The Role of Inventories and Speculative Trading in the Global Market for Crude Oil

Lutz Kilian      Dan Murphy
University of Michigan       University of Michigan
CEPR

PRELIMINARY
February 1, 2010

Abstract: We develop a structural model of the global market for crude oil that explicitly allows for shocks to the speculative demand for oil as well as shocks to the flow demand and flow supply. We show how one can identify the forward-looking element of the real price of oil with the help of data on oil inventories. The model estimates allow us to reject explanations of the 2003-08 oil price surge based on unexpectedly diminishing oil supplies (as in the peak-oil hypothesis) and explanations based on speculative trading. Instead, we find that this surge was caused by fluctuations in the flow demand for oil driven by the global business cycle. There is evidence, however, that speculative demand shifts played an important role during earlier oil price shock episodes including 1979 and 1990. Recently, it has been suggested that it is possible for speculative trading to occur even without any change in inventory holdings, if the short-run price elasticity of oil demand is zero. Our model encompasses that possibility. The use of a structural model allows us to obtain an estimate of this elasticity based on shifts of the supply curve along the demand curve. Even after accounting for the role of inventories in smoothing oil consumption, our estimates of the price elasticity of oil demand are not close to zero and much higher than traditional estimates from dynamic series models that do not account for price endogeneity.

Key words: Oil market; speculation; fundamentals; peak oil; inventories; demand; supply; oil demand elasticity; gasoline demand elasticity; structural model; identification.

JEL: Q41, Q43, Q48, D84

Acknowledgements: We thank Ana María Herrera, Ryan Kellogg, Chris Knittel, and Whitney Newey for helpful discussions.

Correspondence to: Lutz Kilian, Department of Economics, 611 Tappan Street, Ann Arbor, MI 48109-1220, USA. Email: lkilian@umich.edu.
1. Introduction

The dramatic price fluctuations in the global market for crude oil since 2003 have renewed interest in understanding the evolution of the real price of oil. There is much debate in policy circles about whether the surge in the real price of oil between 2003 and mid-2008 was caused by speculative trading. An alternative view is that unexpected reductions in oil supplies are to blame in the form of OPEC withholding oil supplies from the market or because global oil production has peaked, as predicted by the peak oil hypothesis. Yet another view is that this surge in the real price of oil was driven instead by unexpectedly strong economic growth in the global economy, in particular in emerging Asia.

The relative importance of these explanations is important to policymakers. To the extent that speculative trading is perceived to be the core of the problem, for example, there has been considerable political pressure recently to impose regulatory limits on trading in oil futures markets. If dwindling oil supplies are the problem, in contrast, there is little U.S. policymakers can do to avoid similar price surges but to promote energy conservation and the use of alternative sources of energy or perhaps to put political pressure on oil producers. Finally, if surges in the global business cycle are the chief cause of high oil prices, then efforts aimed at reviving the global economy after the financial crisis are likely to cause the real price of oil to recover as well, creating a policy dilemma.

This policy discussion has been accompanied by renewed debate among academic researchers about how much oil supply shocks matter for the real price of oil relative to speculative demand shocks and business-cycle driven demand shocks.1 This debate has far-reaching implications for the specification of empirical models and for the design of theoretical models of the transmission of oil price shocks. Despite much progress in recent years, there is no consensus in the academic literature on how to model the global market for crude oil. One strand of the literature views oil as an asset, the price of which is determined by desired stocks. In this interpretation, shifts in the expectations of forward-looking traders are reflected in changes in the real price of oil and changes in oil inventories. The other strand of the literature views the price of oil as being determined by shocks to the flow supply of oil and flow demand for oil with little attention to the role of inventories in smoothing oil consumption. Much of the research on oil supply shocks is in that tradition, as are economic models linking the real price of oil to

---

1 See, e.g., Baumeister and Peersman (2009); Einloth (2009); Hamilton (2009a,b); Kilian (2009a,b).
fluctuations in the global business cycle.

Recently, there has been increasing recognition that both elements of price determination matter in modeling the market for oil (see, e.g., Dvir and Rogoff 2009; Frankel and Rose 2009; Hamilton 2009a,b; Kilian 2009a; Alquist and Kilian 2010). In section 2, we propose a structural vector autoregressive (VAR) model of the global market for crude oil that explicitly embeds these two explanations of the determination of the real price of oil. Using a new approach to identification, discussed in section 3, we show how the forward-looking element of the real price of oil can be identified with the help of data on crude oil inventories. The proposed dynamic simultaneous equations model allows us to separate the speculative component of the real price of oil from the components driven by flow demand and flow supply and to characterize the relative importance of each type of demand and supply shock for the real price of oil and for changes in oil inventories.

This feature allows to study the role of storage in oil markets and to shed light on the extent of speculation in oil markets in section 4. The model provides several new insights. First, the model estimates rule out speculation as a potential cause of the surge in the real price of oil between 2003 and mid-2008. This finding implies that additional regulation of these markets would not have stemmed the increase in the real price of oil.

Second, although speculative trading does not explain the recent surge in the real price of oil, we show that it played an important role in several earlier oil price shock episodes. For example, it was a central feature of the oil price shock of 1979, following the Iranian Revolution, consistent with the narrative evidence in Barsky and Kilian (2002), and it helps explain the sharp decline in the real price of oil in early 1986 after the collapse of OPEC. It also was an important contributing factor in the decline in the real price of oil in 1997-99, following the Asian financial crisis, and again in late 2008, following the global financial crisis. In both cases, speculative trading amplified the effect of the global business cycle on the real price of oil. Finally, we document that neither negative oil supply shocks nor positive speculative demand shocks alone can explain the oil price spike and oil inventory behavior of 1990/91. In particular, we show that about one third of the oil price increase in August of 1990 was caused by speculative demand and two thirds by oil supply shocks, whereas negative speculative demand shocks were responsible for almost all of the decline in the real price of oil starting in late 1990, when the threat to Saudi oil fields had been alleviated.
Third, we document that unexpected fluctuations in global real activity alone explain virtually the entire surge in the real price of oil between 2003 and mid-2008, even acknowledging that negative oil supply shocks raised the real price of oil slightly. Business cycle factors were also responsible for the bulk of the 1979/80 oil price increase in conjunction with sharply rising speculative demand in 1979. Oil supply shocks played only a minor role in 1979. The continued rise in the real price of oil in 1980 reflected negative oil supply shocks (largely caused by the outbreak of the Iran-Iraq War) as much as continued (if slowing) global growth, amidst declining speculative demand. Finally, there is evidence that the recovery of the real price of oil starting in 1999, following an all-time low in postwar history, was aided by coordinated oil supply cuts. Although our analysis assigns more importance to oil supply shocks than some previous studies, we conclude that, with the exception of 1990, the major oil price shocks were driven primarily by oil demand shocks.

Much of the prima facie case against an important role for speculative trading rests on the fact that there has been no noticeable increase in inventory accumulation in recent years (see, e.g., Kilian 2010). Recently, Hamilton (2009a) suggested that, as a matter of theory, speculative trading may cause a surge in the real price of oil without any change in oil inventory holdings if the short-run price elasticity of gasoline demand (and hence the short-run price elasticity of oil demand) is zero. Thus, it is essential that we pin down the value of these elasticities.

Hamilton observed that existing estimates of this elasticity in the literature are close to zero, lending credence to this model of speculation. These estimates, however, are based on dynamic reduced form regressions that ignore the endogeneity of the real price of oil. They have no structural interpretation and suffer from downward bias. In section 5, we address this limitation with the help of our structural VAR model. That model nests the limiting case discussed in Hamilton. Not only do our response estimates show that speculative demand shocks are associated with systematic inventory building, but the econometric model allows us to construct a direct estimate of the elasticity parameter based on exogenous shifts of the oil supply curve along the oil demand curve. Our median estimate of the short-run price elasticity of oil demand of -0.41 is seven times higher than standard estimates in the literature.

That estimate, however, like all earlier estimates, is misleading because it ignores the role of oil inventories in smoothing oil consumption. Refiners hold crude oil inventories as insurance against unexpected disruptions of crude oil supplies or unexpected increases in the demand for
refined products such as gasoline. The decline in crude oil inventories in response to a negative oil supply shock increases the magnitude of the quantity change relevant for computing the oil demand elasticity. Our structural model is designed to allow the estimation of this alternative demand elasticity. The median estimate of this short-run price elasticity of oil demand in use is -0.20. Although much lower than the corresponding traditional elasticity estimate, even that estimate is three times higher than the consensus view in the literature and distinctly higher than zero.

We provide a theoretical benchmark for thinking about the relationship between the short-run price elasticity of oil demand in use and the short-run price elasticity of gasoline demand. We show that under reasonable assumptions about the oil refining industry, the latter is about as high as the price elasticity of oil demand in use. That gasoline demand elasticity estimate is much larger than some recent estimates in the literature. For example, Hughes, Knittel and Sperling (2008) in an influential study report a short-run U.S. gasoline elasticity estimate of between -0.03 and -0.08, which is well below our estimate, illustrating the importance of structural modeling in constructing estimates of the price elasticity of energy demand. In section 6, we relate our estimate of the short-run price elasticity of gasoline demand to the literature on estimating long-run U.S. gasoline demand elasticities from cross-sectional data. The concluding remarks are in section 7.

