

THE ROLE OF INVENTORIES AND SPECULATIVE TRADING IN THE GLOBAL MARKET FOR CRUDE OIL

LUTZ KILIAN^{a,b,*} AND DANIEL P. MURPHY^a

^a *Department of Economics, University of Michigan, Ann Arbor, MI, USA*

^b *CEPR, London, UK*

SUMMARY

We develop a structural model of the global market for crude oil that for the first time explicitly allows for shocks to the speculative demand for oil as well as shocks to flow demand and flow supply. The speculative component of the real price of oil is identified with the help of data on oil inventories. Our estimates rule out explanations of the 2003–2008 oil price surge based on unexpectedly diminishing oil supplies and based on speculative trading. Instead, this surge was caused by unexpected increases in world oil consumption 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, 1986 and 1990. Our analysis implies that additional regulation of oil markets would not have prevented the 2003–2008 oil price surge. We also show that, even after accounting for the role of inventories in smoothing oil consumption, our estimate of the short-run price elasticity of oil demand is much higher than traditional estimates from dynamic models that do not account for the endogeneity of the price of oil. Copyright © 2013 John Wiley & Sons, Ltd.

Received 3 April 2012; Revised 18 December 2012



Supporting information may be found in the online version of this article.

1. INTRODUCTION

There is no consensus in the academic literature on how to model the global market for crude oil. One strand of the literature suggests that the price of oil 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. Another strand of the literature views the price of oil as being determined by shocks to the amount of oil coming out of the ground ('flow supply of oil') and the amount of oil being consumed ('flow demand for oil') with little attention to the role of inventories. Much of the early literature on oil supply shocks is in that tradition, as are more recent economic models linking the real price of oil to fluctuations in the global business cycle.¹ Recently, there has been increasing recognition that both stock demand and flow demand for oil matter in modeling the real price of oil (see, for example, Hamilton, 2009a,b; Kilian, 2009a; Alquist and Kilian, 2010; Dvir and Rogoff, 2010). In this paper, we propose a structural vector autoregressive (VAR) model of the global market for crude oil that explicitly nests these two explanations of the determination of the real price of oil and allows us to quantify the effects of different oil demand and supply shocks. Drawing on insights from the economic theory for storable commodities, we design a set of identifying assumptions that allows us to estimate jointly the expectations-driven component of the real price of oil and the components driven by flow demand and flow supply.

Constructing such an econometric model is nontrivial because the potential presence of a forward-looking component in the real price of oil considerably complicates the identification of the structural

* Correspondence to: Lutz Kilian, Department of Economics, University of Michigan, 611 Tappan Street, Ann Arbor, MI 48109-1220, USA. E-mail: lkilian@umich.edu

¹ See, for example, Baumeister and Peersman (2013); Hamilton (2009a,b); Kilian (2008a,b, 2009a,b); Kilian and Murphy (2012).

shocks. Traditional oil market VAR models implicitly equate market expectations with the econometric expectations formed on the basis of past data on oil production, global real activity, and the real price of oil. If traders respond to information about future demand and supply conditions not contained in the past data available to the econometrician, however, market expectations will differ from the expectations constructed by the econometrician, rendering traditional models of flow supply and flow demand invalid. The desirability of holding oil stocks may change, for example, in response to news about oil discoveries, or as traders anticipate wars in the Middle East, or as traders respond to increased uncertainty about future oil supply shortfalls. None of these expectations shifts can be captured using standard models of flow demand and flow supply. In Sections 2 and 3, we show that this problem can be overcome with the help of data on above-ground crude oil inventories. The intuition is that – unless the price elasticity of oil demand is zero – any expectation of a shortfall of future oil supply relative to future oil demand not already captured by flow demand and flow supply shocks necessarily causes an increase in the demand for above-ground oil inventories and hence in the real price of oil (see, for example, Hamilton, 2009a; Alquist and Kilian, 2010). We refer to such a shock as a speculative demand shock in the spot market for crude oil. It is this type of shock that many researchers and policymakers explicitly or implicitly appeal to when attributing higher spot prices to speculation. We stress that such speculative demand shocks cannot be inferred directly from observables and can only be identified within the context of a structural econometric model.

