

Demand Effects and Speculation in Oil Markets: Theory and Evidence

Eyal Dvir and Kenneth Rogoff
Boston College and Harvard University

Abstract

We present evidence showing the existence of stable cointegrating vectors connecting four important variables in the U.S. oil market: U.S. oil production, U.S. stocks of crude oil, the real price of oil, and U.S. industrial production. Our data are monthly, and go back to the 1930s, split into sub-samples which correspond to periods before and after the 1973 crisis. We further show that the cointegrating vectors found in the data accord well with an extended commodity storage model which allows for demand growth dynamics and for supply regimes.

1 Introduction

The role of speculation in driving the price of crude oil has been the object of renewed interest recently. The decades-old debate, between those who argue that market developments can be directly attributed to changes in fundamentals and those who believe that speculators are creating price volatility, is showing no signs of abating¹. In this paper we put forward the argument that a simple model with four variables - inventories, production, income, and price - can be useful in capturing important long-run features of the market for crude oil, and in particular elucidating the seemingly unstable relationship between inventories and price. We then estimate the long-run relationships among these four variables for the period 1933 - 2011, taking

¹See Singleton (2012) for a recent review. Recent theoretical contributions include Basak and Pavlova (2012), where institutional investors can cause commodity prices to rise and become more volatile, and Sockin and Xiong (2012), where feedback effects from futures prices affect commodity spot prices through an information channel. However, in a survey recent empirical work Fattouh, Kilian, and Mahadeva (2012) find scant evidence for the effect of speculation on prices.

account of the break in the series which occurred around the first oil crisis of 1973. We show that our model's long-run predictions are borne out by the data, in that long-run relationships between these variables exist, that they are stable before and after the 1973 crisis, and that they comport in sign to the model's predictions. In particular, we find that post - 1973 there is a stable **positive** long-run relationship between U.S. commercial inventories of crude oil and the U.S. price of crude, whereas before 1973 that relationship was **negative**.

In a previous paper, Dvir and Rogoff (2009), we argue that the real price of oil has gone through three distinct periods. First, from 1861 to about 1878 (a period not covered in the current paper), the price of oil was generally high (in real terms), and was moreover highly persistent and volatile. Then came a much less volatile period, between 1878 and 1973, in which prices were also generally lower and not at all persistent. This long period can be further divided into two sub-periods: before and after 1933, where price volatility is significantly lower after 1933 compared with the years 1878-1933. Finally, from 1973 onwards, there is a recurrence of high persistence and volatility accompanied again by higher prices. In that paper, we offered a narrative, based on our reading of the historical events, for the recurrence of high price persistence in the two end-periods mentioned, 1861-1878 and 1973-2010. We argued that in these periods two forces coincided: first, demand (governed by income) was high and very persistent, i.e it was governed by growth shocks. Second, access to supply was restricted by agents who had the capability and incentive to do so.

In that paper we also presented our model, which is an extension of the canonical commodity storage model à la Deaton and Laroque (1992, 1996). Our model introduces income growth dynamics to that framework, and in particular, can accommodate both $I(0)$ and $I(1)$ income processes. In that paper we focused on predictions of the model with regards to demand and supply shocks. In the current paper we emphasize the related but distinct **long-run** relationships which are also predicted by the model. This focus is important when attempting to account for actual market behavior. Perhaps due to the existence of unit roots in the various series - inventories, production, and price - identifying long-run relationships among them has not been a priority in the literature on the market for crude oil. However predictable long-run relationships between these variables are a hallmark of the commodity storage model. Therefore our contribution here is two-fold: first, we show that these long-run relationships do exist in the data, as predicted by the model. Second, we show that the relationships we identify in the data are sign - consistent with our version of the commodity storage model.

Introducing income growth dynamics to the classical model adds considerably to

its predictive capacity, partly by changing some of the classical model's predictions. The model can now predict how inventories and price behave when income rises or falls, conditioning on production behavior. The relationship between inventories and price is therefore no longer simple, becoming a function of production behavior. In particular, inventories may rise or fall with income. Specifically, in periods when production is flexible, i.e. when a rise in income, which raises demand, is predicted to result in a commensurate rise in production, inventories should fall, since the effects of high demand should dissipate quickly. This should help mitigate any rise in price associated with the surge in demand so that inventories and price should exhibit a negative relationship. Conversely, in periods when production is inflexible, i.e. when a rise in income is not predicted to raise production significantly, inventories should rise. This would actually enhance any rise in prices associated with high demand, so that inventories and price should exhibit a positive long run relationship. Note importantly that causality runs in both directions in a commodity storage model: in the flexible supply case a rise in prices will cause inventories to drop, thereby releasing more oil to the market and exerting a downward effect on prices. In the restricted supply case a rise in price leads to a surge in inventories, pulling oil off the market and leading to further price rises. The relationships we identify should therefore be understood as long-run equilibrium conditions, and not given a causality interpretation. It is therefore important to specify the long-run relationship among all four variables - inventories, production, income, and price - at the same time.

