data constructed and updated by Baker, Bloom and Davis (2016), which includes several sub-indexes based on news data. These indices include the overall economic policy uncertainty, as well as fiscal policy, monetary policy, healthcare, national security, regulation, sovereign debt & currency crises, entitlement programs and trade policy uncertainty. Each of the indices could affect crude oil returns – and hence the latent factors – in a different manner.

### 4.2 Empirical Methodology

To address the problem of selecting the variables that drive the mean and volatility latent factors, we initially consider two regularization methods: adaptive LASSO (least absolute shrinkage and selection operator) and adaptive Elastic-net. The adaptive LASSO estimator proposed by Zou (2006) is given by

\[
\hat{\beta}_{AL} = \arg \min_{\beta} \| y - X\beta \|^2_2 + \lambda \sum_{j=1}^{p} \hat{w}_j |\beta_j| 
\]

where \( \{\hat{w}_j\}_{j=1}^{p} \) are data-driven weights, and \( \lambda \) is the penalty parameter which we choose by cross-validation. We compute the initial weights using the Ridge regression estimators

¹⁹Note that because we exclude the FRED-MD oil prices as well as variables with missing observations, the IDs reported in the Tables 2-3 do not correspond to the IDs in McCracken and Ng (2016).
\[ \hat{\beta}_{j}^{\text{ridge}} \] as \( \hat{w}_j = (|\hat{\beta}_{j}^{\text{ridge}}|)^{-\gamma} \) where the parameter \( \gamma \) is a positive constant, and also chosen by cross-validation.

As Zou (2006) shows, the adaptive LASSO has the advantage of allowing us to obtain a sparse solution while at the same time having an oracle property. However, in high-dimensional data when the variables are highly correlated, the adaptive LASSO might perform poorly (see Zou and Hastie 2005). For this reason, we also implement the adaptive Elastic-net of Zou and Zhang (2009); a procedure that can be viewed as a combination of Elastic-net, which was developed to tackle the issue of highly correlated variables but does not have an oracle property, and the adaptive LASSO. The adaptive Elastic-net is implemented by solving the following optimization problem:

\[
\hat{\beta}_{\text{AEn}} = \left(1 + \frac{\lambda_2}{n}\right) \left\{ \arg \min_{\beta} \|y - X\beta\|_2^2 + \lambda_2\|\beta\|_2^2 + \lambda_1^* \sum_{j=1}^{p} \hat{w}_j |\beta_j| \right\}
\]

where the data-driven weights \( \{\hat{w}_j\}_{j=1}^{p} \) here are constructed with the Elastic-net estimators \( \hat{\beta}_{j}^{\text{enet}}, j = 1, \ldots, p, \hat{w}_j = (|\hat{\beta}_{j}^{\text{enet}}|)^{-\gamma} \) where \( \gamma \) is a positive constant chosen by cross-validation, \( \lambda_2 \) is the mixing parameter, and \( \lambda_1^* \) is the penalty parameter.

The estimation procedure we use relies on the R package msaenet (Xiao and Xu, 2015): we estimate a Ridge regression in the first step and an Elastic-net in the second step. The parameter \( \gamma \) in the Ridge regression is chosen by cross-validation. In the second step, we use the multiplicative reciprocal of the first step coefficients as weights on the penalty. For the Elastic-net results, we set the mixing parameter for the quadratic regularization term to 0.2.\(^{20}\) The penalty parameter is then also chosen by cross-validation.

### 4.3 What can “Big Data” Tell Us About the Behavior of Crude Oil Prices?

We begin this section by noting that both regularization methods select the same set of variables for the mean and volatility factors. The only noticeable differences are slight divergences in the magnitudes attached to each variable. The variables selected by these two regularization methods for the extracted latent factors from the mean-volatility switching model are reported in Table 2.

Regarding the volatility factor, Panel A of Table 2 reports the twelve selected variables, as well as the group where they belong. Interestingly, all of the variables are included

\(^{20}\)The results are consistent across different values of the mixing parameter.
in the FRED-MD monthly variables for macroeconomic research. They comprise labor
market variables, orders and inventories, money and credit, prices and stock market variables.
These variables are tightly linked to production in the oil sector (employment in mining,
V31), indicators of economic activity (business inventories, V57; inventories to sales ratio,
V58), consumer sentiment (V110), indicators of the Federal Reserve action (real M2 money
stock, V62; St. Louis adjusted monetary base, V63), as well as prices and the stock market
composite index (S&P500, V109).