2. VAR Methodology
Our analysis is based on a four-variable dynamic simultaneous equations model in the form of a structural VAR. Let $y_t$ be a vector of endogenous variables including the percent change in global crude oil production, a measure of global real activity expressed in percent deviations from trend, the real price of crude oil, and the change in oil inventories. All data are monthly and the sample period is 1973.2-2008.9. We remove seasonal variation by including seasonal dummies in the VAR model.

2.1. Data
Data on global crude oil production are available in the Monthly Energy Review of the Energy Information Administration (EIA). Our measure of fluctuations in global real activity is the dry

---

2 For further discussion of this convenience yield provided by oil inventories see Alquist and Kilian (2010) and the references therein.
cargo shipping rate index developed in Kilian (2009a). For more details on the rationale, construction and interpretation of this index the reader is referred to the related literature. While there are other indices of global real activity available, none of these alternative proxies is as appropriate for our purpose of measuring shifts in the global demand for industrial commodities and none is available at monthly frequency for our sample period.

The real price of oil is defined as the U.S. refiners’ acquisition cost for imported crude oil, as reported by the EIA, extrapolated from 1974.1 back to 1973.1 as in Barsky and Kilian (2002) and deflated by the U.S. consumer price index. We use the refiners’ acquisition cost for imported crude oil because that price is likely to be a better proxy for the price of oil in global markets than the U.S. price of domestic crude oil which was regulated during the 1970s and early 1980s.

Given the lack of data on crude oil inventories for other countries, we follow Hamilton (2009a) in using the data for U.S. crude oil inventories provided by the EIA. These data are scaled by the inverse ratio of U.S. petroleum stocks in OECD petroleum stocks for each time period. That scale factor ranges from about 2.23 to 2.59 in our sample. We express the resulting proxy for global crude oil inventories in changes rather than percent changes. One reason is that the percent change in inventories does not appear to be covariance stationary, whereas the change in inventories does. The other reason is that the proper computation of the oil demand

---

3 This index has also been employed by Alquist and Kilian (2010), Bahattin and Robe (2009), Fukunaga, Hirakata and Sudo (2009), Kilian (2009a,b, 2010), Kilian, Rebucci and Spatafora (2009), and Kilian and Park (2009), among others.

4 The idea of using fluctuations in dry cargo freight rates as indicators of shifts in the global real activity dates back to Isserlis (1938) and Tinbergen (1959); also see Stopford (1997). One advantage of using this type of index is that it automatically accounts for any additional demand for industrial commodities generated by the depreciation of the U.S. dollar in recent years. The panel of monthly freight-rate data underlying the global real activity index was collected manually from *Drewry’s Shipping Monthly* using various issues since 1970. The data set is restricted to dry cargo rates. The earliest raw data are indices of iron ore, coal and grain shipping rates compiled by *Drewry’s*. The remaining series are differentiated by cargo, route and ship size and may include in addition shipping rates for oilseeds, fertilizer and scrap metal. In the 1980s, there are about 15 different rates for each month; by 2000 that number rises to about 25; more recently that number has dropped to about 15. The index was constructed by extracting the common component in the nominal spot rates. The resulting nominal index is expressed in dollars per metric ton and was deflated using the U.S. CPI and detrended to account for the secular decline in shipping rates. For this paper, this series has been extended based on the Baltic Exchange Dry Index, which is available from Bloomberg. The latter index, which is commonly discussed in the financial press, is essentially identical to the nominal index in Kilian (2009a), but only available since 1985.

5 Petroleum stocks as measured by the EIA include crude oil (including strategic reserves) as well as unfinished oils, natural gas plant liquids, and refined products. The EIA does not provide petroleum inventory data for non-OECD economies. We treat the OECD data as a proxy for global petroleum inventories. Consistent series for OECD petroleum stocks are not available prior to 1987.12. We therefore extrapolate the percent change in OECD inventories backwards at the rate of growth of U.S. petroleum inventories. For the period 1987.12-2009.5, the U.S. and OECD petroleum inventory growth rates are reasonably close with a correlation of about 80%.
elasticity, as discussed below, requires an explicit expression for the change in global crude oil inventories in barrels. This computation is only possible if oil inventories are specified in changes rather than percent changes.6

2.2. A Model of the Global Market for Crude Oil

The reduced-form model allows for two years worth of lags. This approach is consistent with evidence in Hamilton and Herrera (2004) and Kilian (2009a) on the importance of allowing for long lags in the transmission of oil price shocks. The real activity index of Kilian (2009a) is stationary by construction. We measure fluctuations in real activity in percent deviations from trend. Oil production is expressed in percent changes in the model, whereas oil inventories are expressed in differences. Preliminary tests provided no evidence of cointegration between oil production and oil inventories. Following Kilian (2009a), the real price of oil is expressed in log-levels.7 The corresponding structural model of the global oil market may be written as

\[ B_0 y_t = \sum_{i=1}^{24} B_i y_{t-i} + \epsilon_t, \]

where \( \epsilon_t \) is a 4×1 vector of orthogonal structural innovations and \( B_i, i = 0, \ldots, 24 \), denotes the coefficient matrices. The seasonal dummies have been suppressed for notational convenience.

The vector \( \epsilon_t \) consists of a shock to the flow of the production of crude oil ("flow supply shock"), a shock to the flow demand for crude oil and other industrial commodities ("flow demand shock"), a shock to the demand for oil inventories arising from forward-looking behavior ("speculative demand shock"), and a residual shock that captures all structural shocks not

---

6 Note that we focus on above-ground inventories in this paper. A different strand of the literature has considered oil below the ground as another form of oil inventories. We do not pursue this possibility, because oil below the ground is not fungible with oil above ground in the short run and because no reliable time series data exist on the quantity of oil below the ground. We do discuss, however, how speculation based on below-ground inventories would be recorded within our model framework and how it may be detected in section 4.3.

7 One reason is that economic theory suggests a link between cyclical fluctuations in global real activity and the real price of oil and hence the real price of gasoline (see Barsky and Kilian 2002). Differencing the real price series would remove that slow-moving component and eliminate any chance of detecting persistent effects of global shocks to the demand for all industrial commodities including crude oil. Moreover, even granting that the real price of oil is well approximated by a random walk in prediction, it is not clear a priori whether there is a unit root in the real price of oil. The level specification adopted in this paper has the advantage that the impulse response estimates are not only asymptotically valid under the maintained assumption of a stationary real price of oil, but robust to departures from that assumption, whereas incorrectly differencing the real price of oil would cause these estimates to be inconsistent. The potential cost of not imposing unit roots in estimation is a loss of asymptotic efficiency, which would be reflected in wider error bands. Since the impulse response estimates presented below are reasonably precisely estimated, this is not a concern in this study. It should be noted, however, that historical decompositions for the real price of oil rely on the assumption of covariance stationarity and would not be valid in the presence of unit roots.
otherwise accounted for and has no direct economic interpretation.

The flow supply shock corresponds to the classical notion of an oil supply shock, as discussed in the literature. The flow demand shock can be thought of as a demand shock reflecting the state of the global business cycle. The speculative demand shock is designed to capture innovations to the demand for crude oil that reflect revisions to expectations about future demand and future supply as distinct from innovations to the current flow supply or flow demand. We identify these shocks based on a combination of sign restrictions on the structural impulse response functions and other economically motivated restrictions.

Our model allows for alternative views of the determination of the real price of oil conditional on lagged data. One view is that the real price of crude oil is determined by the current flow supply of oil and the current flow demand for oil. The flow supply of crude oil is measured by the global production of crude oil. An unexpected disruption of that flow (embodied in a shift to the left of the contemporaneous oil supply curve along the oil demand curve) within the month will raise the real price of crude oil, will cause oil inventories to be drawn down in an effort to smooth consumption, and will lower global real activity, if it has any effect on real activity at all.

The flow demand for crude oil is driven by unexpected fluctuations in global real activity. These represent shifts in the demand for all industrial commodities including crude oil associated with the global business cycle. An unanticipated increase in global real activity (embodied in a shift to the right of the contemporaneous oil demand curve along the oil supply curve) within the month will raise the real price of oil, will cause oil inventories to be drawn down, and will stimulate global oil production, if it has any effect on oil production at all. The latter stimulus, however, is bound to be small on impact, given the consensus in the literature that the short-run price elasticity of oil supply is near zero.

This standard view of the global crude oil market is incomplete, however. Given that crude oil is storable, it may also be viewed as an asset, the real price of which is determined by the demand for inventories (see, e.g., Frankel and Rose 2009; Kilian 2009a; Alquist and Kilian 2010). This means that we must allow the price of oil to jump in response to any news about future oil supply or future oil demand. For example, upward revisions to expected future demand

---

8 A well documented fact is that there is considerable comovement between the real price of crude oil and the real price of other industrial commodities during times of major fluctuations in global real activity (see, e.g., Kilian 2009b).
for crude oil (or downward revisions to expected future production of crude oil), all else equal, will increase the demand for crude oil inventories in the current period, resulting in an instantaneous shift of the contemporaneous demand curve for oil along the oil supply curve and an increase in the real price of oil. Although there is no direct effect of such a shock on global real activity within the month by construction because the shock is defined as an innovation about future global real activity or oil production, to the extent that the real price of oil jumps on impact, one would expect this shock to lower global real activity, if it has any impact on real activity at all. One would also expect the impact response of global oil production to such an expectations shock to be positive, if oil production responds at all. This follows from the premise that any news about future oil production shortfalls will not have any direct effect on current oil production by construction.9

News about future oil production and future demand for crude oil are but one example of shocks to expectations in the global market for crude oil. An unexpected increase in the uncertainty about future oil supply shortfalls would have the much same effect. This point has been demonstrated formally in a general equilibrium model by Alquist and Kilian (2010). The main difference is that uncertainty shocks would not be associated with expected changes in future oil production or real activity.10 In this paper we refer to any oil demand shock that reflects shifts in expectations about future oil production or future real activity as a speculative oil demand shock. The next section discusses how we distinguish speculative demand shocks from flow demand shocks and flow supply shocks in practice.