We use U.S. monthly data on crude oil stocks, production, and prices, as well as on the size of the economy, going back to 1/1931. Demand is assumed to rise with income, or the size of the economy. Our chosen variable is monthly U.S. industrial production, which is available from the Federal Reserve. Petroleum is used, directly or indirectly, by every sector of the economy; therefore demand for oil is commensurate with overall economic activity. The measure we use is the most complete measure of economic activity in the U.S. at a monthly frequency which is available for the entire period. Our measures of U.S. crude oil stocks and production are from the Energy Information Administration. Oil price in our dataset is a composite series of monthly prices quoted in Texas and Oklahoma.

A common feature of these series is that they exhibit unit roots, at least for some sub-period. The most pressing empirical question then becomes: Are these series cointegrated? For our commodity storage model to be a reasonable account of the market for crude oil, these series must, in the long run, co-move in a predictable way. If there is little evidence of the existence of a cointegrating vector, then the commodity storage model cannot be the simplest way of describing the market. Our first task then is to establish that stationary cointegrating vectors do indeed exist.

Since our model leads us to expect that the market should behave differently before and after the crisis of 1973, we believe it is necessary to split the sample around that time. In Section 3 we show that using standard Johansen tests we can establish quite clearly that, yes, these series are cointegrated.

Our second task is to estimate these cointegrating vectors to see whether the variables interact in the way predicted by the model. We run vector error correction regressions and arrive at statistically and economically significant long-run relationships among the four variables, in two separate sub-samples: 1933/1 - 1973/12, and 1975/1 - 2011/12. In the earlier sub-sample, supply was quite flexible, as described in detail in our previous paper. We would therefore expect inventories to increase with production, and to decrease with demand and with price. That is indeed what we find. In the later sub-sample, supply was inflexible, since U.S. production was at its limit (wells were operating at capacity) and U.S. firms had limited access to additional oil. We would expect then to see inventories **increase** with demand and with price. Here as well our estimates of the coefficients of the cointegrating vector accord well with the model's expectations. Experimenting with different break points leads to different magnitudes for the coefficients, but importantly their sign remains stable, and they remain statistically significant.

The paper proceeds as follows: Section 2 presents our model. Section 3 presents our empirical findings. Section 4 concludes.

2 An Extended Commodity Storage Model

Our model is an extension of the classic commodity storage framework. Chambers and Bailey (1996) and Deaton and Laroque (1996) extend the model to allow for autoregressive shocks. We extend it further to explicitly incorporate demand, and to allow for growth shocks².

2.1 Availability and Storage

Time is discrete, indexed by t . The market for oil consists of consumers, producers, and risk neutral arbitrageurs. The latter have at their disposal a costly storage technology which may be used to transfer any positive amount of oil from period $t - 1$ to period t . Storage technology is limited by a non-negativity constraint, i.e. the amount stored at any period cannot drop below zero. This implies that

²This is essentially the same model we presented in Dvir and Rogoff (2009). We include it here for completeness and also because here we emphasize its predictions of stable relationships between the constituent series. Empirical work on commodity futures curves has shown the need for a permanent shock component: see Routledge, Seppi, and Spatt (2000), and Schwartz and Smith (2000).

intertemporal arbitrage, although potentially profitable, cannot always be achieved. In these cases the market is "stocked out". Let A_t denote *oil availability*, the amount of oil that can potentially be consumed at time t . This amount has already been extracted from the ground, either in period t or at some point in the past, and has not been consumed before period t . It is given by

$$A_t = X_{t-1} + Z_t, \tag{1}$$

where X_{t-1} denotes the stock of oil transferred from period $t-1$ to t , and Z_t denotes the amount of oil that is produced at time t . For simplicity, we assume that no oil is lost due to storage³. Decisions concerning both variables - how much to store, how much to produce - are assumed to have been made before period t began. In period t agents decide how to divide A_t between current consumption Q_t and future consumption, so that demand - the sum of current consumption and the amount stored for the future - must always equal current availability:

$$A_t = Q_t + X_t. \tag{2}$$

2.2 Demand for Oil

Let P_t denote the price of crude oil, and let Y_t be a demand parameter, which should be thought of as capturing the economy's derived demand for energy stemming from industrial, residential, and transportation uses. For simplicity, we will refer to Y_t as income. We can then write an inverse demand function for oil as follows:

$$P_t = P(Q_t, Y_t), \tag{3}$$

where inverse demand is decreasing in its first argument, and increasing in its second. This constitutes a mild departure from the canonical model, where demand for the commodity is a function of its price alone. This departure is a natural one to make, however, in the context of oil, as oil consumption and income are very highly correlated. We posit an inverse demand function in which only the *ratio* of consumption to income matters, i.e. inverse demand is homogeneous of degree zero:

$$P_t = P(Q_t, Y_t) = P\left(\frac{Q_t}{Y_t}, 1\right) = p(q_t), \tag{4}$$

where lowercase letters denote variables normalized by Y_t . We will refer to normalized variables as "effective" amounts, in the sense that a growing economy leads to higher energy needs, spreading any given amount of oil more thinly.

³Alternatively, we could have specified storage costs by a given loss percentage, as in Deaton and Laroque (1996).

We will use a CES inverse demand function:

$$P_t = q_t^{-\gamma} = (a_t - x_t)^{-\gamma}, \quad (5)$$

where $\gamma > 1$ is the inverse elasticity of demand, and a_t, x_t denote effective availability and storage in period t , respectively. It is natural to assume that the effective demand for oil is inelastic with respect to price. As equation (5) makes clear, for a given supply of oil, price is a function of the competing demands of current and future consumption. If the desire to consume more in the future grows (driven by expectations of future conditions), more oil is stored rather than consumed today, resulting in a price rise today even though supply has not changed.