**TABLE 2 HERE**

Panel B of Table 2 lists the three variables selected by the two regularization methods for
the mean latent factor. Note that two of the variables, the business inventories to sales ratio
(V58) and the St. Louis adjusted monetary base (V63) – a variable that captures monetary
policy actions taken by the Federal Reserve – play an important role in driving both the
volatility and the mean extracted latent factors. Additionally for the mean factor, personal
consumption expenditure in durable goods (V106), an indicator of prices, is also selected by
the adaptive LASSO and the adaptive Elastic-net.

In brief, over the full sample for the benchmark database, two indicators are key in
explaining the variation in both extracted latent factors: the total business inventories to
sales ratio and the St. Louis adjusted monetary base. That is, changes in economic activity
and monetary policy appear to be correlated with the factors leading the transition between
high and low mean, and high and low volatility states. In addition, stock prices and labor
market conditions have an impact on the transition between high and low volatility regimes.

**TABLE 3 HERE**

### 4.4 Time-variation in the Relationship between Latent Factors
and Macroeconomic Variables

While the previous section puts in evidence the relevance of particular macroeconomic vari-
ables for the evolution of crude oil returns, the reader may wonder whether the importance
of some factors in explaining the mean and volatility factors could have changed over time.
After all, the sample used in our estimation covers a period where the crude oil market and
the global economy experienced substantial changes. Arguably, the increased financializa-
tion of the oil market since the early 2000s could have accounted for increased volatility\textsuperscript{21} and the fracking revolution may have contributed to changes in crude oil returns. While the Great Recession constituted a period of increased volatility and lower crude oil prices, it also brought about the use of monetary policy tools that had never been employed in the U.S. In other words, in the period of time analyzed in the previous sections, agents and policy makers could well have experienced changes in preferences, policies and technology.

To further investigate the possible time-variation that such changes could have implied and, having noted that the differences between the adaptive LASSO and adaptive Elastic-net are minimal, we re-estimate the adaptive LASSO regressions over a series of subsamples. More specifically, we construct 25 rolling windows of ten years (120 months) with the first window spanning the period between 1986:2 and 1996:2. We then advance the window by one year so that the second subsample covers the period between 1987:2 and 1997:2. We keep rolling the window by a year until the last subsample: 2000:2-2020:2. Then, we use the adaptive LASSO to select the variables that explain the mean or the volatility latent factor in each of the subsamples.

Figure 8 depicts the results of this rolling window exercise. For each sample ending on the month noted in the horizontal axis, we plot the ratio of the $\hat{\beta}$ (AL) coefficients on the selected variables relative to the largest coefficient in the subsample. The darker the shaded area associated with a variable, the greater its relative importance in explaining the factor in the corresponding subsample.

FIGURE 8 HERE

Four insights stand out from the heat map on the top panel of Figure 8 (Volatility) . First, not surprisingly, some key variables selected for the whole sample are also selected in several subsamples. For instance, this is the case of total business inventories to sales ratio (V58) and the S&P500 composite index. Second, the relevance of different macroeconomic factors varies over time. In particular, total non-farm employment (V29) is ranked higher – relative to the size of the largest coefficient – in the samples that predate the 2001 recession and the onset of the financialization of the oil market around 2003, but is not selected during most of the 2000s. Yet, it becomes increasingly relevant by the end of the Great Recession. In contrast, the S&P500 composite index (V109) is key in explaining the volatility factor during the 2000s. As for employment in mining (V31), it is selected in most subsamples that include

\textsuperscript{21}Yet, work by Fattouh, Killian and Mahadeva (2013) finds no evidence that the surge in the real price of oil during the 2003-2008 period was due to increased financialization of oil futures markets. Instead, they show that economic fundamentals (e.g., increases in global demand) account for the price increase.
the fracking revolution and gains importance during the high volatility period that ensued the 2014 collapse in oil prices. Third, since the Great Recession, additional variables reflecting changes in aggregate demand (e.g., the inventory to sales ratio, V58), employment in financial activities (V40) and manufacturing (V3), and changes in prices (V106) have gained relevance in explaining the volatility factor. Fourth, an interesting result that emerges from the heat map, is the contribution of specific variables during periods of turmoil. Monetary policy uncertainty (V115) is selected around the time when the U.S. economy hit the zero lower bound and indicators of output and income, labor market conditions and prices are selected in the sample that contains the Covid-19 pandemic.