3. Identification

The structural VAR model is set-identified based on a combination of sign restrictions, shape restrictions, and bounds on the implied price elasticities of oil demand and oil supply. Some of these restrictions are implied by the economic model discussed in section 2, while others can be motivated based on extraneous information. We impose six sets of identifying restrictions, each of which is discussed in turn.

---

9 If there is an effect, it must arise from the price incentive for higher production created by the speculative demand shock. That effect is likely to be small, given the consensus in the literature that the short-run price elasticity of oil production is near zero.

10 Our approach is consistent with the observation in Dvir and Rogoff (2009) that the role of storage is speculative when rapid industrialization coincides with uncertain oil supplies, as has been the case in the post-1973 period that we focus on in this paper.
3.1. Impact Sign Restrictions

The sign restrictions on the impact responses of oil production, real activity, the real price of oil and oil inventories are summarized in Table 1. These restrictions directly follow from the model discussed in section 2. All sign restrictions involve weak inequalities and allow for the response to be zero. Implicitly, these restrictions also identify the residual innovation. Given the difficulty of interpreting this residual economically, we do not report results for the residual shock.

Sign restrictions alone are typically too weak to be informative about the effects of oil demand and oil supply shocks. As demonstrated in Kilian and Murphy (2009) in the context of a simpler model, it is important to impose all credible identifying restrictions to allow us to narrow down the set of admissible structural models.11 One such set of restrictions relates to bounds on impact price elasticities of oil demand and oil supply.

3.2. Bound on the Impact Price Elasticity of Oil Supply

An estimate of the impact price elasticity of oil supply may be constructed from the dynamic simultaneous equations model (1) by evaluating the ratio of the impact responses of oil production and of the real price of oil to an unexpected increase in aggregate demand or in speculative demand. There is a consensus in the literature that this short-run price elasticity of oil supply is close to zero (see Hamilton 2009a, Kilian 2009a). Neither Hamilton (2009a,b) nor Kilian (2009a) provide econometric estimates of the short-run elasticity of oil supply. Here we follow Kilian and Murphy (2009) in imposing an upper bound on the impact oil supply elasticity. This bound is important because many models identified purely based on sign restrictions imply an impact oil supply elasticity that is far too large to be economically plausible.

An empirically plausible bound may be obtained by focusing on the historical episodes of well-defined and exogenous oil price shocks such as the outbreak of the Persian Gulf War on August 2, 1990. That event caused a sharp spike in the real price of oil. Given the timing of that event, production data for August of 1990 are informative about the oil supply response in the rest of the world. Because prior to the war there had been spare capacity in oil production, as discussed in Kilian (2008a), and because there was a concerted effort to increase global oil production following the invasion, this response provides a plausible upper bound for the with-in-the-month elasticity of oil supply. In August of 1990, the global production of crude oil from all oil producers excluding Iraq and Kuwait increased by 1.17%, whereas the real price jumped by

11 For a similar point also see Canova and De Nicolo (2002), Uhlig (2005), and Canova and Paustian (2007).
45.46% which suggests an oil supply elasticity of 0.0257.\textsuperscript{12} We impose this upper bound on the impact price elasticity of oil supply in selecting the set of admissible models. It is important to stress that this additional identifying restriction only relates to the impact elasticity rather than the dynamic responses and that it does not constrain the level of the impact responses, but merely imposes a bound on their relative magnitude. While we prefer to work with an empirically grounded bound on this elasticity, it can be shown that relaxing this bound somewhat, say to 0.05, has little effect on the impulse response estimates. Even for a bound of 0.10 our results remain qualitatively similar.

### 3.3. Bound on the Impact Price Elasticity of Oil Demand

A preliminary estimate of the impact price elasticity of oil demand may be constructed from the estimated model (1) by evaluating the ratio of the impact responses of oil production and of the real price of oil to an unexpected oil supply disruption. This oil demand elasticity in production corresponds to the standard definition of the oil demand elasticity used in the literature. It equates the production of oil with the consumption of oil. In the presence of changes in oil inventories that assumption is inappropriate. The relevant quantity measure instead is the sum of the flow of oil production and the depletion of oil inventories triggered by an oil supply shock. To our knowledge, this distinction has not been discussed in the literature nor has there been any attempt in the literature to estimate this oil demand elasticity in use. A natural additional identifying assumption is that the impact elasticity of oil demand in use must be weakly negative on average over the sample. We defer a formal definition of this elasticity to section 5.2. Note that we do not need to restrict the oil demand elasticity in production. Our impact sign restrictions ensure that this elasticity is weakly negative on impact.

### 3.4. Relative Order of Magnitude Restrictions on Impact Responses

The impact elasticity of oil supply is estimated based on a shift of the oil demand curve along the supply curve. Because there are two oil demand shocks in the model, the model allows the construction of two estimates of the impact price elasticity of oil supply. In population, these elasticities must be identical. The reason is that the effect of both oil demand shocks on oil production in the impact period by construction operates exclusively through the real price of oil.

\textsuperscript{12} Source: Energy Information Administration. The real price of oil increased from $16.19 per barrel to $23.55 per barrel from July to August, which translates to $12.67 and $18.43, respectively, in real terms. Oil supply from all producers excluding Iraq and Kuwait increased from 55,197 barrels per day to 55,844 barrels per day.
Although we would expect some sampling variation in these estimates, it seems reasonable to discard all structural models that imply estimates of the same elasticity that are different by an order of magnitude. This means, for example, that we would discard a model solution in which one of the impact elasticities of oil supply is 0.001, while the other is 0.013.

Likewise the model implies that the ratio of the impact response of real activity to the impact response of the real price of oil should be the same in response to negative oil supply shocks as in response to positive speculative demand shocks. In both cases, the impact response of real activity is driven by the response of the real price of oil. In particular, a speculative demand shock by construction will not affect real activity directly within the month. It is important to stress that these additional identifying assumptions impose no further restrictions on the ratio or the magnitude of the level of the impact responses. They merely represent cross-equation restrictions.

3.5. Shape Restrictions
Rather than impose any dynamic sign restrictions on these responses, we exploit the fact that expectations shocks should result in a jump of the real price of oil on impact followed by a decline in the long run (see Alquist and Kilian 2010). This motivates the additional identifying restriction that the response of the real price of oil to a speculative demand shock must not be higher at the highest horizon considered than it is on impact. We do not take a stand on the evolution of the response function at intermediate horizons.

3.6. Dynamic Sign Restrictions
Our final set of restrictions relates to the dynamic response of oil production to an oil supply shock. This set of identifying restrictions is necessary to rule out structural models in which unanticipated oil supply disruptions cause a strong decline in the real price of oil below its starting level which is at odds with the standard view in the literature. We therefore impose the additional restriction that the response of the real price of oil to a negative oil supply shock must be positive for at least twelve months, starting in the impact period. Because the positive response of the real price of oil tends to be accompanied by a persistently negative response of oil production, once we impose this additional dynamic sign restriction, it furthermore must be the case that global real activity responds negatively to oil supply shocks. This is the only way for the oil market to experience higher prices and lower quantities in practice, because in the data
the decline of inventories triggered by an oil supply disruption is much smaller than the shortfall of oil production. This implies a joint set of sign restrictions such that the responses of oil production and global real activity to an unanticipated oil supply disruption are negative for the first twelve months, while the response of the real price of oil is positive.

In contrast, we do not impose any dynamic sign restrictions on the responses to oil demand shocks. In this respect our approach differs from Baumeister and Peersman (2009), for example. In particular, we do not impose any dynamic sign restrictions on the responses of global real activity and oil production to speculative oil demand shocks. The reasoning is as follows. The speculative demand shock in our model is a composite of three distinct types of expectations shocks. For example, a pure uncertainty shock (such as a mean-preserving increase in the spread) that raises precautionary demand will be associated with an increase in the real price of oil and higher oil inventories without any change in expected oil production or expected real activity (see Alquist and Kilian 2010). As a result, one would expect no response in oil production or in real activity, except to the extent that higher oil prices all else equal stimulate oil production over time or may reduce global real activity.

To the extent that a speculative shock involves upward revisions to expected future demand for crude oil instead, one would expect that shock to increase the real price of oil, oil inventories and future global real activity, but not to change oil production, except to the extent that higher oil prices stimulate global oil production over time. A classical example of such a situation would be speculation in anticipation of a booming world economy (or – with reverse signs – the anticipation of a major recession driven by a financial crisis).