Let \bar{Y}_t denote trend income, i.e. the level of income that would prevail at time t in a world without income shocks. \bar{Y}_t , which we think of as a measure of current production technology, is assumed to increase over time at a constant rate $\bar{\mu} > 0$. We now consider two alternative stochastic processes for Y_t : one where income moves around a deterministic trend, and another where the trend itself is stochastic. The former is a simple AR(1) process, analogous to the stochastic process that Deaton and Laroque (1996) consider for supply. Under this assumption we have:

$$\frac{Y_{t+1}}{\bar{Y}_{t+1}} = \left(\frac{Y_t}{\bar{Y}_t} \right)^\rho e^{\varepsilon_{t+1}}, \quad (6)$$

where $\rho \in (0, 1)$ and $\varepsilon_{t+1} \sim N(0, \sigma_\varepsilon^2)$ is an iid shock. We think of this case as more closely relevant to income shocks in developed economies, where the economy exhibits business cycles around a stable trend. In the latter case, we assume instead:

$$Y_{t+1} = e^{\mu_{t+1}} Y_t, \quad (7)$$

such that

$$\mu_{t+1} = (1 - \phi)\bar{\mu} + \phi\mu_t + v_{t+1}, \quad (8)$$

where $\phi \in (0, 1)$ and $v_{t+1} \sim N(0, \sigma_v^2)$ is an iid shock. Dividing both sides of (7) by \bar{Y}_{t+1} we get:

$$\frac{Y_{t+1}}{\bar{Y}_{t+1}} = e^{\mu_{t+1} - \bar{\mu}} \frac{Y_t}{\bar{Y}_t}. \quad (9)$$

We think of this case as more relevant to income shocks in some developing countries, in particular quickly industrializing economies where very high growth rates can be quite persistent.

2.3 Supply of Oil

In the canonical commodity storage model, supply Z_t varies according to some stochastic process ψ_t around a predetermined mean \tilde{Z}_t , and it is this variability in supply that creates an incentive for inter-temporal smoothing by the large pool of risk neutral arbitrageurs. As the literature has long recognized, demand and supply shocks in the canonical model are isomorphic: one can think of a negative realization of ψ_t as representing an especially cold winter (demand) or a breakdown in a major pipe (supply). For this reason, since we model demand shocks explicitly, it would be redundant to model supply shocks separately.

We do model supply *choices*, however. In particular, we assume that either of the following two regimes holds: a regime where oil supply does not react at all to demand shocks due to capacity constraints (such as railroad infrastructure or number of operational wells), and a regime in which oil supply fully accommodates any shock to demand (for example, when potential production is much higher than current production). We think of the former regime as describing supply behavior when access to excess supply sources is restricted, so that suppliers are constrained to produce at their installed capacity⁴. Under the latter regime, suppliers seek to stabilize prices by varying quantities as needed. We think of this regime as representing either perfectly competitive supply, where producers will offer any amount at a given price, or else the effect of purposeful government intervention, seeking to control market prices by adjusting supply.

Formally, in the former regime we assume that supply grows at the trend income rate $\bar{\mu}$, so that

$$Z_{t+1} = \tilde{Z}\bar{Y}_t, \quad (10)$$

where \tilde{Z} is a supply parameter. Next period's oil supply depends then on current technology, since overall technological progress, which drives global GDP growth, applies to the oil extraction and exploration sectors as well, and therefore determines overall capacity.

This assumption deserves some comment. The total amount of oil existing in the earth's crust is finite. However technological progress is key to exploiting an increasing fraction of it over time. The global ratio of oil production to known oil reserves is slightly less than 2.5% , and has been quite steady at that level since 1985 (BP Statistical Review), even though global production has increased by about 39%

⁴Naturally, capacity constraints can be relaxed in the medium run. However, as long as capacity does not fully accommodate all demand shocks, dynamic behavior will be qualitatively similar to the case where it does not react at all. A similar point has been made by Williams and Wright (1991).

from 1985 to 2010. The world economy is no closer to running out of oil now than it was in 1985 due to the rate at which new reserves are discovered and known reserves become exploitable due to better technology. This is the context which drives our modeling choice, since it suggests that a stationary equilibrium relationship among the important variables might exist.

Note that in this regime oil supply depends on the technology driving income growth, but not on income growth itself. Therefore shocks to demand will drive a wedge between supply and demand, causing a shift in equilibrium price. In contrast, under the alternative supply regime oil suppliers will accommodate all income shocks, i.e. oil supply will be perfectly elastic. Next period's supply then will also depend on current income level (and growth rate if appropriate). Supply is then given by:

$$Z_{t+1} = \tilde{Z}\bar{Y}_t \left(\frac{Y_t}{\bar{Y}_t} \right)^\rho, \quad (11)$$

for the AR(1) case or by:

$$Z_{t+1} = \tilde{Z}e^{(1-\phi)\bar{\mu}+\phi\mu_t}Y_t, \quad (12)$$

for the stochastic trend case.