Our definition of speculation in this paper is general in that we treat anyone buying crude oil not for current consumption but for future use as a speculator from an economic point of view. Speculative purchases of oil usually occur because the buyer is anticipating rising oil prices. This anticipation may arise because of changes in expected fundamentals, for example, or because the buyer is anticipating other market participants' actions. Speculative purchases may also be precautionary in that they reflect increased uncertainty about future demand or supply conditions (see Alquist and Kilian, 2010).

We do not take a stand on whether such speculative behavior is desirable from a social point of view. In particular, we do not attempt to distinguish between normal and excessive levels of speculation, nor do we define speculation on the basis of who is trading or what positions these traders take. As discussed in Fattouh *et al.* (2013), there is no operational definition of excessive speculation in the literature. Rather, we will show in Section 4 that all speculative trading in the spot market for crude oil combined had little effect on the real price of oil between 2003 and mid 2008, making the distinction between normal and excessive levels of speculation moot, at least for this episode.

Our analysis allows new insights into the genesis of historical oil price fluctuations. It is of particular relevance for recent policy discussions about the potential role of speculation in oil markets after 2003. First, as discussed in Section 4, our estimates rule out speculation as a cause of the surge in the real price of oil between 2003 and mid 2008. Furthermore, under the maintained assumption of arbitrage between spot and futures markets, the absence of speculative pressures in the spot market implies that there cannot have been speculative pressures in the oil futures market either.² Instead, our model implies that both spot and futures prices during 2003–2008 were driven by unexpected increases in world oil consumption. From this result we infer that additional regulation of oil futures markets would not have prevented the increase in the real price of oil in the physical market for oil, which is the ultimate concern of policymakers.³

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

² This implication is fully consistent with other, more direct evidence on the potential impact of financial investors on oil futures prices (see Fattouh *et al.*, 2013).

³ It is important to note that our conclusion regarding the determinants of the spot price of oil would remain equally valid if there were limits to arbitrage between spot and futures markets. Limits to arbitrage, of course, would undermine the argument that financial speculation is driving the spot price.

a central feature of the oil price surge 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 played a central role in 1990, following Iraq's invasion of Kuwait. Although neither negative flow supply shocks nor positive speculative demand shocks alone can explain the oil price spike and oil inventory behavior of 1990/91, their combined effects do.

Third, our analysis sheds new light on the evolution of the real price of oil since 1978. We document that unexpected fluctuations in global real activity explain most of the surge in the real price of oil between 2003 and mid 2008, even acknowledging that negative flow supply shocks also 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 the second half of 1979. In contrast, flow supply shocks played only a minor role in 1979. The continued rise in the real price of oil in 1980 reflected negative flow supply shocks (caused in part 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 post-war history, was aided by coordinated 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 speculation in oil markets rests on the fact that there has been no noticeable increase in the rate of inventory accumulation in recent years. Our model suggests that even after controlling for the effect of other shocks on crude oil inventories there is no evidence of rising speculative demand after 2003. Recently, Hamilton (2009a) pointed out that, as a matter of theory, speculative trading in oil futures markets may cause a surge in the real price of oil even without any change in oil inventory holdings if the short-run price elasticity of demand for gasoline is literally zero. Thus it is essential that we pin down the value of this elasticity. We provide a theoretical model that shows that under reasonable assumptions about the oil refining industry the short-run price elasticity of gasoline demand is about as high as the short-run price elasticity of oil demand. Hamilton (2009a) observed that existing estimates of the latter elasticity are so close to zero that the possibility of an elasticity of zero deserves further examination. These existing elasticity 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 concern with the help of our structural VAR model. Not only does this model allow the construction of a direct estimate of the elasticity parameter based on exogenous shifts of the oil supply curve along the oil demand curve, but it also incorporates for the first time changes in oil inventories in computing the price elasticity of oil demand. Our posterior median estimate of the short-run price elasticity of oil demand of -0.26 is four times higher than standard estimates in the literature and there is little probability mass on values close to zero.⁴ Thus the limiting case of a zero price elasticity of demand discussed by Hamilton (2009a) is an unlikely explanation of the 2003–2008 surge in the real price of oil. The concluding remarks are in Section 6.

2. VAR METHODOLOGY

Our analysis is based on a dynamic simultaneous equation 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, the real price of crude oil and the change in global crude oil inventories

⁴ Even higher oil demand elasticity estimates have been obtained independently by other recent studies employing structural estimation methods. None of these studies accounts for changes in inventories, however.

above the ground.⁵ All data are monthly. The sample period is 1973:2–2009:8. We remove seasonal variation by including seasonal dummies in the VAR model.