In either of these first two cases, the response of oil production is expected to be weakly positive over time. In contrast, a speculative shock involving downward revisions to expected future crude oil production would be expected to cause higher oil prices, higher inventories, and lower future oil production, and possibly a decline in global real activity over time driven by higher oil prices. Thus, unlike in the earlier two cases, there is reason to expect the response of oil production to be negative over time. For example, it has been argued in the literature that the oil price spike of 1990, following the invasion of Kuwait by Saddam Hussein, did not only reflect the oil supply shock associated with the cessation of oil production in Iraq and Kuwait, but the specter of Iraq invading Saudi Arabia and future Saudi oil production ceasing (see Kilian 2008a). Likewise, Barsky and Kilian (2002) made the case that the increase in the price of oil
following the Iranian Revolution of 1979 reflected in part the fear of a wider regional conflagration involving the U.S. and/or moderate Gulf oil producers. A third example would be speculation by traders driven by the anticipation of declining future oil supplies (as implied by the peak oil hypothesis). Such speculation, if it did indeed occur, would have been associated with higher oil prices, higher inventories, and lower future oil production.\footnote{In contrast, speculation by oil producers, who choose to leave oil below the ground in anticipation of rising prices, as discussed in Hamilton (2009b), in our modeling framework would be classified as a shock to the flow supply of crude oil.}

Given the fact that historically expectations of future oil production reductions more often than not have failed to materialize within the sample, one would not expect the estimated response of oil production to the speculative demand shock to be strongly negative and it is conceivable that this response may actually be positive. Likewise, it is not clear whether the dynamic response of real activity should be positive, reflecting the realization of expectations of increased real activity, or negative, reflecting declines in real activity triggered by higher oil and other industrial commodity prices. Hence, we do not impose restrictions on the sign of the responses of real activity and oil production to speculative oil demand shocks beyond the impact period.

3.7. Implementation of the Identification Procedure

Given the set of identifying restrictions and consistent estimates of the reduced-form VAR model, the construction of the set of admissible structural models follows the standard approach in the literature on VAR models identified based on sign restrictions (see, e.g., Canova and De Nicolo 2002; Uhlig 2005). Consider the reduced-form VAR model $A(L) y_t = e_t$, where $y_t$ is the $N$-dimensional vector of variables, $A(L)$ is a finite-order autoregressive lag polynomial, and $e_t$ is the vector of white noise reduced-form innovations with variance-covariance matrix $\Sigma_{e_t}$. Let $\varepsilon_t$ denote the corresponding structural VAR model innovations. The construction of structural impulse response functions requires an estimate of the $N \times N$ matrix $\tilde{B}$ in $\varepsilon_t = \tilde{B} e_t$.\footnote{For a review of the construction of these structural impulse responses the reader is referred to Fry and Pagan (2010), for example.}

Let $\Sigma_{\varepsilon_t} = P \Lambda P'$ and $B = P \Lambda^{0.5}$ such that $B$ satisfies $\Sigma_{\varepsilon_t} = BB'$. Then $\tilde{B} = BD$ also satisfies $\tilde{B}B' = \Sigma_{\varepsilon_t}$ for any orthonormal $N \times N$ matrix $D$. One can examine a wide range of possibilities for $\tilde{B}$ by repeatedly drawing at random from the set $D$ of orthonormal rotation matrices $D$. 

---

13 In contrast, speculation by oil producers, who choose to leave oil below the ground in anticipation of rising prices, as discussed in Hamilton (2009b), in our modeling framework would be classified as a shock to the flow supply of crude oil.

14 For a review of the construction of these structural impulse responses the reader is referred to Fry and Pagan (2010), for example.
Following Rubio-Ramirez, Waggoner and Zha (2005) we construct the set $\tilde{B}$ of admissible models by drawing from the set $D$ of rotation matrices and discarding candidate solutions for $\tilde{B}$ that do not satisfy a set of a priori restrictions on the implied impulse responses functions. The procedure consists of the following steps:

1) Draw an $N \times N$ matrix $K$ of $NID(0,1)$ random variables. Derive the $QR$ decomposition of $K$ such that $K = Q \cdot R$ and $QQ' = I_N$.

2) Let $D = Q'$. Compute impulse responses using the orthogonalization $\tilde{B} = BD$. If all implied impulse response functions satisfy the sign restrictions, retain $D$. Otherwise discard $D$.

3) Repeat the first two steps a large number of times, recording each $D$ that satisfies the restrictions (and the corresponding impulse response functions).

The resulting set $\tilde{B}$ comprises the set of admissible structural VAR models.

4. Estimation Results

The identifying restrictions described in section 3 do not necessarily yield point-identified structural impulse responses. We generated 60 million rotations based on the reduced-form VAR estimate. 13 candidate models satisfied all identifying restrictions. For further analysis, we select the model that yields an impact price elasticity of oil demand in use (defined in detail in section 5.2) closest to the posterior median of this elasticity. This can be done without loss of generality, as the other admissible models yield virtually indistinguishable empirical results.

4.1. Responses to Oil Supply and Oil Demand Shocks

Figure 1 plots the responses of each variable in this benchmark model to the three oil supply and oil demand shocks (together with the corresponding pointwise 68% percentiles of the posterior distribution). All shocks have been normalized such that they imply an increase in the real price of oil. In particular, the oil supply shock refers to an unanticipated oil supply disruption.

Figure 1 illustrates that the role of storage differs depending on the nature of the shock. An oil supply disruption causes inventories to be drawn down in an effort to smooth production of refined products. A positive flow demand shock causes an initial decline in oil inventories, again reflecting production smoothing, followed by a complete rebuilding of inventories within the first year. A positive speculative demand shock causes a persistent increase in oil inventories.
An oil supply shock is associated with a reduction in global real activity and a persistent drop in oil production, but much of the initial drop in is reversed within the first year. The real price of oil rises only temporarily. It peaks after two months and falls as oil production recovers. After one year, the real price of oil falls below its starting value, as global real activity drops further. A positive shock to the flow demand for crude oil, in contrast, is associated with a persistent increase in global real activity. It causes a persistent hump-shaped increase in the real price of oil with a peak after one year. Oil production also rises somewhat, but only temporarily. Finally, a positive speculative demand is associated with an immediate jump in the real price of oil. The real price response overshoots, before declining gradually. The effects on global real activity and global oil production are largely negative, but small. These estimates imply a larger role for oil supply shocks than the structural VAR model in Kilian (2009a), for example, illustrating the importance of including oil inventories, and they produce economically more plausible responses of global real activity in response to oil supply and speculative demand shocks.

4.2. What Drives Fluctuations in Oil Inventories and in the Real Price of Oil?

Table 2 illustrates the relative importance of each of the structural shocks in driving the change in oil inventories. In the short run, 46% of the variation is driven by speculative demand shocks, followed by oil supply shocks with 27%. Flow demand shocks have a negligible impact with 1%. At long horizons, in contrast, the role of speculative demand shocks declines to 39% and that of flow demand shocks increases to 15%, while the contribution of oil supply shocks remains essentially unchanged at 23%. This evidence suggests that, on average, fluctuations in oil inventories mainly reflect speculation, but there also is an important element of production smoothing by refiners in response to oil supply shocks. This contrasts with a much larger role of flow demand shocks in explaining the variability of the real price of oil. For example, in the long run, 56% of the variation in the real price of oil can be attributed to flow demand shocks, compared with 32% due to speculative demand shocks and 3% due to flow supply shocks.

Impulse responses and variance decompositions are useful in studying average behavior. To understand the historical evolution of the real price of oil, especially following major exogenous events in oil markets, it is more useful to compute the cumulative effect of each shock on the real price of oil and on the change in oil inventories. Figure 2 allows us to answer not only the question of how important the speculative component in the real price of oil was between
2003 and mid-2008, but we can assess the quantitative importance of speculation at each point in time since the late 1970s.\textsuperscript{15}

4.3. Did Speculators Cause the Oil Price Shock of 2003-2008?

A common view in the literature is that speculators caused part or all of the run-up in the real price of oil between 2003 and mid-2008. Especially the sharp increase in the real price of oil in 2007/08 has been attributed to speculation. The standard interpretation is that oil traders in spot markets buy crude oil now in anticipation of higher oil prices with the expectation of selling later at a profit. Unless the oil demand elasticity is zero – a possibility that we discuss below – a necessary implication is that all else equal oil inventories must increase. The response estimates in Figure 1 confirm this prediction. A positive shock to speculative demand increases both the real price of oil and oil inventories.

Notwithstanding the popular perception that speculative demand helped cause the run-up in the real price of oil after 2003 and in particular in 2007/08 – either in anticipation of stronger economic growth or in anticipation of declining oil supplies as predicted by the peak oil hypothesis – there is no indication in Figure 2 that this oil price surge had anything to do with speculative demand shocks. Instead, the model supports the substantive conclusions in Kilian (2009a) and Kilian and Hicks (2009) – based on alternative methodologies – that the surge in the real price of oil between 2003 and mid-2008 was mainly caused by shifts in the flow demand for crude oil driven by the global business cycle. This finding is important because it tells us that further regulation of these markets would have done nothing to stem the increase in the real price of oil.