2.4 Storage of Oil

The defining characteristic of the canonical model is the availability of storage technology, i.e. the ability to perform intertemporal arbitrage. Here we follow the literature closely. We assume free entry into the storage sector as well as risk neutrality, implying that the actions of arbitrageurs will raise or lower the current price until it is at a level which renders the strategy unprofitable in expectation, unless that would require holding negative stocks, at which case inter-temporal arbitrage will be incomplete. In all other cases, i.e. when equilibrium at time t is fully optimal, the price of oil must obey the following arbitrage condition:

$$P_t = \beta E_t[P_{t+1}] - C, \quad (13)$$

where $\beta = 1/(1+r)$ is the discount factor, and $r > 0$ is the exogenously given interest rate. The parameter $C > 0$ denotes the per barrel cost of storage. Equilibrium price P_t must be such that there is no incentive to increase or decrease X_t , the amount stored⁵.

Note that storage involves an intertemporal choice, whereas the production decision does not. This is worth mentioning since models of the oil market which

⁵The inter-temporal price condition (13) does not hold in the case of a stockout, i.e. the case where $X_t = 0$ because the storage non-negativity constraint is binding; every barrel of extracted oil

emphasize non-renewability imply that producers must decide whether to extract a barrel of oil today or tomorrow. That is not the case here: in our model, as in the canonical storage model, production decisions are made based on current and expected market conditions. Hence the real interest rate enters the storage equation, but does not enter the production equations.

2.5 The Rational Expectations Equilibrium

The canonical commodity storage model is a rational expectations model with one state variable - availability of oil A_t - and one choice variables - storage of oil X_t . A solution of the model - the rational expectations equilibrium - consists of a storage rule, which specifies the level of storage for every possible value of the state variable. Determination of price and consumption follows immediately from this rule. In our extended version of the model the rule retains its salient characteristics, well known from the literature (see below). However in the extended version, as in the AR(1) case considered by Chambers and Bailey (1996), storage is also the function of one (or two) exogenous variables, depending on assumptions regarding the income process. Relative income Y_t/\bar{Y}_t - how far above or below its mean is the current level of income - serves as the second state variable of the model when we assume that income follows a stable trend. For the case where income is subject to growth shocks, we need a third state variable: the current growth rate of income, denoted by μ_t .

In order to solve the model we express all quantity variables in their normalized forms. The model can be then be summarized by two (or three) transition functions which govern the state variables, and one response equation which determines storage, the decision variable. We therefore arrive at a 2×2 framework: two alternatives for the demand process and two for the supply regime. Agents in the model observe all the state variables every period, and decide on storage accordingly, taking into consideration expectations regarding the next period's price, and implicitly producers' behavior.

The transition functions for the stable trend case are:

$$a_{t+1} = \frac{x_t + z_{t+1}}{(Y_t/\bar{Y}_t)^{\rho-1} e^{\bar{\mu} + \varepsilon_{t+1}}}, \quad (15)$$

$$\frac{Y_{t+1}}{\bar{Y}_{t+1}} = \left(\frac{Y_t}{\bar{Y}_t} \right)^\rho e^{\varepsilon_{t+1}}, \quad (16)$$

is being used for consumption. As a result, current price is above its unconstrained level:

$$P_t > \beta E_t[P_{t+1}] - C. \quad (14)$$

where equation (15) is derived by normalizing equation (1) by Y_{t+1} and using (6). Effective supply z_{t+1} is arrived at by dividing either equation (10) or (11) through by Y_t , depending on the supply regime in effect.

For the stochastic trend case, there are three transition functions:

$$a_{t+1} = (x_t + z_{t+1})/e^{\mu_{t+1}}, \quad (17)$$

$$\frac{Y_{t+1}}{\bar{Y}_{t+1}} = e^{\mu_{t+1} - \bar{\mu}} \frac{Y_t}{\bar{Y}_t}, \quad (18)$$

$$\mu_{t+1} = (1 - \varphi)\bar{\mu} + \varphi\mu_t + v_t, \quad (19)$$

where the transition function (17) is derived again by normalizing equation (1) by Y_{t+1} , now using (7) instead. Here as well, the supply regime in effect determines how we arrive at z_{t+1} : dividing either equation (10) or (12), as appropriate, by Y_t .

The response equation for both cases is:

$$(a_t - x_t)^{-\gamma} = \beta E_t[P_{t+1}] - C. \quad (20)$$

Note importantly that equation (20), which determines optimal storage, holds only when the state variables are such that the optimal storage is non-negative. If the state variables dictate negative storage, this response condition breaks down and we have simply $P_t = a_t^{-\gamma}$.

Commodity storage models generally cannot be solved analytically even in their most simple form (Newbury and Stiglitz, 1981, Williams and Wright, 1991). We therefore follow the literature since Gustafson's (1958) original contribution and proceed to solve the model numerically⁶. It turns out from our numerical solutions that the storage rules which result from any of our four sets of assumptions regarding supply and demand are very similar. All four of these rules are essentially identical in form to the rule that results from the canonical model. The difference is that in our extended model these rules hold for the *normalized* variables instead of the original quantities. In other words, effective storage has a relationship with effective availability in the extended model, under both sets of assumptions regarding demand, and both supply regimes, that is qualitatively similar to the relationship between actual storage and actual availability in the canonical model. As far as we know this is a new result as well.