2.1. Data

Our measure of fluctuations in global real activity is the dry cargo shipping rate index developed in Kilian (2009a). This real activity index is a business cycle index and stationary by construction. It is designed to capture shifts in the global use of industrial commodities. 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 capturing shifts in the global demand for industrial commodities. Data on global crude oil production are available in the *Monthly Energy Review* of the Energy Information Administration (EIA). These data also include lease condensates but exclude natural gas plant liquids. Oil production is expressed in percent changes in the model. The real price of oil is defined as the US 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 US 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 US price of domestic crude oil, which was regulated during the 1970s and early 1980s. Following Kilian (2009a), the real price of oil is expressed in log-levels.⁶

Given the lack of data on crude oil inventories for other countries, we follow Hamilton (2009a) in using the data for total US crude oil inventories provided by the EIA. These data are scaled by the ratio of OECD petroleum stocks over US petroleum stocks, also obtained from the EIA. That scale factor ranges from about 2.23 to 2.59 in our sample.⁷ 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 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. Preliminary tests provided no evidence of cointegration between oil production and oil inventories.

⁵ Unlike above-ground oil inventories that can be drawn down at short notice, oil below the ground (also known as reserves) is inaccessible in the short run and not available for consumption smoothing. Thus it must be differentiated from oil inventories in the usual sense. We do not utilize data for reserves because no reliable time series data exist on the quantity of oil below the ground and because reserves data are not required for our identification. We 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.

⁶ It is not clear a priori whether the real price of oil should be modeled in log-levels or log-differences. The level specification adopted in this paper has the advantage that the impulse response estimates are asymptotically valid not only 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.

⁷ 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 US petroleum inventories. For the period 1987:12–2009:8, the US and OECD petroleum inventory growth rates are reasonably close, with a correlation of about 80%.

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 and in modeling business cycles in commodity markets. The corresponding structural model of the global oil market may be written as

$$B_0 y_t = \sum_{i=1}^{24} B_i y_{t-i} + \varepsilon_t \quad (1)$$

where ε_t is the vector of orthogonal structural innovations and B_i , $i=0, \dots, 24$ denotes the coefficient matrices. The seasonal dummies have been suppressed for notational convenience. The vector ε_t consists of four structural shocks. The first shock corresponds to the classical notion of an oil supply shock as discussed in the literature ('flow supply shock'). This shock incorporates supply disruptions associated with exogenous political events in oil-producing countries as well as unexpected politically motivated supply decisions by OPEC members and other flow supply shocks. Second, we include a shock to the demand for crude oil and other industrial commodities that is associated with unexpected fluctuations in the global business cycle ('flow demand shock'). The third shock captures shifts in the demand for above-ground oil inventories arising from forward-looking behavior not otherwise captured by the model ('speculative demand shock').⁸ Finally, we include a residual shock designed to capture idiosyncratic oil demand shocks not otherwise accounted for (such as weather shocks, changes in inventory technology or preferences, or politically motivated releases of the US Strategic Petroleum Reserve).

Each of these shocks has unique characteristics. For example, an unexpected disruption of the flow of oil production (embodied in a shift to the left of the contemporaneous oil supply curve along the oil demand curve, conditional on all past data) raises the real price of crude oil and lowers global real activity within the same month. The impact effect on oil inventories is ambiguous. On the one hand, a negative flow supply shock will cause oil inventories to be drawn down in an effort to smooth consumption. On the other hand, the same shock may raise demand for inventories to the extent that a negative flow supply shock triggers a predictable increase in the real price of oil. Which effect dominates is unclear *ex ante*, so we do not restrict the sign of the impact response of inventories.

In contrast, a positive flow demand shock (embodied in a shift to the right of the contemporaneous oil demand curve along the oil supply curve, conditional on all past data) raises the real price of oil and stimulates global oil production within the same month. As in the case of a negative flow supply shock, the impact effect on inventories is ambiguous *ex ante*.