An alternative view of speculation is that OPEC in anticipation of even higher oil prices held back its production after 2001, using oil below ground effectively as inventories (see, e.g., Hamilton 2009a, p. 239).\textsuperscript{16} One way of testing this hypothesis is through the lens of our structural model. In the model, OPEC holding back oil production in anticipation of rising oil prices would be classified as a negative oil supply shock. Figure 2 provides no indication that negative flow supply shocks played an important role between 2003 and mid-2008. Nor does the

\textsuperscript{15} We do not include the contribution of the residual inventory shock because that shock makes no large systematic contribution to the evolution of the real price of oil.

\textsuperscript{16} We do not consider oil below the ground as part of oil inventories in this paper. Unlike above-ground oil inventories that can be drawn down at short notice, oil below the ground is inaccessible in the short run and not available for consumption smoothing. Thus, it must be differentiated from inventories in the usual sense.
evidence provide any support for the notion that negative oil supply shocks associated with the peak oil hypothesis explain the surge in the real price of oil after 2003.\textsuperscript{17} Any unanticipated exogenous decline in oil supplies associated with that hypothesis likewise would be captured by an unexpected flow supply reduction, yet there is no indication of such shocks persistently driving up the real price of oil toward the end of the sample. What evidence there is of a small supply-side driven increase in the real price of oil is dwarfed by the price increases associated with flow demand shocks. Moreover, the sharp V-shaped dip in the real price of oil in late 2008 is unambiguously driven by a similar dip in the global real activity measure associated with the global financial crisis. Interestingly, the fall in the real price of oil in late 2008 was amplified by negative speculative demand shocks. This makes sense to the extent that oil traders were anticipating a collapse of future real activity and to the extent that uncertainty about the level of future real activity increased.

A similar phenomenon also occurred following the Asian financial crisis of mid-1997, when speculative demand shocks reinforced the decline in the real price of oil caused by the reduced real activity. The resulting drop in the real price of oil to an all-time low since the 1970s was overcome in 1999-2000 by a combination of coordinated OPEC oil supply cuts, a gradual increase in real activity (often associated with the U.S. productivity boom) and increased speculative demand in anticipation of increased future real activity and/or further oil supply reductions (see Figure 2).

4.4. The Inventory Puzzle of 1990

Although speculative motives played no important role after 2003, there are other oil price shock episodes when they did. One particularly interesting example is the oil price shock associated with Persian Gulf War of 1990/91. In related work, Kilian (2009a) presented evidence based on a model without oil inventories that the 1990 oil price increase was driven mainly by a shift in speculative demand (reflecting the uncertainty about future oil supplies from neighboring Saudi Arabia) rather than the physical reduction in oil supplies associated with the war. As noted by Hamilton (2009a), this result is puzzling upon reflection because oil inventories moved little and, if anything, slightly declined following the invasion of Kuwait, which seems at odds with the impulse response results in Figure 2. This observation prompted Hamilton to reject the hypothesis that shifts in speculative demand were behind the sharp increase in the real price of

\textsuperscript{17} For a critical review of the peak oil hypothesis see Einloth (2009).
oil in mid-1990 and its fall after late 1990. Given the consensus that flow demand did not move sharply in mid-1990, Hamilton suggested that perhaps this price increase must be attributed to oil supply shocks after all. The inventory data, however, seem just as inconsistent with this alternative hypothesis. Inventories declined in August of 1990, but only by one third of a standard deviation of the change in inventories. Given one of the largest unexpected oil supply disruptions in history, one would have expected a much larger decline in oil inventories given the impulse response estimates in Figure 1. In light of this evidence, neither the supply shock explanation nor the speculative demand shock explanation by itself seems compelling.

Our econometric model resolves this inventory puzzle. The explanation is that the invasion of Kuwait in August of 1990 represented two shocks that occurred simultaneously. On the one hand it involved an unexpected flow supply disruption and on the other an unexpected increase in speculative demand. Whereas the flow supply shock caused a decline in oil inventories, increased speculative demand in August caused an increase in oil inventories, with the net effect being a slight decline in oil inventories. At the same time, the observed increase in the real price of oil was caused by both shocks working in the same direction.

The historical decomposition in Figure 3 contrasts the price and inventory movements caused by each shock during 1990/91. It shows that overall two thirds of the price increase in August of 1990 was caused by the oil supply shock and only one third by the speculative demand shock. This result is in sharp contrast to the estimates in Kilian (2009) who found no evidence of oil supply shocks contributing to this increase, illustrating again that the inclusion of inventories in the structural model matters. Figure 3 also highlights that the decline in the real price of oil, starting in late 1990 when the threat of Saudi oil fields being captured by Iraq had been removed by the presence of U.S. troops, was almost entirely caused by a reduction in speculative demand rather than increased oil supplies. The latter observation is consistent with evidence in Kilian (2008a) that it is difficult to reconcile the sharp decline in the real price of oil starting in late 1990 with data on oil production.

It is also noteworthy that according to Figure 2 there was a substantial increase in speculative demand and a substantial increase in oil supplies in the months leading up to the invasion with offsetting effects on the real price of oil. This result is consistent with a sharp increase in oil inventories in the months leading up to the invasion. One interpretation is that the invasion was anticipated by informed oil traders or, more likely, that traders responded to
evidence of increased political tension in the Middle East. Gause (2002) notes a shift in Iraqi foreign policy toward a more aggressive stance in early 1990. This shift was caused in part by economic and financial pressures which Iraq attributed to rising oil production in excess of OPEC quotas and hence falling oil prices, consistent with the evidence in the upper panel of Figure 2.

4.5. What Caused the 1979 and 1980 Oil Price Shocks?
Speculative demand played an even more important role in 1979. The traditional interpretation is that this oil price increase was driven by flow supply disruptions associated with the Iranian Revolution of late 1978 and early 1979. Much of the observed increase in the real price of oil, however, only occurred after Iranian oil production had resumed later in 1979. Barsky and Kilian (2002) therefore attribute the price increase starting in May of 1979 and extending into 1980 in part to increased flow demand for oil and in part to a substantial increase in speculative demand for oil, consistent with anecdotal evidence from oil industry sources and with the perception of a noticeable increase in the risk of an oil supply disruption in the Persian Gulf in 1979.18

This hypothesis is testable in our model. Figure 2 shows that not only was there a dramatic and persistent increase in the real price of oil driven by positive flow demand shocks in 1979 and 1980 (not unlike the persistent price increase after 2003), but that increase was reinforced in 1979 by a sharp increase in speculative demand, exactly as conjectured in Barsky and Kilian (2002). Whereas flow demand pressures on the real price of oil gradually receded starting in 1981, speculative demand pressures remained high until the collapse of OPEC in late 1985. In contrast, there is little evidence of flow supply shocks being responsible for the oil price surge of 1979, consistent with the fact that overall global oil production increased in 1979, reflecting additional oil production outside of Iran. Only in late 1980 and early 1981 is there a moderate spike in the real price of oil driven by flow supply shocks, largely associated with the outbreak of the Iran-Iraq War (see Figure 2).

It is useful to explore the price and inventory dynamics in 1979 in more detail. The

---

18 For example, Terzian (1985, p. 260) writes that in 1979 “spot deals became more and more infrequent. The independent refineries, with no access to direct supply from producers, began to look desperately for oil on the so-called ‘free market’. But from the beginning of November, most of the big oil companies invoked force majeure and reduced their oil deliveries to third parties by 10% to 30%, when they did not cut them off altogether. Everybody was anxious to hang on to as much of their own oil as possible, until the situation had become clearer. The shortage was purely psychological, or ‘precautionary’ as one dealer put it.”
historical decompositions in Figure 4 show that indeed flow supply shocks caused a temporary drop in oil inventories in December of 1978 and January of 1979, but for the next half year positive flow supply shocks increased oil inventories. This result is also consistent with the fact that global oil production starting in April of 1979 exceeded its level prior to the Iranian Revolution. At the same time, after March of 1979, repeated speculative demand shocks caused a persistent accumulation of inventories, while driving up the real price of oil. The inventory accumulation continues into 1980. There is no indication that flow supply shocks had much to do with the oil price surge of 1979.

It was not until September of 1980 when the Iran-Iraq War broke out that the oil market experienced another major disruption of flow supply. This event is once again associated with declining oil inventories initially and subsequently rising inventories driven by unexpected flow supply increases, reflecting in part the growing importance of new non-OPEC oil producers. As Figure 5 shows, the increase in the real price of oil in response to this flow supply shock is larger than the price response to the 1979 shock. There is also evidence of a resurgence of speculative demand in the months following the outbreak of the war, reflected in rising inventories and a higher real price of oil.

4.6. The Collapse of OPEC in 1986

In late 1985, Saudi Arabia decided that it would no longer attempt to prop up the price of oil by reducing its oil production, creating a major positive shock to the flow supply of oil. The same event also markedly changed market perceptions about OPEC’s market power. Figure 6 shows that, as expected, the positive flow supply shock in early 1986 drove down the real price of oil, while oil inventories rose in response. Simultaneously, a drop in speculative demand reinforced the decline in the real price of oil, while lowering inventory holdings. Although OPEC attempted to reunite and control production in 1987, amidst increased speculation, as shown in Figure 2, these attempts proved unsuccessful in the long run.

4.7. The Venezuelan Crisis and Iraq War of 2002/03

Figure 7 focuses on the oil supply shock of 2002/2003 when within months first Venezuelan oil production slowed considerably at the end of 2002 and then Iraqi oil production ceased altogether in early 2003. The combined cutback in oil production was of similar magnitude to the oil supply shocks of the 1970s (see Kilian 2008a). Figure 7 shows that this event reflected a
combination of flow supply shocks and speculative demand shocks.