Figure 1 shows a typical storage rule as well as the corresponding equilibrium price, both as functions of effective oil availability a_t (on the horizontal axis)⁷. Both curves are qualitatively similar regardless of our assumption on income's stochastic

⁶See Dvir and Rogoff (2009), appendix B for details of the solution method.

⁷Certain assumptions need to be made regarding the model's parameters in order to solve the

process or the supply regime. Together these curves signify the location of equilibrium at every possible level of effective availability. As in the canonical model, storage is a positive function of availability beyond a certain point (below this point the non-negativity constraint is binding), whereas price is a negative function of availability, the curve becoming less steep once storage is positive.

Figure 2 exhibits the novel results of our model. In its two panels we show the effect of a rise in relative income Y_t/\bar{Y}_t (horizontal axis) on effective storage x_t . In the upper panel we show the rational expectations equilibrium where supply is flexible and demand grows around a deterministic trend. In the lower panel we show the RE equilibrium where supply is restricted and demand exhibits a stochastic trend. Our model predicts that in the former case (flexible supply, deterministic trend), a rise in relative income will be accompanied by a reduction in inventories. The reason is as follows: as income rises above its long-run trend, production will increase to accommodate the higher demand, and also income will be expected to revert back to its trend. Both forces imply that any rise in price will be short-lived, and therefore rational agents will sell some of their inventories in order to profit from the relatively higher price. On the other hand, when supply is restricted and demand exhibits a unit root (lower panel), a rise in income is not predicted to induce a rise in production or any mean reversion. For this reason rising prices due to rising demand can be seen as a process which is likely to continue, and rational agents will accumulate inventories as a result. Note that in both panels we also show that higher availability (i.e. higher production for any given relative income) will in both cases be associated with higher inventories, as already seen in Figure 1.

3 Stocks, Production, Demand, and Price: Empirical Links Over Time

We have monthly production and stocks data from the U.S. Dept. of Energy going back to 1920/1, covering the entire U.S. Our oil price series, reflecting prices in Oklahoma (what became in the 1980's the West Texas Intermediate price), is constructed from Commodity Research Bureau (1940, 1950, 1960), for 1931/1 - 1958/12, and from the IMF's International Financial Statistics database for 1959/1 - 2011/12. We deflate this series by the U.S. CPI (Bureau of Labor Statistics), to arrive at the real price of oil. For industrial production, we use the Federal Reserve Board's Indus-

model numerically. Demand elasticity $-1/\gamma$ is set at -0.5. The cost of storage C is 0.02 per barrel. The discount factor β is set at 0.97. The trend income growth rate $\bar{\mu}$ is set at 0.02, the income persistence parameter ρ is set at 0.6, and the growth persistence parameter ϕ is set at 0.45. Effective supply capacity \tilde{Z} is set at $e^{\bar{\mu}}$. Lastly, the income shock's standard deviation σ is set at 0.1, and the growth shock's standard deviation v is set at 0.02.

trial Production series, which starts in 1919/1. We utilize the most inclusive index available. Since our price series starts at 1931/1, the data we use in our regressions covers the period 1931/1 - 2011/12. Figures 3 and 4 present the four series for the sub-periods 1920/1 - 1972/12 and 1973/1 - 2011/12.

Preliminary tests cannot reject the null that all of the series contain a unit root⁸. Our model, while able to accommodate stochastic trends, nevertheless posits a stationary relationship between the variables of interest. Note however that the model posits a different relationship in the sub-period when supply was flexible - before U.S. access to global supplies was severely restricted in 1973 - than the relationship which should exist in the sub-period when supply is restricted - anytime after 1973⁹. Our first task is therefore to test whether a stationary cointegrating vector exists. We split the sample in the following way: the former sub - period, when supply was flexible, includes observations up to and including 1972/12. This is an arbitrary choice, but one which turns out not to make much difference to the results. For the latter period, when supply was restricted, we will include all of 1973 to conform with the extensive literature which has looked at the oil market starting in 1973/1. To allow for 24 lags within the same sub - period, we will formally examine the series' behavior from 1975/1 - 2011/12. For both sub-periods our variables are: log of oil production, log of oil inventories, log of industrial production index, and log of the real price of oil.

Table 1 presents the results of standard Johansen tests conducted on the two sub-samples, with all variables included, as well as a constant and seasonal dummies. The number of lags included is determined by the HQ information criterion, since it is a consistent statistic of the true number of lags¹⁰. We see that for both sub-samples the null hypothesis of no cointegrating vector (rank zero) is strongly rejected. In 1933/5 - 1973/12, there is no evidence of more than one cointegrating vector. However in 1975/1 - 2011/12 there is some evidence of more than one cointegrating vectors. We find very little support for that in further testing, and do not explore this here. Changing the beginning and ending months, within the limits detailed above, does not qualitatively change the test results.

Table 2 proceeds to estimate vector error correction models, under the assumption that in each sub-period there is exactly one cointegrating vector. The number of lags

⁸Results available upon request. All series were tested using the GLS version of Dickey-Fuller, separately for each sub-period. We could not reject the null of a unit root at the 5% level for any of the series.

⁹See Dvir and Rogoff (2009) for an extensive discussion of the 1973 crisis and for details on the identification of 1973 as the break point.