The results for the mean factor using the rolling window scheme described above are reported in the bottom panel of Figure 8. As in the case of the volatility factor, we observe (a) the prevalence of the variables selected for the whole sample in the majority of subsamples (see for instance V63, St. Louis adjusted monetary base), (b) time variation in the importance of different macroeconomic factors, notably with the S&P500 having high explanatory power during the 2000s, and (c) indicators of aggregate demand, labor market conditions and prices having a predominant role after the Great Recession. Yet, in contrast with the volatility factor, some of the variables in the price group (V90-V18) also have explanatory power in the subsamples that span the years before the Great Recession.

4.5 Is More International Data Better?

A question that arises when using large data sets is whether more data contain additional explanatory power for the extracted latent factors. In particular, the reader may argue that a large proportion of the time series used in this section come from the FRED-MD data set and thus, a-priori, we place less importance on international data. But, is more international data better? On the one hand, as mentioned earlier, our emphasis on the U.S. data is partially motivated by our use of the West Texas Intermediate oil price. Yet, as oil is traded in international markets, global supply and demand forces should influence the behavior of the WTI returns. On the other hand, as noted by McCracken and Ng(2016), the FRED-MD database has three appealing features: (1) Updates based on the FRED database are made at a monthly frequency; (2) It is publicly available; (3) Revisions to the data are made by the Fed. In contrast, databases that mimic the coverage used in the macroeconomic literature for other countries tend not to be readily available. Yet, in this section, we inquire into the robustness of our results by adding additional macroeconomic variables for a group of countries where the data is publicly available. We retain the time series for which the
sample period coincides with that of our benchmark database. As a result, the database contains more data for the U.S. and Canada than for the other two regions.

Large-scale macroeconomic databases comparable to FRED-MD have been recently built and made publicly available for the U.K. (Goulet Coulombe, Marcellino on Stevanovic, 2021) and Canada (Fortin-Gagnon et al, 2018). A smaller real time database constructed by the Euro Area Business Cycle Network (EABCN) in cooperation with several European central banks was available to the EABCN members since October 2005, but has only recently been made available for the public (Giannone, et al. 2010). To inquire whether adding more international data aids in explaining the latent factors, we augment the large data set described in the previous section with the variables from these three databases.

Interestingly, as Table 4 illustrates as we add the variables from these databases the set of variables selected for the volatility latent factor becomes more sparse. Indeed, using the same regularization parameters as in the previous section, the adaptive LASSO and adaptive Elastic-net procedures select only two variables from the benchmark database (Employees in the mining sector and total business inventories to sales ratio) and one variable from the additional data sets (Total employment in the UK).

In contrast, the set of variables selected for the mean latent factor becomes somewhat less sparse. In addition to three variables selected from FRED-MD when we use only the benchmark data set (total business inventories to sales ratio, the St. Louis adjusted monetary base, and personal consumption expenditure in durable goods), a handful of additional prices variables (V93, V96, V92, V105, V108) and the S&P500 common stock price index (V109) are selected from the U.S. data. When we extend the database to include data from Canada, the U.K. and the Euro area only M3 and several price variables for Canada are selected (see Panel B of Table 4).

All in all, while we recognize that even the extended database is limited in its coverage of international data, the estimation results reported in this section point towards two U.S. variable playing a key role in explaining the mean latent factor, total business inventories to sales ratio, St. Louis adjusted monetary base, and the former having explanatory power for the volatility latent factor. In addition, we find employment in the U.S. mining sector to have explanatory power for both the mean and the volatility latent factors.  

Before we conclude, let us turn to the question posed in this section: Is more international data better for explaining the latent factors in WTI returns? Clearly, the answer we

---

22Heat maps included in the online appendix indicate the selection of the key variables is robust to including additional international data.
can provide is limited by the sets of large-scale macroeconomic databases that are publicly available. Yet, the fact that the adaptive LASSO and adaptive Elastic-net tend to select the same key variables present in the benchmark and the extended database suggests adding more variables is not necessarily better.

5 Conclusions

We employed an endogenous regime switching model to study the behavior of the crude oil returns. To gain some intuition as to the role of allowing for endogenous switching, we started our investigation using a volatility switching model. We then build on the model by allowing for unsynchronized switching in the mean and volatility of crude oil returns. Finally, we use regularization methods to explore which variables, among a big data set used by policy analysts, explain fluctuations in the extracted mean and volatility factors that determine mean and volatility regimes.