Given that crude oil is storable, the real price of oil also depends on the demand for oil stocks or inventories. 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 that is not already embodied in flow supply and flow demand shocks. For example, upward revisions to expected future demand 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. Such shifts could arise, for example, because of the anticipation of political unrest in oil-producing countries in the Middle East,

⁸ An alternative and less common view is that speculation may also be conducted by oil producers who choose to leave oil below the ground in anticipation of rising prices. The latter form of 'speculative supply shock' would be associated with a negative flow supply shock in our framework rather than the building of above-ground inventories. Both forms of speculation are permitted by our model, but only speculation that involves the accumulation of inventories above the ground is explicitly identified.

because of the anticipation of peak oil effects, or because of the anticipated depletion of oil reserves. Likewise, traders may anticipate a global recession in the wake of a financial crisis, may anticipate higher future oil production as new deep-sea oil is discovered off the shores of Brazil, or may anticipate the resumption of oil production in Iraq, as the stability of that country improves.

Rather than being associated only with future oil supply conditions or only with future oil demand conditions, speculative demand shocks reflect the expected shortfall of future oil supply relative to future oil demand. A positive speculative demand shock will shift the demand for above-ground oil inventories, causing in equilibrium the level of inventories and the real price of oil to increase on impact. The accumulation of inventories is accomplished by a reduction in oil consumption (reflected in lower global real activity) and an increase in oil production, as the real price increases.⁹ Both flow demand shocks and speculative demand shocks have an expectational component. The feature that distinguishes flow demand shocks from speculative demand shocks is that positive flow demand shocks necessarily involve an increase in the demand for consumption in the current period, whereas speculative demand shocks do not.¹⁰

News about the level of future oil supplies and the level of 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 much the 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. Finally, speculative demand shocks may also arise because traders' perception of what other traders think evolves or simply because of changes in beliefs not related to expected fundamentals. One of the attractive features of the econometric model is that it does not require the user to specify how expectations are formed.

2.3. Why Do We not Include the Oil Futures Spread?

The focus in this paper is on modeling the real price of oil in the physical market for oil. We do not explicitly model the oil futures market. Indeed, conceptually the futures market is distinct from the physical (or spot) market. As discussed in Alquist and Kilian (2010) and Hamilton (2009a,b), there is an arbitrage condition linking the oil futures market and the spot market for crude oil. To the extent that speculation drives up the price in the oil futures market, arbitrage will ensure that oil traders buy inventories in the spot market in response. Thus we can focus on quantifying speculation in the spot market with the help of the oil inventory data without loss of generality.¹¹ In fact, the analysis in Alquist and Kilian implies that data on the oil futures spread are redundant in our structural VAR model, given that we have already included changes in above-ground oil inventories. The fact that the inclusion of oil inventory data makes oil futures prices redundant is particularly advantageous considering that oil futures markets were created only in the 1980s, and thus oil futures prices do not exist for a large part of our sample. Equally importantly, our model remains well specified even if the arbitrage between spot and futures markets were less than perfect at times, whereas a model

⁹ Although oil producers could conceivably react to the same news that triggers a positive speculative demand shock by lowering oil production in anticipation of predictable increases of the real price, there is no evidence that oil producers have responded systematically in this way. Instead, anecdotal evidence suggests that oil producers such as Saudi Arabia have often increased their production levels following positive speculative demand shocks, consistent with the view that the expected impact response should be weakly positive. Our analysis is based on the premise that these shocks are mutually uncorrelated, but allows multiple shocks to occur at the same time.

¹⁰ For a theoretical analysis of flow demand shocks and how their effect on the real price of oil may be amplified by index funds trading in oil futures markets also see Sockin and Xiong (2012).

¹¹ This result breaks down only if demand for oil is completely price inelastic, a case that we discuss in more detail in section 5.

including the oil futures spread would become invalid in that case. Finally, inventory data allow us to impose identifying information about the price elasticity of oil demand, which could not be imposed when using the oil futures spread.

2.4. How Accurate Are the Oil Inventory Data?

While our structural VAR specification has many advantages, it relies on the global crude oil inventory data being accurate. Two concerns regarding the reliability of these data stand out. First, much has been made of media reports that some speculators in 2007/08 leased oil tankers to store oil. Although the EIA does not provide data on the use of tankers for storage, leasing tankers is expensive and the extent to which tankers have been used for storage appears small and limited to the very end of our sample. Second, there has been concern about the expansion of strategic reserves in non-OECD countries such as China. Non-OECD strategic reserves are not included in our inventory dataset. However, most of the new Chinese oil storage facilities had yet to be filled by the end of our sample, so our analysis is not likely to be affected much. Thus there is reason to believe that our inventory data, while less than perfect, are still informative. There are several ways of testing this premise.