¹⁰Results of the Johansen tests are not sensitive to the choice of lag number. The VECM estimates are sensitive to this choice, in size but not in sign or significance.

and periods is the same as in Table 1. Note that the coefficient for log of inventories is normalized to one. The table shows the coefficients of the lagged variables in the estimated cointegration equation only¹¹. Note that for both sub-periods the cointegrating equations are extremely significant. All coefficients are significant at the 1% or 5% level.

A number of interesting relationships are shown in the table. Note that a negative sign implies that the variable has a positive long-run relationship with inventories:

1. In both sub-periods, we see that inventories and oil production co-move: as production increases, inventories tend to increase as well.

2. In the early sub-period, demand and inventories move in opposite directions, however in the late sub-period they co-move in the same direction.

3. In the early sub-period, price and inventories move in opposite directions, however in the late sub-period they co-move in the same direction.

These long-run relationships are consistent with our model's predictions. First, inventories increase with production since higher production implies lower price relative to the future. Second, inventories tend to fall with a rise in income when supply is flexible (early period), however they tend to rise with income when supply is inflexible (late period). Finally, inventories fall when prices rise if supply is flexible (since the rise is expected to be temporary), but rise with prices if supply is restricted (since the rise is expected to persist).

It is important to stress that these estimates represent the long-run relationship among the variables, i.e. there is no claim here of causality from any one variable to the other, rather the finding is of a long-run stationary link. This strongly supports the relevance of a model which posits such a link among the variables. The fact that the signs seem to accord well with our model is encouraging. Experimenting with different starting and ending points, as well as varying the lag order, do not change the signs of the coefficients, nor the cointegration rank, nor the significance of the cointegrating equation or the estimated coefficients.

4 Conclusion

This paper presents evidence that important variables in the market for crude oil are connected by stable relationships, and have been at least since the 1930s. This evidence, of a single cointegrating vector connecting U.S. oil production, U.S. oil stocks, U.S. industrial production, and the real price of crude oil, turns out to accord quite well with an extended storage model which allows for income growth dynamics and

¹¹More results are available upon request. In particular, both VECMs are stable, and the we can reject the null of nonstationarity for both estimated cointegrating vectors at the 5% level using Dickey-Fuller GLS.

for changes in supply regimes. In particular, before 1973, when supply was unrestricted, stocks were negatively associated with demand (as measured by industrial production) and with price, and positively associated with U.S. oil production. After 1973, when supply became restricted, the relationship changed, and inventories became positively associated with demand and with price, while still positively associated with oil production. These stable relationships which exist in the data have so far not been used for forecasting and analysis purposes (See Alquist et al. [2011]), a fact which presents a potential opportunity to increase forecast accuracy. This is a subject for future research.

References

- [1] Alquist, Ron, Lutz Kilian, and Robert J. Vigfusson. 2011. "Forecasting the Price of Oil." Working paper.
- [2] Barsky, Robert B. and Lutz Kilian. 2004. "Oil and the Macroeconomy Since the 1970s." *Journal of Economic Perspectives*, 18(1): 115-134.
- [3] Chambers, Marcus J. and Roy E. Bailey. 1996. "A Theory of Commodity Price Fluctuations." *Journal of Political Economy*, 104(5): 924-957.
- [4] Commodity Research Bureau. 1940. *Commodities in Industry*. New York: CRB, Inc.
- [5] Commodity Research Bureau. 1950, 1960. *Commodity Yearbook 1950, 1960*. New York: CRB, Inc.
- [6] Deaton, Angus and Guy Laroque. 1992. "Behaviour of Commodity Prices." *Review of Economic Studies*, 59(1): 1-23.
- [7] Deaton, Angus and Guy Laroque. 1996. "Competitive Storage and Commodity Price Dynamics." *Journal of Political Economy*, 104(5): 896-923.
- [8] Dvir, Eyal and Kenneth Rogoff, 2009. "Three Epochs of Oil." NBER Working Paper No. 14927.
- [9] Fattouh, Bassam, Lutz Kilian, and Lavan Mahadeva, 2012. "The Role of Speculation in Oil Markets: What Have We Learned So Far?", Working paper.
- [10] Hamilton, James D. 2008. "Oil and the Macroeconomy." in *The New Palgrave Dictionary of Economics, 2nd Edition*, eds. Steven N. Durlauf and Lawrence E. Blume. New York: Palgrave Macmillan.
- [11] Hamilton, James D. 2009a. "Causes and Consequences of the Oil Shock of 2007-08." *Brookings Papers on Economic Activity*, Spring 2009: 215-59.
- [12] Hamilton, James D. 2009b. "Understanding Crude Oil Prices." *Energy Journal*, 30(2): 179-206.
- [13] Hotelling, Harold. 1931. "The Economics of Exhaustible Resources." *Journal of Political Economy* 39(2): 137-175.