Four key results are derived from our paper. First, we demonstrated that a model that allows the processes that govern the switching between volatility and mean regimes to evolve in an endogenous manner produces a better in-sample fit than an exogenous regime switching model. Moreover, conditional on knowing the regime, an endogenous regime switching model provides useful information regarding the time-varying nature of the volatility and, hence, could be useful in assessing risk. Second, we showed that the extracted latent volatility factor from the unsynchronized mean-volatility switching model can be used to gauge risk in crude oil markets. Third, we found that, when considering the whole 1986:1-2020:1 sample, the evolution of the mean and volatility factors is largely explained by two indicators of economic activity and monetary policy; namely, the business inventories to sales ratio and the St. Louis adjusted monetary base. Additional labor market and price variable are useful in explaining the volatility factor. Lastly, we found significant time-variation in the variables that account for the fluctuations in the latent regime factors. In particular, whereas the S&P500 composite index plays a key role during the 2000s, employment in the mining sector and other indicators of economic activity and prices gained importance since the Great Recession.

To summarize, estimation results presented in this paper suggest the use of endogenous regime switching models could be useful for policy analysts and economic agents interested in understanding fluctuations in crude oil returns. In fact, the estimated volatility regime factor can be used as a measure of risk or stress in oil markets. Moreover, tracking the
The evolution of a handful of macroeconomic variables available through the St. Louis Federal Reserve Bank, can provide useful information in understanding the evolution of these latent factors and, hence, form part of an early warning system of oil market risk.
6 References


<table>
<thead>
<tr>
<th>Parameters</th>
<th>Switching volatility</th>
<th>68% CI</th>
<th>Unsynchronized switching</th>
<th>68% CI</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\tau$</td>
<td>5.691</td>
<td>[2.893, 5.900]</td>
<td></td>
<td></td>
</tr>
<tr>
<td>$\tau_1$</td>
<td>-1.033</td>
<td>[-1.165, -0.983]</td>
<td></td>
<td></td>
</tr>
<tr>
<td>$\tau_2$</td>
<td>9.873</td>
<td>[8.339, 10.945]</td>
<td></td>
<td></td>
</tr>
<tr>
<td>$\rho_{u,v}$</td>
<td>-0.974</td>
<td>[-0.992, -0.333]</td>
<td></td>
<td></td>
</tr>
<tr>
<td>$\rho_{u,v_m}$</td>
<td>0.529</td>
<td>[0.335, 0.540]</td>
<td></td>
<td></td>
</tr>
<tr>
<td>$\rho_{u,v_v}$</td>
<td>-0.917</td>
<td>[-0.956, -0.340]</td>
<td></td>
<td></td>
</tr>
<tr>
<td>$\rho_{v_m,v_v}$</td>
<td>-0.824</td>
<td>[-0.825, -0.414]</td>
<td></td>
<td></td>
</tr>
<tr>
<td>$\rho_{v_m,v_v,u}$</td>
<td>-0.339</td>
<td>[-0.609, -0.191]</td>
<td></td>
<td></td>
</tr>
<tr>
<td>$\alpha$</td>
<td>0.971</td>
<td>[0.912, 0.978]</td>
<td></td>
<td></td>
</tr>
<tr>
<td>$a_{mm}$</td>
<td>0.133</td>
<td>[0.060, 0.242]</td>
<td></td>
<td></td>
</tr>
<tr>
<td>$a_{vm}$</td>
<td>-2.177</td>
<td>[-2.278, -1.868]</td>
<td></td>
<td></td>
</tr>
<tr>
<td>$a_{mv}$</td>
<td>-0.024</td>
<td>[-0.024, -0.024]</td>
<td></td>
<td></td>
</tr>
<tr>
<td>$a_{vv}$</td>
<td>0.847</td>
<td>[0.843, 0.848]</td>
<td></td>
<td></td>
</tr>
<tr>
<td>$\sigma_0$</td>
<td>0.069</td>
<td>[0.064, 0.072]</td>
<td>0.059</td>
<td>[0.056, 0.059]</td>
</tr>
<tr>
<td>$\sigma_1$</td>
<td>0.222</td>
<td>[0.160, 0.277]</td>
<td>0.219</td>
<td>[0.151, 0.257]</td>
</tr>
<tr>
<td>$\mu_0$</td>
<td>-0.085</td>
<td>[-0.093, -0.085]</td>
<td></td>
<td></td>
</tr>
<tr>
<td>$\mu_1$</td>
<td>0.018</td>
<td>[0.015, 0.022]</td>
<td></td>
<td></td>
</tr>
<tr>
<td>$\gamma_1$</td>
<td>0.084</td>
<td>[0.084, 0.185]</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Table 2. Selected Variables for Extracted Latent Factor in the Benchmark Database