First, if there were additional information in the oil futures spread that is not already contained in our inventory proxy, rendering our VAR model informationally misspecified, the oil futures spread should Granger cause the remaining model variables (see Giannone and Reichlin, 2006). We formally tested this proposition and were unable to reject the null of no Granger causality at conventional significance levels for maturities of 1, 3, 6, 9 and 12 months, consistent with the view that the inventory data are as accurate as the oil futures spread.¹²

A second test is based on extraneous information about the time periods during which speculation is known to have taken place. For example, Yergin (1992) and other oil market historians have described a speculative frenzy in oil markets in the second half of 1979 with heavy inventory buying. This provides an opportunity to test whether our structural model estimates correctly identify this episode. Similarly, we have a strong presumption that speculation mattered in 1986, 1990 and 2002. In Section 4, we will show that our inventory data in all these cases generate results consistent with conventional wisdom.

Finally, Baumeister and Kilian (2012a) show that reduced-form VAR models based on our oil inventory proxy generate more accurate real-time out-of-sample forecasts of the real price of oil than other methods even during 2009–2011, further strengthening our claim that the oil inventory proxy is reasonably accurate.

3. IDENTIFICATION

An important question is how to distinguish speculative demand shocks from flow demand and flow supply shocks in practice. Our structural VAR model is set-identified based on a combination of sign restrictions and bounds on the implied price elasticities of oil demand and oil supply.¹³ Some of these restrictions are implied by economic theory, while others can be motivated based on extraneous information. We impose four sets of identifying restrictions, each of which is discussed in turn.

¹² Similar results also hold for *ex ante* real interest rates.

¹³ The use of sign restrictions in oil market VAR models was pioneered by Baumeister and Peersman (2013) and refined by Kilian and Murphy (2012). For a general exposition also see Fry and Pagan (2011) and Inoue and Kilian (2012).

Table I. Sign restrictions on impact responses in VAR model

	Flow supply shock	Flow demand shock	Speculative demand shock
Oil production	–	+	+
Real activity	–	+	–
Real price of oil	+	+	+
Inventories			+

Note: All structural shocks have been normalized to imply an increase in the real price of oil. Missing entries mean that no sign restriction is imposed.

3.1. Impact Sign Restrictions

The sign restrictions on the impact responses of oil production, real activity, the real price of oil and crude oil inventories are summarized in Table I. These restrictions follow directly from the economic model outlined in Section 2. Implicitly, these restrictions also identify the fourth innovation, which can be thought of as a conglomerate of idiosyncratic oil demand shocks. Given the difficulty of interpreting this residual shock economically, we do not report results for this fourth shock but merely note that it is not an important determinant of the real price of oil.

Impact 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 (2012) in the context of a simpler model, it is essential that we impose all credible identifying restrictions for identification for the estimates to be economically meaningful. 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

The price elasticity of oil supply depends on the slope of the oil supply curve. A vertical short-run oil supply curve would imply a price elasticity of zero, for example. An estimate of the impact price elasticity of oil supply may be constructed from the dynamic simultaneous equation 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 flow 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, if not effectively zero.¹⁴ This fact suggests the need for an upper bound on this elasticity in selecting the admissible models that allows for steep, but not quite vertical short-run oil supply curves (see Kilian and Murphy, 2012). It is important to stress that this additional identifying restriction does not constrain the levels of the impact responses but merely imposes a bound on their relative magnitude. In our baseline model, we impose a fairly stringent bound of 0.025 on the impact price elasticity of oil supply. Because any such bound is suggestive only, we also experimented with higher bounds. It can be shown that doubling this bound, while increasing the number of admissible models, has little effect on the shape of the posterior distribution of the impulse responses. Even for a bound of 0.1 the 68% quantiles of the posterior distribution of the impulse responses remain qualitatively similar to the baseline model. Moreover, the estimates of the posterior median price elasticity of oil demand reported in Section 5 are remarkably robust to this change.

¹⁴ For example, Hamilton (2009b, p. 25) observes that ‘in the absence of significant excess production capacity, the short-run price elasticity of oil supply is very low’. In practice, it will often take years for significant production increases. Kilian (2009a) makes the case that, even in the presence of spare capacity, the response of oil supply within the month to price signals will be negligible because changing oil production is costly. Kellogg (2011) using monthly well-level oil production data from Texas finds essentially no response of oil production to either the spot price or the oil futures price.