- [14] International Monetary Fund. 2008. *World economic Outlook*. Washington, DC.
- [15] Gustafson, Robert L. 1958. "Carryover Levels for Grains: A Method for Determining Amounts That Are Optimal under Specified Conditions." *USDA Technical Bulletin* 1178.
- [16] Judd, Kenneth L. 1998. *Numerical Methods in Economics*. Cambridge, MA: MIT Press.
- [17] Kilian, Lutz. 2008a. "The Economic Effects of Energy Price Shocks." *Journal of Economic Literature*, 46(4): 871-909.
- [18] Kilian, Lutz. 2008b. "Exogenous Oil Supply Shocks: How Big Are They and How Much Do They Matter for the U.S. Economy?" *Review of Economics and Statistics*, 90(2): 216-240.
- [19] Kilian, Lutz. 2009. "Not All Oil Price Shocks Are Alike: Disentangling demand and Supply Shocks in the Crude Oil Market." *American Economic Review* 99(3): 1053-69.
- [20] Miranda, Mario J. and Paul L. Fackler. 2002. *Applied Computational Economics and Finance*. Cambridge, MA: MIT Press.
- [21] NBER. 2008. "NBER Macrohistory Database". <http://www.nber.org/databases/macroeconomy/contents/>
- [22] Pavlova, Anna, and Suleyman Basak, 2012. "A Model of Financialization of Commodities", Working paper, London Business School
- [23] Pirrong, Craig. 2009. "Stochastic Fundamental Volatility, Speculation, and Commodity Storage." U. of Houston.
- [24] Routledge, Bryan R., Duane J. Seppi, and Chester S. Spatt. 2000. "Equilibrium Forward Curves for Commodities." *Journal of Finance*, 55(3): 1297-1338.
- [25] Schwartz, Eduardo, and James E. Smith. 2000. "Short-Term Variations and Long-Term Dynamics in Commodity Prices." *Management Science*, 46(7): 893-911.
- [26] Singleton, Kenneth J. 2012. "Investor Flows and the 2008 Boom/Bust in Oil Prices." Stanford GSB.
- [27] Sockin, Michael, and Wei Xiong, 2012. "Feedback Effects of Commodity Futures Prices," Working paper, Princeton.
- [28] Stock, James H. and Mark W. Watson. 1993. "A Simple Estimator of Cointegrated Vectors in Higher Order Integrated Systems." *Econometrica*, 61(4): 783-820.
- [29] Williams, Jeffrey C. and Brian D. Wright. 1991. *Storage and Commodity Markets*. Cambridge, England: Cambridge University Press.
- [30] Wright, Brian D. 2001. "Storage and Price Stabilization." in *Handbook of Agricultural Economics*. Vol 1B, *Marketing, Distribution, and Consumers*, eds.

Bruce L. Gardner and Gordon C. Rausser, 817-862. New York: Elsevier.

Table 1: Johansen Tests for the Existence of Cointegration Vectors

	Column I	Column II
Period	1931/5 - 1972/12	1975/1 - 2011/12
Cointegrating Rank	0	1
Trace Statistic	98.89***	54.70***
5% Critical Value	47.21	29.68
1% Critical Value	54.46	35.65
Obs.	500	444
Differenced Lags	3	1

Tests include a constant and seasonal dummies. Number of lags chosen by Akaike information criterion. (***) denotes that the trace statistic for the applicable rank is larger than the 1% critical value. (**) denotes that the trace statistic for the applicable rank is larger than the 5% critical value.

Table 2: Long-Run Relationships of Stocks, Production, Demand, and Price

	Column I	Column II
Period	1931/5 - 1972/12	1975/1 - 2011/12
ln Stocks _t	1	1
ln Oil_Production _t	-6.80*** (1.12)	-1.02*** (0.27)
ln Industrial_Production _t	3.58*** (0.68)	-0.65*** (0.20)
ln Price _t	3.98*** (0.47)	-0.10** (0.04)
Obs.	500	444
Differenced Lags	3	1
χ^2 (p-value)	75.54 (<0.0001)	18.27 (0.0004)

Data sources: see text. Three asterisks (***) denote significance at the 1% level, two asterisks (**) denote significance at the 5% level. Standard errors are shown in parentheses. See text for definition of variables. All regressions include a constant and seasonal dummies (not shown).

Figure 1: RE Equilibrium: Storage and Price As Functions of Effective Availability