### Panel A: Volatility Latent Factor

<table>
<thead>
<tr>
<th>ID</th>
<th>Name</th>
<th>Group</th>
</tr>
</thead>
<tbody>
<tr>
<td>V31</td>
<td>All Employees: Mining and Logging - Mining</td>
<td>U.S. Labor Market</td>
</tr>
<tr>
<td>V40</td>
<td>All Employees: Financial Activities</td>
<td>U.S. Labor Market</td>
</tr>
<tr>
<td>V45</td>
<td>Avg Hourly Earnings: Goods-Producing</td>
<td>U.S. Labor Market</td>
</tr>
<tr>
<td>V57</td>
<td>Total Business Inventories</td>
<td>U.S. Orders and Inventories</td>
</tr>
<tr>
<td>V58</td>
<td>Total Business Inventories to Sales Ratio</td>
<td>U.S. Orders and Inventories</td>
</tr>
<tr>
<td>V59</td>
<td>Consumer Sentiment Index</td>
<td>U.S. Orders and Inventories</td>
</tr>
<tr>
<td>V62</td>
<td>U.S. Real M2 Money Stock</td>
<td>U.S. Money and Credit</td>
</tr>
<tr>
<td>V63</td>
<td>St. Louis Adjusted Monetary Base</td>
<td>U.S. Money and Credit</td>
</tr>
<tr>
<td>V67</td>
<td>Real Estate Loans at All Commercial Banks</td>
<td>U.S. Money and Credit</td>
</tr>
<tr>
<td>V96</td>
<td>CPI: U.S. Apparel</td>
<td>U.S. Prices</td>
</tr>
<tr>
<td>V103</td>
<td>CPI: All Items Less Medical Care</td>
<td>U.S. Prices</td>
</tr>
</tbody>
</table>

### Panel B: Mean Latent Factor

<table>
<thead>
<tr>
<th>ID</th>
<th>Name</th>
<th>Group</th>
</tr>
</thead>
<tbody>
<tr>
<td>V58</td>
<td>Total Business Inventories to Sales Ratio</td>
<td>U.S. Orders and Inventories</td>
</tr>
<tr>
<td>V63</td>
<td>St. Louis Adjusted Monetary Base</td>
<td>U.S. Money and Credit</td>
</tr>
<tr>
<td>V106</td>
<td>Personal Consumption Expenditure: Durable Goods</td>
<td>U.S. Prices</td>
</tr>
</tbody>
</table>

Note: Variables in teal are selected for both the volatility and mean factors.
Table 3. Groups of Variables in the Benchmark Database

<table>
<thead>
<tr>
<th>ID</th>
<th>Group</th>
</tr>
</thead>
<tbody>
<tr>
<td>V1-V16</td>
<td>U.S. Output and Income</td>
</tr>
<tr>
<td>V17-V47</td>
<td>U.S. Labor Market</td>
</tr>
<tr>
<td>V48-V52</td>
<td>U.S. Consumption and Housing</td>
</tr>
<tr>
<td>V53-V59</td>
<td>U.S. Orders and Inventories</td>
</tr>
<tr>
<td>V60-V71</td>
<td>U.S. Money and Credit</td>
</tr>
<tr>
<td>V72-V89</td>
<td>U.S. Interest Rates and Exchange Rates</td>
</tr>
<tr>
<td>V90-V108</td>
<td>U.S. Prices</td>
</tr>
<tr>
<td>V109-V113</td>
<td>U.S. Stock Market</td>
</tr>
<tr>
<td>V114-V125</td>
<td>U.S. Economic Policy Uncertainty</td>
</tr>
<tr>
<td>V126-130</td>
<td>Global Oil Market and World Economic Activity</td>
</tr>
</tbody>
</table>
Table 4. Selected Variables for Extracted Latent Factors in the Extended Database