The Venezuelan oil supply crisis of late 2002 was associated with declining oil inventories, consistent with an unexpected oil supply disruption, but it also was associated with an increase in speculative demand in anticipation of the Iraq War that dampened the decline in inventories, while reinforcing the increase in the real price of oil. The military conflict in Iraq lasted from late March 2003 until the end of April 2003. Despite the additional loss of Iraqi output in early 2003, global oil production unexpectedly increased. The production shortfalls in Iraq and Venezuela were more than offset at the global level by increased oil production elsewhere. These positive flow supply shocks lowered the real price of oil starting in early 2003 and resulted in positive inventory accumulation. At the same time, as early as March 2003, lower speculative demand caused the real price of oil to drop and oil inventories to fall. Again the effect of the two shocks on inventories was offsetting, whereas the effect on the price worked in the same direction. This last example again underscores that geopolitical events in the Middle East matter not merely because of the disruptions of the flow supply of oil they may create, but also because of their effect on speculative demand.

5. Can Speculation Occur Without a Change in Oil Inventories?

Our structural model of the oil market not only sheds light on the historical evolution of prices and quantities, but it may be used to obtain direct estimates of the short-run price elasticities of oil supply and oil demand. The elasticity of oil demand in particular plays a central role in assessing the empirical content of recently proposed models of speculation in oil markets. In related work, Hamilton (2009a) shows that speculation could result in a surge in the real price of oil without any oil inventory accumulation, provided the short-run price elasticity of gasoline demand is zero. As shown below, that elasticity is closely related to the short-run price elasticity of oil demand.

5.1. The Short-Run Oil Demand Elasticity in Production

While there is little doubt that the price elasticity of oil supply is near zero in the short run, the literature does not offer much direct evidence on the price elasticity of oil demand.\(^{19}\) This elasticity can only be estimated with the help of a structural model such as model (1). The identification of this parameter requires an exogenous shift of the contemporaneous oil supply

---

\(^{19}\) Indeed our estimate of the price elasticity of oil supply is only about 0.01 or 0.02 on impact, consistent with the conventional view that the short-run oil supply curve is nearly vertical.
curve along the contemporaneous oil demand curve. In contrast, much of the existing literature on estimating oil demand elasticities has been based on models that do not distinguish between oil demand and oil supply shocks (see, e.g., Dahl 1993; Cooper 2003).  

The price elasticity of oil demand can be expressed as the percent change in quantity divided by the percent change in price, as one moves along the oil demand curve. The existing literature on estimating this elasticity has traditionally measured the change in quantity based on the percent change in the production of crude oil. In this paper, we refer to the resulting elasticity as an oil demand elasticity in production, denoted by $\eta^{O,\text{Production}}$. In model (1), this elasticity corresponds to the ratio of the impact response of oil production in Figure 2 relative to the impact response of the real price of oil, given an oil supply shock.

There is a consensus in the literature that the short-run price elasticity of oil demand in production is very low. For example, Hamilton (2009b) conjectures that this elasticity is -0.06. Our posterior median estimate of the oil demand elasticity estimate, as shown in the first column of Table 3, is -0.41. This estimate is seven times higher than typical conjectures in the recent literature. It is also much higher than conventional regression estimates of this elasticity. For example, Dahl (1993) and Cooper (2003) report estimates between -0.05 and -0.07.

The reason for this difference is that standard econometric estimates of the crude oil demand elasticity fail to account for the endogeneity of the price of crude oil. Standard concerns about price endogeneity with respect to quantity suggest that these elasticity estimates are biased toward zero. For example, if we employed the conventional log-level specification used in the earlier literature on our data the implied elasticity estimate would be only -0.02. The use of a fully structural econometric model allows us to overcome this bias problem. 21 While the precise value of the oil demand elasticity estimate may be sensitive to the number of rotations and the choice of seed, accounting for estimation uncertainty does not overturn this result. Table 3 shows the posterior median of the oil demand elasticity estimate along with the 68% error band. The model assigns substantial probability mass to values between -0.59 and -0.26. The posterior

---

20 A recent exception is Baumeister and Peersman (2009) who independently proposed a structural VAR model of the oil market to estimate the oil demand elasticity.

21 Baumeister and Peersman (2009) using a time-varying structural VAR model obtain a median estimate of the short-run oil demand elasticity in production of -0.15 using quarterly data. Their model differs from ours in several dimensions including the choice of the data, data frequency, and sample period, in the variables included, the dynamic specification and the identifying assumptions. Their analysis abstracts from oil inventories. Nevertheless, their elasticity estimate like ours is much higher than conventional estimates from non-structural models.
standard error is 0.15.

5.2. The Short-Run Oil Demand Elasticity in Use

The oil demand elasticity in production in Table 3 is based on the conventional definition in empirical work. The definition of the price elasticity of oil demand that matters for policy questions typically is the oil demand elasticity in use, however. The latter elasticity is based on the change in quantity from both oil production and the depletion of oil inventories and hence more accurately captures the behavior of oil consumers. To our knowledge, this distinction has not been discussed in the literature nor has there been any attempt to estimate the elasticity in use. Doing so requires a structural model that explicitly includes oil inventories. In this section we show how the oil demand elasticity in use can be approximated with the help of our structural dynamic model of the oil market. By construction, allowing for inventory responses will tend to lower the price elasticity of oil demand. An important question is how much of a difference the inclusion of oil inventories makes for the elasticity estimate.

The amount of oil used in period $t$, denoted by $U_t$, equals the oil produced in that period ($P_t$) minus the oil that is added to the stock of inventories ($\Delta S_t$):

$$U_t = P_t - \Delta S_t.$$  

The change in oil used over time therefore equals the change in oil produced minus the change in the addition to inventory stocks: $\Delta U_t = \Delta P_t - \Delta^2 S_t$. The price elasticity of oil demand in use is defined as:

$$\eta_{t}^{Use} = \frac{\%\Delta U_t}{\%\Delta Price_t} = \frac{\Delta P_t - \Delta^2 S_t}{\frac{P_{t-1} - \Delta S_{t-1}}{\%\Delta Price_t}},$$

where $\Delta$ represents changes and $\%\Delta$ indicates percent changes in response to an oil supply shock in period $t$. Denote by $\hat{B}_{11}$ the impact response of the percent change in oil production to an oil supply shock, where $\hat{B}_{ij}$ refers to the $ij$th element of $\hat{B}$. Then the implied change in oil production is $\Delta P_t = P_t - P_{t-1} = P_{t-1} \times \hat{B}_{11} / 100 - P_{t-1} = P_{t-1} \times \hat{B}_{11} / 100$. Moreover, 

$$\Delta^2 S_t = \Delta S_t - \Delta S_{t-1} = \Delta S + \hat{B}_{4t} - \Delta S = \hat{B}_{4t},$$

where the change in oil inventories in response to the oil supply shock equals the impact response $\hat{B}_{41}$ and, prior to the shock, the change in crude oil
inventories is equal to its mean $\overline{\Delta S}$, which is observable. Finally, the impact percent change in
the real price of oil in response to an oil supply shock is $\hat{B}_{31}$. Hence, the demand elasticity in use
can be expressed equivalently as

$$\eta_{t}^{\text{Use}} = \frac{(P_{t-1} \times \hat{B}_{11} / 100) - \hat{B}_{41}}{P_{t-1} - \Delta S - \hat{B}_{31} / 100}.$$  

Note that by construction $\eta_{t}^{\text{Use}}$ depends on $P_{t-1}$ and hence will be time-varying even though the
oil demand elasticity in production is not. We therefore report the average oil demand elasticity
in use over the sample period throughout this paper, denoted by $\eta^{O, \text{Use}}$. This allows for the
possibility that the estimate of $\eta_{t}^{\text{Use}}$ may be slightly positive merely by chance, which seems
especially likely if $\eta_{t}^{\text{Use}}$ is close to zero in population because this elasticity depends on reduced-
form parameters that may be imprecisely estimated.

The second column of Table 3 shows that, as expected, the demand elasticity in use is
much lower than the elasticity in production. The median estimate is only -0.20 compared with
an estimate of -0.41 for the demand elasticity in oil production. The 68% error bands range from
-0.38 to -0.09. This finding is important in that it suggests that even the inclusion of inventories
does not overturn our findings that the short-run price elasticity of oil demand is much higher
than previously thought.

5.3. Bounding the Short-Run Gasoline Demand Elasticity

The magnitude of the short-run price elasticity of oil demand has important implications for
theoretical models of speculative demand. In the absence of changes to oil production, standard
models of speculation imply that oil inventories must increase to enable the price of oil to
increase. Recently, it has been suggested that under certain conditions speculation may drive up
the real price of oil without any change in oil inventories (see Hamilton 2009a). Specifically, this
may occur if refiners are able to pass on fully to gasoline consumers an exogenous increase in
the real price of oil driven by speculation. This requires the demand for gasoline to be
completely price-inelastic. Whether the alternative model of speculation in Hamilton (2009a)
could explain our data depends on the magnitude of the short-run price elasticity of gasoline
demand.
That magnitude is directly related to the impact price elasticity of oil demand in use. Rather than extend the econometric model to permit a joint analysis of the crude oil and gasoline markets, which does not seem feasible given the limited degrees of freedom, in this section we derive an explicit relationship between consumers’ demand for gasoline and refiners’ demand for crude oil in a model in which refiners are allowed to, but not required to have market power in the gasoline market. Refiners are treated as price-takers in the crude oil market. Our analysis is strictly short-term, as is appropriate in constructing impact price elasticities. In the interest of tractability, we abstract from the fact that gasoline is only one of several refined products jointly produced from crude oil.