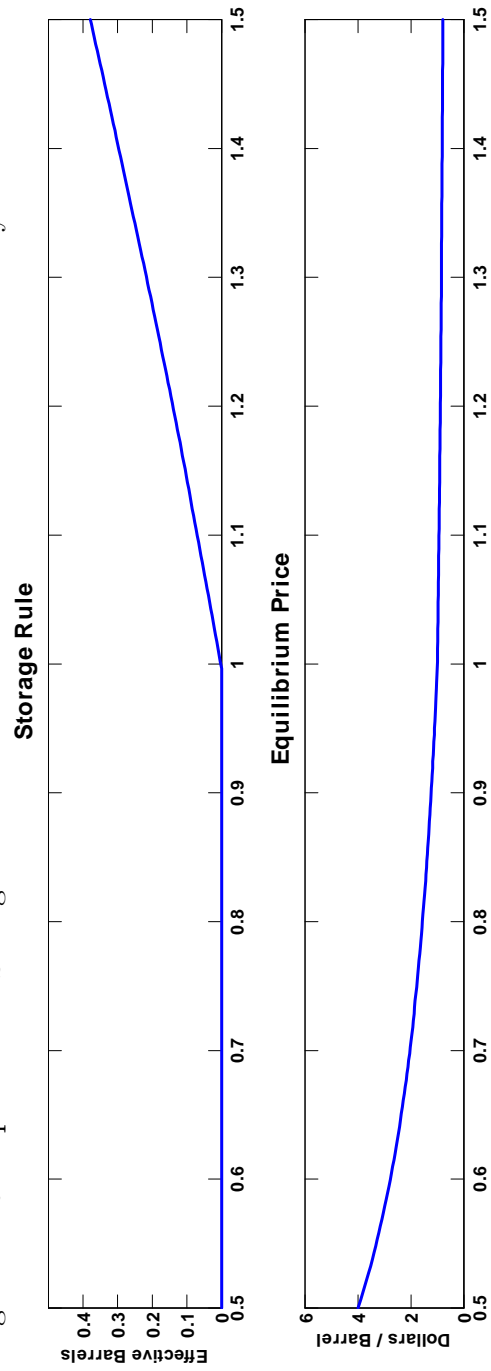


Figure 2: Effect of Change in Relative Income on Storage Across Models

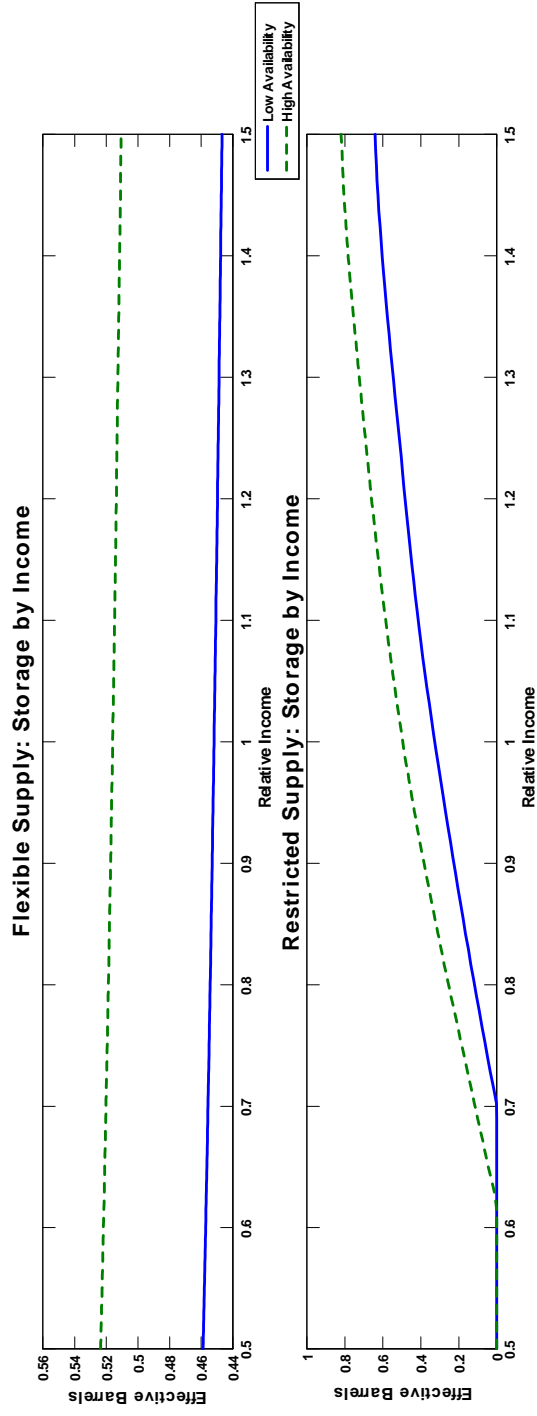


Figure 3: 1920/1 - 1972/12

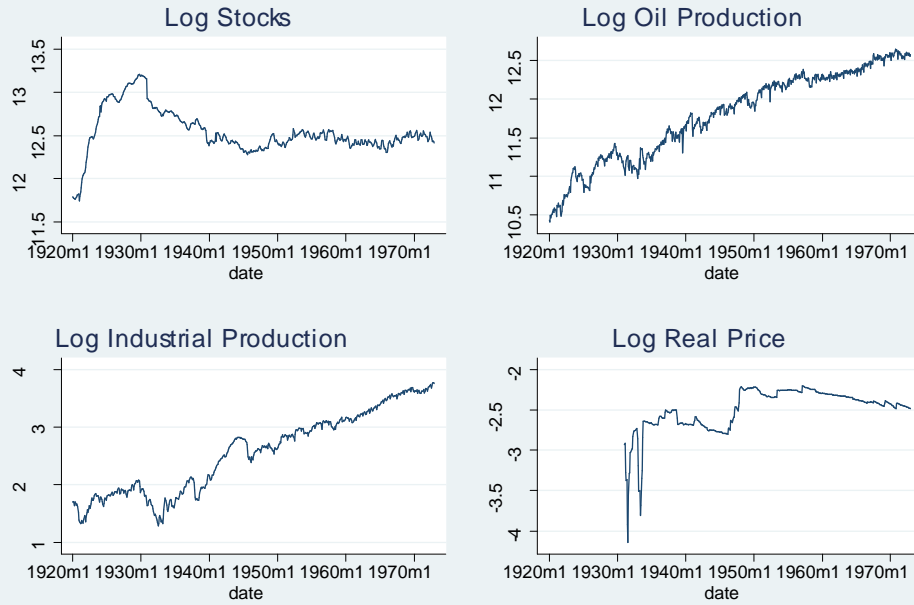


Figure 4: 1973/1 - 2011/12