<table>
<thead>
<tr>
<th>Panel A: Volatility Latent Factor</th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>ID</td>
<td>Name</td>
<td>Group</td>
</tr>
<tr>
<td>V31</td>
<td>All Employees: Mining and Logging-Mining</td>
<td>U.S. Labor Market</td>
</tr>
<tr>
<td>V58</td>
<td>Total Business Inventories to Sales Ratio</td>
<td>U.S. Orders and Inventories</td>
</tr>
<tr>
<td>V518</td>
<td>Number of people in employment (16 and over)</td>
<td>U.K. Labor Market</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Panel B: Mean Latent Factor</th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>ID</td>
<td>Name</td>
<td>Group</td>
</tr>
<tr>
<td>V31</td>
<td>All Employees: Mining and Logging - U.S. Mining</td>
<td>U.S. Labor Market</td>
</tr>
<tr>
<td>V58</td>
<td>Total Business Inventories to Sales Ratio</td>
<td>U.S. Orders and Inventories</td>
</tr>
<tr>
<td>V63</td>
<td>St. Louis Adjusted Monetary Base</td>
<td>U.S. Money and Credit</td>
</tr>
<tr>
<td>V93</td>
<td>Production Price Index: Crude Materials</td>
<td>U.S. Prices</td>
</tr>
<tr>
<td>V96</td>
<td>CPI: Apparel</td>
<td>U.S. Prices</td>
</tr>
<tr>
<td>V102</td>
<td>CPI: All Items Less Food</td>
<td>U.S. Prices</td>
</tr>
<tr>
<td>V105</td>
<td>Personal Consumption Expenditure: Chained Index</td>
<td>U.S. Prices</td>
</tr>
<tr>
<td>V106</td>
<td>Personal Consumption Expenditure: Durable Goods</td>
<td>U.S. Prices</td>
</tr>
<tr>
<td>V108</td>
<td>Personal Consumption Expenditure: Services</td>
<td>U.S. Prices</td>
</tr>
<tr>
<td>V185</td>
<td>New Housing Price Index</td>
<td>Canada Prices</td>
</tr>
<tr>
<td>V239</td>
<td>M3</td>
<td>Canada Money and Credit</td>
</tr>
<tr>
<td>V292</td>
<td>IP Price Index: Metal and Construction Materials</td>
<td>Canada Prices</td>
</tr>
<tr>
<td>V294</td>
<td>IP Price Index: Machinery and Equipment</td>
<td>Canada Prices</td>
</tr>
<tr>
<td>V436</td>
<td>CPI: Services</td>
<td>Canada Prices</td>
</tr>
<tr>
<td>V437</td>
<td>CPI: All</td>
<td>Canada Prices</td>
</tr>
<tr>
<td>V438</td>
<td>CPI: Shelter</td>
<td>Canada Prices</td>
</tr>
<tr>
<td>V441</td>
<td>CPI: All minus Food</td>
<td>Canada Prices</td>
</tr>
<tr>
<td>V447</td>
<td>CPI: Clothing and Footwear</td>
<td>Canada Prices</td>
</tr>
<tr>
<td>V458</td>
<td>CPI: Health and Personal Care</td>
<td>Canada Prices</td>
</tr>
<tr>
<td>V460</td>
<td>CPI: All minus Food and Energy</td>
<td>Canada Prices</td>
</tr>
<tr>
<td>V465</td>
<td>CPI: Shelter (Quebec)</td>
<td>Canada Prices</td>
</tr>
<tr>
<td>V473</td>
<td>CPI: Clothing and Footwear (Ontario)</td>
<td>Canada Prices</td>
</tr>
<tr>
<td>V496</td>
<td>CPI: Clothing and Footwear (Saskatchewan)</td>
<td>Canada Prices</td>
</tr>
<tr>
<td>V507</td>
<td>CPI: Durable Goods (Alberta)</td>
<td>Canada Prices</td>
</tr>
</tbody>
</table>

Note: Variables in blue are selected in the benchmark and extended databases.
Figure 1: Evolution of WTI Prices, Returns and Volatility

(a) Monthly WTI prices and returns

(b) Monthly volatility

Notes: Panel (a) plots monthly West Texas Intermediate (WTI) prices (blue line) and returns (black line). Panel (b) plots the monthly volatility of the WTI returns.
Figure 2: Extracted Latent Factor: Volatility Switching Model

Notes: This figure depicts the extracted latent factor from the volatility switching model (solid black line) and the estimated threshold $\tau$ (dashed red line).
Figure 3: Transition Probabilities: Volatility Switching Model

(a) High to High  
(b) Low to High

Notes: This figure presents the transition probabilities from the volatility switching model. Panel (a) depicts the transition probability from high to high volatility state: the black solid line corresponds to the time-varying transition probabilities from the endogenous volatility switching model, while the red dashed line represents the constant transition probability from the exogenous model. Similarly, panel (b) depicts the transition probabilities of switching from the low to the high volatility regime.