We postulate that gasoline is produced according to a Leontief production function over capital, labor, and oil, \( G = \min (K, L, \alpha O) \). If capital is fixed in the short run and refiners’ labor input can be varied on the intensive margin, which seems plausible in practice, refiners produce gasoline in fixed proportion to the quantity of oil consumed, \( G = \alpha O \), and pay a marginal cost equal to the price of oil, \( P_o \), plus the marginal cost of labor, \( MC = P_o + c \).

Consumers maximize

\[
U(B, G) = \left[ \gamma^{\sigma} B^{(\sigma - 1)/\sigma} + (1 - \gamma)^{\sigma} G^{(\sigma - 1)/\sigma} \right]^{\sigma/(\sigma - 1)},
\]

where \( B \) is a bundle of non-gasoline consumer goods and \( G \) is gasoline consumption at price \( P_g \). The term \( \sigma \) is the constant elasticity of substitution and \( 1 - \gamma \) denotes the expenditure share of gasoline. Consumers choose \( G(P_g) = X(1 - \gamma)P_o^{\sigma} / P^{1 - \sigma} \), where \( X \) is the total expenditure on all goods, \( P \) is the consumer price index, and the price elasticity of demand for gasoline, \( \eta^G \), equals \(-\sigma\). Equivalently, \( P_g(G) = \omega G^{-1/\sigma} \) where \( \omega = (1 - \gamma)X / P^{1-\sigma} \).

In the Cournot-Nash equilibrium, each of \( J \) identical refinery firms will choose its own quantity of gasoline output, \( g_j \), \( j = 1, ..., J \), given the outputs of other firms, to maximize profits

\[
\pi_j = P_g(G) g_j - (c + P_o) g_j - f = \omega G^{-1/\sigma} g_j - (c + P_o) g_j - f
\]

with respect to \( g_j \), where \( G = \sum_j g_j \) and \( f \) represents short-run fixed costs. This yields the first order conditions

\[
P_g + g_j P_g' = \omega G^{-1/\sigma} - \omega g_j G^{(\sigma - 1)/\sigma} / \sigma - (c + P_o) = 0 \quad j = 1, ..., J.
\]
Summing over $j$ and solving for the market price and gasoline production yields

$$P_G = \frac{J\sigma (P_o + c)}{J\sigma - 1} \quad G = \left( \frac{\omega(J\sigma - 1)}{J\sigma(P_o + c)} \right)^\sigma.$$  

Given $G = \alpha O$, we obtain

$$\alpha O = \left( \frac{\omega(J\sigma - 1)}{J\sigma(P_o + c)} \right)^\sigma.$$  

Log-linearization yields

$$\eta^{O,\text{use}} \approx \frac{P_o}{P_o + c} \eta^G,$$

where $\eta^{O,\text{use}}$ denotes the price elasticity of demand for crude oil in use. Thus, if $\eta^{O,\text{use}}$ were zero, then by construction so would be $\eta^G$.

Given the evidence in section 5 that $\eta^{O,\text{use}}$ is not even close to zero, we need to be more specific about the link. If the price of oil constitutes the largest component of refiners’ marginal cost, consistent with the evidence presented in Considine (1997), then $0 \leq c \leq P_o$, which in turn implies $\eta^{O,\text{use}} \leq \eta^G \leq 2\eta^{O,\text{use}}$. Thus, the price elasticity of gasoline demand must be between one and two times as high as the price elasticity of the demand for crude oil. This result includes as a limiting case the view expressed in Hamilton (2009a) that the gasoline demand elasticity is approximately twice as high as the oil demand elasticity. Given the marginal cost estimates presented in Considine (1997), the factor of proportionality is likely to be close to unity.

This allows us to conclude based on Table 3 that the median estimate of the short-run price elasticity of gasoline demand must be about -0.20.\footnote{Based on the 68% error band, the price elasticity of gasoline demand could be anywhere between -0.09 and -0.38.} That estimate is much larger than some recent estimates in the literature. For example, Hughes, Knittel and Sperling (2008) in an influential recent study report a baseline elasticity estimate of only -0.04 for the United States based on data for 2001-06. Their estimate, however, like earlier studies, fails to account for price endogeneity and cannot be interpreted as an elasticity in the textbook sense.\footnote{The baseline results in Hughes et al. (2009) are ordinary least squares estimates of monthly reduced-form time series regressions of quantity on price in log-levels (with additional controls included). As part of their sensitivity analysis for 2001-06, they also use as an instrument unpredictable changes in global crude oil production driven by}
sensitive to the sample period. Similar estimates of this elasticity based on older data are near -0.20 (see, e.g., Dahl and Sterner 1991). The average of the least-squares estimates in Hughes et al. for 1975-80 and for 2001-06 is -0.19.24

Our bounds on the impact price elasticity of gasoline demand allows us to discount the possibility discussed in Hamilton (2009a) that speculation may arise without any change in inventories. This finding is also consistent with the large positive inventory response to speculative demand shocks in Figure 1. It is important to stress that our identifying assumptions encompass the model discussed in Hamilton (2009a) as a special case. Hence, we can unambiguously reject the hypothesis that increases in the real price of oil in 2003-08, in particular, were driven by speculation.

6. Related Estimates of the Price Elasticity of Gasoline Demand in the Literature

Although there is no alternative to the use of dynamic simultaneous equations models in estimating the short-run price elasticity of oil demand or gasoline demand, it is useful to compare our findings to the related literature on estimating long-run price elasticities of gasoline demand. There is widespread recognition that the long-run adjustments to changes in energy prices are potentially much greater than the adjustments occurring within the first month. It is commonly thought that the adjustment toward more energy-efficient technologies occurs over a horizon of perhaps five or ten years (see, e.g., Sweeney 1984).

Such long-run price elasticities may be estimated from cross-sectional household data.25 The premise of cross-sectional estimates of the price elasticity of gasoline demand is that, controlling for region-specific and time effects (as well as other relevant controls such as demographics or income), consumers on average are in steady state. Identification in these regressions arises from the within-region variation in U.S. gasoline prices. Thus, an obvious events such as Hurricanes Rita and Katrina. This results in gasoline demand elasticity estimates as high as -0.08 for 2001-2006. However, the weak instrument diagnostics in Hughes et al. are not informative, raising concerns about downward bias in the elasticity estimates.

24 Indeed, one of the main points in Hughes et al. (2009) is that there appears to have been a structural change with recent estimates of the price elasticity of gasoline demand being lower. This interpretation overlooks that their model makes no allowance for changes in the composition of demand and supply shocks in the gasoline market. Recent work by Kilian (2010) has demonstrated that ignoring the distinction between shifts in the contemporaneous oil supply curve and shifts in the contemporaneous demand curve for oil can account for the apparent instability in the expected response of gasoline consumption to gasoline price shocks. In the model proposed in section 3, we similarly control for this source of time variation.

25 This is not an issue that can be addressed with our VAR methodology. That methodology is not suitable for assessing responses at horizons of five or ten years. Given the short time series available, estimates of dynamic responses from structural VAR models tend to become unreliable at horizons beyond one or two years.
plausibility check is that estimates of the short-run price elasticity of gasoline demand from structural dynamic models must not exceed the corresponding long-run estimates form structural cross-sectional models. It can be shown that our short-run estimates are fully consistent with the cross-sectional evidence. For example, studies of nonparametric gasoline demand functions based on U.S. household survey data such as Hausman and Newey (1995) have consistently produced long-run price elasticity estimates near -0.8, far in excess of our median estimate of -0.20 of the short-run elasticity.26

7. Conclusion

Standard structural VAR models of the market for crude oil implicitly equate oil production with oil consumption and ignore the role of oil inventories. Traditionally these models have focused on shocks to the contemporaneous flow supply of oil and the contemporaneous flow demand for oil. In this paper we augmented the structural model to include shocks to expectations about future oil supplies and future oil demand. Such shocks must be represented as shifts of the contemporaneous oil demand curve rather than the contemporaneous oil supply curve, even if the shift in expectations is about a cut in future oil supplies. The reason is that traders in anticipation of the expected oil shortage will buy and store crude oil now with the expectation of selling later at a profit. Such speculative demand shocks will typically be associated with inventory building. In contrast, production smoothing in response to positive flow demand shocks and negative flow supply shocks will involve a drawing down of oil inventories. We proposed a dynamic simultaneous equations model including oil inventories that allows the identification of all three types of shocks.

The inclusion of oil inventories matters. It not only resolves some anomalies in earlier impulse response estimates, but the structural model proposed in this paper implies a larger role for oil supply shocks in explaining fluctuations in the real price of oil. The added explanatory power of oil supply shocks in explaining fluctuations in the real price of oil, especially in 1990, comes at the expense of the explanatory power of speculative demand shocks. Nevertheless, the largest and most persistent fluctuations in the real price of oil since the 1970s such as the oil price increase in 1979/80 were driven primarily business cycle fluctuations affecting the demand

26 Schmalensee and Stoker (1999) question the reliability of the price data and the specification used in Hausman and Newey (1995), but Yatchew and No (2001) address those concerns using more detailed Canadian data and arrive at a gasoline demand elasticity estimate of -0.9, very close to Hausman and Newey’s original estimate.
for crude oil. Of particular interest in this paper has been the oil price increase from 2003 until mid-2008. We were able to provide direct evidence against the popular view that the sharp increase in the real price of oil during this period (and in 2007/08 in particular) was driven by speculation among oil traders. Shifts in speculative demand played a more important role during several earlier oil price shock episodes, however, notably in 1979, 1986 and in 1999/2000. We showed that, without accounting for shifts in the speculative demand for oil, it is not possible to understand the evolution of the real price of oil during these episodes.

Our analysis also suggests that there is no evidence that the peak oil hypothesis or deliberate production cutbacks by oil producers had much bearing on the recent oil price shock. Rather our results support recent findings in the literature that the run-up in the real price of oil between 2003 and mid-2008 was caused almost entirely by shifts in the global flow demand for oil. This implies that the real price of oil is expected to rise, as the global economy recovers from the financial crisis, creating a policy dilemma, unless energy consumption can be reduced or new energy sources can be found. In contrast, additional regulation of oil traders is unlikely to prevent the price of oil from rising again in the future.

Hamilton (2009a,b) recently has cast doubt on explanations of major oil price increases based on shifts in speculative demand during previous oil price shocks episodes. He observed in particular that following the outbreak of the Persian Gulf War in August 1990, oil inventories did not increase as one would have expected in response to a positive speculative demand shock. At the same time, the absence of a sharp decline in oil inventories in August of 1990 is inconsistent with the view that the price increase reflected a negative oil supply shock. We demonstrated that this inventory puzzle can be resolved with the help of a structural oil market model. Our analysis showed that the price and inventory data can be explained only based on a combination of these two shocks. Because the implied inventory responses are of opposite sign, the net effect in inventories is close to zero, where the sharp price increase reflects the fact that the implied price responses are of the same sign. Similar relationships were shown to hold during other key historical episodes. These examples illustrate that it is essential to rely on structural models in interpreting the price and quantity data.

The use of a structural regression model also is important for the construction of the short-run price elasticity of oil demand. For example, Hamilton (2009a,b) suggests that 1978-81 is one episode where one might clearly and without a regression model attribute cumulative
changes in the price of oil to exogenous oil supply shifts only, allowing one to construct a
demand elasticity estimate from the ratio of cumulative changes in quantities and prices for that
period. The structural model we have analyzed suggests otherwise. We showed that oil demand
shocks were the main cause of the observed oil price increase in 1978-81. Oil supply shocks
played a small role only, violating the premise of Hamilton’s calculations.

We observed that there are no credible regression-based estimates of the short-run price
elasticity of oil demand in the literature. Conventional estimates of this elasticity from dynamic
regressions, as in Dahl (1993) and Cooper (2003), in particular, not only ignore the endogeneity
of the real price of oil, causing the elasticity estimate to be downward biased, but they also
ignore the use of inventories in smoothing oil consumption in response to oil supply shocks. We
showed how both the traditional oil demand elasticity in production and the more relevant oil
demand elasticity in use can be estimated from our structural model. Our elasticity estimates are
substantially higher than existing estimates, helping us to reject models of speculative trading
based on a zero short-run price elasticity of oil demand. We also provided a theoretical
benchmark for the relationship between the short-run price elasticity of oil demand in use and the
corresponding price elasticity of gasoline demand. The latter was shown to be about as high as
the oil demand elasticity in use under plausible assumptions about the refining industry. The
resulting short-run gasoline demand elasticity estimates are compatible with alternative estimates
of the long-run price elasticity of gasoline demand from cross-sectional data.
References


Table 1: Sign Restrictions on Impact Responses in VAR Model

<table>
<thead>
<tr>
<th></th>
<th>Flow supply shock</th>
<th>Flow demand shock</th>
<th>Speculative demand shock</th>
</tr>
</thead>
<tbody>
<tr>
<td>Oil production</td>
<td>-</td>
<td>+</td>
<td>+</td>
</tr>
<tr>
<td>Real activity</td>
<td>-</td>
<td>+</td>
<td>-</td>
</tr>
<tr>
<td>Real price of oil</td>
<td>+</td>
<td>+</td>
<td>+</td>
</tr>
<tr>
<td>Inventories</td>
<td>-</td>
<td>-</td>
<td>+</td>
</tr>
</tbody>
</table>

Note: All sign restrictions involve weak inequalities. All shocks have been normalized to imply an increase in the real price of oil.

Table 2: Percent Contribution to Variability of the Change in Inventories

<table>
<thead>
<tr>
<th>Horizon</th>
<th>Flow supply shock</th>
<th>Flow demand shock</th>
<th>Speculative demand shock</th>
</tr>
</thead>
<tbody>
<tr>
<td>3</td>
<td>26.5</td>
<td>1.2</td>
<td>45.8</td>
</tr>
<tr>
<td>6</td>
<td>26.1</td>
<td>1.9</td>
<td>45.6</td>
</tr>
<tr>
<td>9</td>
<td>26.0</td>
<td>3.1</td>
<td>45.4</td>
</tr>
<tr>
<td>12</td>
<td>25.6</td>
<td>3.6</td>
<td>45.6</td>
</tr>
<tr>
<td>∞</td>
<td>23.4</td>
<td>14.5</td>
<td>39.1</td>
</tr>
</tbody>
</table>

Note: Results from a structural forecast error variance decomposition based on the benchmark estimate of the structural model (1) as in Figure 1. The infinite horizon is approximated by 600.

Table 3: Posterior Distribution of the Short-Run Price Elasticity of Demand for Crude Oil

\[
\eta^{O, \text{Production}} \quad \eta^{O, \text{Use}}
\]

<table>
<thead>
<tr>
<th>Percentile</th>
<th>(\eta^{O, \text{Production}})</th>
<th>(\eta^{O, \text{Use}})</th>
</tr>
</thead>
<tbody>
<tr>
<td>16th</td>
<td>-0.59</td>
<td>-0.38</td>
</tr>
<tr>
<td>50th</td>
<td>-0.41</td>
<td>-0.20</td>
</tr>
<tr>
<td>84th</td>
<td>-0.26</td>
<td>-0.09</td>
</tr>
<tr>
<td>Standard Deviation</td>
<td>0.15</td>
<td>0.13</td>
</tr>
</tbody>
</table>

Note: Based on 58 admissible structural models drawn from the posterior. In constructing the posterior we impose a uniform prior on the elasticity in use that bounds this elasticity at -0.5. This restriction is not binding in estimates based on actual data. \(\eta^{O, \text{Production}}\) refers to the price elasticity of oil demand in production and \(\eta^{O, \text{Use}}\) to the price elasticity of oil demand in use. The latter definition accounts for the role of inventories in smoothing oil consumption.
Figure 1: Structural Impulse Responses
1973.2-2009.8

Note: Results for the admissible structural model with an impact price elasticity of oil demand in use closest to the posterior median of that elasticity. Dashed lines indicate pointwise 16% and 84% posterior quantiles based on 58 admissible draws from the posterior.
Figure 2: Historical Decompositions for 1978.6-2009.8

Cumulative Effect of Flow Supply Shock on Real Price of Crude Oil

Cumulative Effect of Flow Demand Shock on Real Price of Crude Oil

Cumulative Effect of Speculative Demand Shock on Real Price of Crude Oil

Note: Based on benchmark estimate as in Figure 1. The vertical bars indicate major exogenous events in oil markets, notably the outbreak of the Iranian Revolution in 1978.9 and of the Iran-Iraq War in 1980.9, the collapse of OPEC in 1985.12, the outbreak of the Persian Gulf War in 1990.8, the Asian Financial Crisis of 1997.7, and the Venezuelan crisis in 2002.11, which was followed by the Iraq War in early 2003. In constructing the historical decomposition we discard the first five years of data in an effort to remove the transition dynamics.
Figure 3: Historical Decompositions for the Persian Gulf War Episode of 1990/91

Note: Based on estimates of structural model (1) on data for 1973.2-2009.8
Figure 4: Historical Decompositions for the Iranian Revolution of 1978/79

Real price of oil

Change in oil inventories

Note: Based on estimates of structural model (1) on data for 1973.2-2009.8
Figure 5: Historical Decompositions for the Outbreak of the Iran-Iraq War in 1980

Real price of oil

- Cumulative effect of flow supply shock
- Cumulative effect of speculative demand shock

Change in oil inventories

Note: Based on estimates of structural model (1) on data for 1973.2-2009.8
Figure 6: Historical Decompositions for the Collapse of OPEC in 1986

Real price of oil

- Cumulative effect of flow supply shock
- Cumulative effect of speculative demand shock

Change in oil inventories

Note: Based on estimates of structural model (1) on data for 1973.2-2009.8
Figure 7: Historical Decompositions for Venezuelan Crisis and Iraq War in 2002/03

Note: Based on estimates of structural model (1) on data for 1973.2-2009.8