Figure 4: Extracted Factors - Unsynchronized Mean and Volatility Switching

(a) Mean Factor  
(b) Volatility Factor

Notes: This figure depicts the extracted latent factors from the unsynchronized mean-volatility switching model. The solid black line in panel (a) corresponds to the mean latent factor and the red dashed line represents the estimated threshold $\tau_m$. Similarly, the solid black line in panel (b) represents to the volatility latent factor and the red dashed line depicts the corresponding threshold $\tau_v$. 
Figure 5: Correlation and Coherence between Extracted Latent Factors

(a) Correlation
(b) Coherence

Notes: Panel (a) depicts the correlation between the mean and volatility latent factors computed using a 24-month rolling window. Panel (b) depicts the coherence between the two latent factors computed using the full sample.
Figure 6: LH-to-LH Transition Probabilities and WTI Returns

(a) Persian Gulf War

(b) Great Recession

(c) Oil Price Collapse

(d) Covid-19 Pandemic

Notes: This figure illustrates the probability to remain in low return-high volatility state for four episodes of turbulence in oil markets. The black solid line represents the transition probability $P(s_{m,t} = 0, s_{v,t} = 1 | s_{m,t-1} = 0, s_{v,t-1} = 1, y_{t-1})$ estimated from the endogenous unsynchronized switching model; the dashed red line corresponds to the constant transition probability $P(s_{m,t} = 0, s_{v,t} = 1 | s_{m,t-1} = 0, s_{v,t-1} = 1)$ estimated from the exogenous switching model; the solid blue line corresponds to the WTI returns. The shaded areas denote periods of low returns and high volatility.
Figure 7: Extracted Volatility Regime Factor from Mean-Volatility Switching Model and CBOE Crude Oil Volatility Index (OVX)

Notes: This figure plots the volatility latent factor extracted from the unsynchronized mean-volatility switching model (solid blue line) and the Chicago Board Options Exchange (CBOE) Crude Oil Volatility Index OVX (dashed red line).
Figure 8: Adaptive LASSO Heat Map: Importance of Different Variables in Explaining the Volatility and Mean Latent Factors

(a) Volatility

(b) Mean

Notes: This figure reports the normalized coefficients obtained with the adaptive LASSO. Panel (a) reports the normalized coefficients for the volatility factor. Panel (b) reflects the normalized coefficients for the mean factor.
Online Appendix for Oil Price Volatility, Endogenous Regime Switching and Macroeconomic Factors *

Yoosoon Chang †  Ana María Herrera ‡  Elena Pesavento §
Indiana University  University of Kentucky  Emory University

November 8, 2021

*We are grateful to participants in the 2019 Workshop on Energy Economics and the 2019 Midwest Econometrics Group meetings for very useful comments. Our special thanks go to Sangmyung Ha for his superb research assistance and to Joon Y. Park and Shi Qiu for many helpful discussions. We also thank Diego Rojas Baez for his excellent research assistance.
†Indiana University, Email: yoosoon@indiana.edu.
‡University of Kentucky, Email: amherrera@uky.edu.
§Emory University, Email: epesave@emory.edu.
A Estimation with Brent Prices

This section contains the results from the estimation of the endogenous regime switching model with one and two (asynchronous) Latent factors for returns computed with Brent Oil prices. The data are monthly data from May 1987 to January 2021. Volatility is computed as the sum of daily squared returns. The estimated models are described in Section 2 of the main paper. 68% confidence intervals are computed using Stationary Block Bootstrap.

Table A.1: Descriptive Statistics

<table>
<thead>
<tr>
<th>Monthly (demeaned) Brent return</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>mean</td>
<td>0.003</td>
</tr>
<tr>
<td>std</td>
<td>0.100</td>
</tr>
<tr>
<td>max</td>
<td>0.469</td>
</tr>
<tr>
<td>75 quantile</td>
<td>0.062</td>
</tr>
<tr>
<td>median</td>
<td>0.009</td>
</tr>
<tr>
<td>25 quantile</td>
<td>-0.050</td>
</tr>
<tr>
<td>min</td>
<td>-0.555</td>
</tr>
</tbody>
</table>

Table A.2: One Factor Volatility Switching Model

<table>
<thead>
<tr>
<th>Parameters</th>
<th>Estimates</th>
<th>68% CI</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\tau$</td>
<td>3.617</td>
<td>[1.896, 4.739]</td>
</tr>
<tr>
<td>$\rho$</td>
<td>-0.794</td>
<td>[-1.000, -0.423]</td>
</tr>
<tr>
<td>$\alpha$</td>
<td>0.902</td>
<td>[0.760, 0.945]</td>
</tr>
<tr>
<td>$\sigma_0$</td>
<td>0.075</td>
<td>[0.068, 0.080]</td>
</tr>
<tr>
<td>$\sigma_1$</td>
<td>0.281</td>
<td>[0.173, 0.350]</td>
</tr>
</tbody>
</table>

log-likelihood | 415.484 | [400.181, 469.012]|
Table A.3: Two factors model: Unsynchronized Switching Model

<table>
<thead>
<tr>
<th>Parameters</th>
<th>Estimates</th>
<th>68% CI</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\tau_1$</td>
<td>-1.090</td>
<td>[-1.346, -1.090]</td>
</tr>
<tr>
<td>$\tau_2$</td>
<td>7.727</td>
<td>[6.324, 8.067]</td>
</tr>
<tr>
<td>$\rho_{u,v_1}$</td>
<td>0.426</td>
<td>[0.245, 0.426]</td>
</tr>
<tr>
<td>$\rho_{u,v_2}$</td>
<td>-0.791</td>
<td>[-0.927, -0.723]</td>
</tr>
<tr>
<td>$\rho_{v_1,v_2}$</td>
<td>-0.891</td>
<td>[-0.891, -0.542]</td>
</tr>
<tr>
<td>$\alpha_{11}$</td>
<td>0.307</td>
<td>[0.307, 0.546]</td>
</tr>
<tr>
<td>$\alpha_{21}$</td>
<td>-1.544</td>
<td>[-1.632, -1.412]</td>
</tr>
<tr>
<td>$\alpha_{12}$</td>
<td>0.008</td>
<td>[0.008, 0.014]</td>
</tr>
<tr>
<td>$\alpha_{22}$</td>
<td>0.840</td>
<td>[0.777, 0.840]</td>
</tr>
<tr>
<td>$\mu_0$</td>
<td>-0.099</td>
<td>[-0.103, -0.099]</td>
</tr>
<tr>
<td>$\mu_1$</td>
<td>0.020</td>
<td>[0.016, 0.022]</td>
</tr>
<tr>
<td>$\gamma_1$</td>
<td>0.069</td>
<td>[0.069, 0.194]</td>
</tr>
<tr>
<td>$\sigma_0$</td>
<td>0.064</td>
<td>[0.060, 0.064]</td>
</tr>
<tr>
<td>$\sigma_1$</td>
<td>0.255</td>
<td>[0.174, 0.267]</td>
</tr>
</tbody>
</table>

log-likelihood 430.205 [392.652, 458.160]

(a) Returns

(b) Volatility

Figure A.1: Evolution of Monthly Brent Prices Returns and Volatility
Figure A.2: Extracted Latent Factor: Volatility Switching Model

Figure A.3: Transition Probabilities for One Factor Model - Brent

(a) High to High

(b) Low to High
Figure A.4: Extracted Factors - Unsynchronized Mean and Volatility for Brent

Figure A.5: Correlation and Coherence between Extracted Latent Factors
Figure A.6: High-to-High Transition Probabilities and Brent Returns

(a) Persian Gulf War
(b) Great Recession
(c) Oil Price Collapse
(d) Covid-19 Pandemic
Figure A.7: Adaptive LASSO Heat Map: Importance of Different Variables in Explaining the Volatility and Mean Latent Factors - Additional International Data

(a) Volatility

(b) Mean

Notes: This figure reports the normalized coefficients obtained with the adaptive LASSO. Panel (a) reports the normalized coefficients for the volatility factor. Panel (b) reflects the normalized coefficients for the mean factor.