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 flow 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 definition is inappropriate. The relevant quantity measure is instead 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 – with the exception of Considine (1997) – has not been discussed in the literature, nor has there been any attempt in the literature to estimate this *oil demand elasticity in use*. The reader is referred to the Appendix for a formal discussion of how this elasticity can be derived from the structural VAR model.

A natural additional identifying assumption is that the impact elasticity of oil demand in use, η_t^{Use} , must be weakly negative on average over the sample. In addition to bounding the demand elasticity in use at zero from above, we also impose a lower bound.¹⁵ It is reasonable to presume that the impact price elasticity of oil demand is lower in magnitude than the corresponding long-run price elasticity of oil demand (see, for example, Sweeney, 1984). A benchmark for that long-run elasticity is provided by studies of nonparametric gasoline demand functions based on US household survey data such as Hausman and Newey (1995), which have consistently produced long-run price elasticity estimates near -0.8 . Their estimate suggests a bound of $-0.8 \leq \eta_t^{\text{Use}} \leq 0$.¹⁶

3.4. Dynamic Sign Restrictions

Our final set of restrictions relates to the dynamic responses to a flow supply shock. We impose the additional restriction that the response of the real price of oil to a negative flow supply shock must be positive for at least 12 months, starting in the impact period. This restriction is necessary to rule out structural models in which unanticipated flow supply disruptions cause a decline in the real price of oil below its starting level. Such a decline would be at odds with conventional views of the effects of unanticipated oil supply disruptions. 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 flow supply disruption are negative for the first 12 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 particular, we do not impose any dynamic sign restrictions on the responses of global real activity and oil production to speculative oil demand shocks. Given that this shock is a composite of expectations shocks related to shifts in uncertainty and to the anticipation of rising oil demand and/or falling oil supplies, it is impossible to determine the sign of these responses *ex ante* beyond the impact period.

¹⁵ 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.

¹⁶ In related work, Yatchew and No (2001) using more detailed Canadian data arrive at a long-run gasoline demand elasticity estimate of -0.9 , very close to Hausman and Newey's original estimate.

3.5. 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, for example, Canova and De Nicolò, 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 Σ_e . Let ε_t denote the corresponding structural VAR model innovations. The construction of structural impulse response functions requires an estimate of the $N \times N$ matrix $\tilde{B} \equiv B_0^{-1}$ in $e_t = \tilde{B}\varepsilon_t$. Let $\Sigma_e = PAP'$ and $B = PA^{0.5}$ such that B satisfies $\Sigma_e = BB'$. Then $\tilde{B} = BD$ also satisfies $\tilde{B}\tilde{B}' = \Sigma_e$ for any orthogonal $N \times N$ matrix D . One can examine a wide range of possibilities for \tilde{B} by repeatedly drawing at random from the set \mathbf{D} of orthogonal matrices D . Following Rubio-Ramirez *et al.* (2010) we construct the set $\tilde{\mathbf{B}}$ of admissible models by drawing from the set \mathbf{D} and discarding candidate solutions for \tilde{B} that do not satisfy a set of a priori restrictions on the implied impulse response functions. In practice, this procedure may be implemented conditional on the conventional maximum likelihood/least squares estimator of $A(L)$ and Σ_e in the reduced-form VAR model. This allows the resulting impulse response estimates to be given a frequentist interpretation. To summarize, 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 identifying 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 record the corresponding impulse response functions.

The resulting set $\tilde{\mathbf{B}}$ corresponds to the set of all admissible structural VAR models.

The estimation uncertainty underlying these structural impulse response estimates may be assessed by frequentist or Bayesian methods. We adopt the latter approach and follow the standard approach in the literature of specifying a diffuse Gaussian-inverse Wishart prior distribution for the reduced-form parameters and a Haar distribution for the rotation matrix (see, for example, Inoue and Kilian, 2012). The posterior distribution of the structural impulse responses is obtained by applying our identification procedure to each draw of $A(L)$ and Σ_e from their posterior distribution.

4. ESTIMATION RESULTS

For expository purposes, in the analysis below we focus on one model among the admissible structural models obtained conditional on the least squares estimate of the reduced-form. The results shown are for the model that yields an impact price elasticity of oil demand in use closest to the posterior median of this elasticity among the candidate models that satisfy all identifying restrictions. We also conducted the same analysis with every other admissible structural model and verified that our main results are robust to the choice of admissible model. The only difference is that some admissible models assign even more explanatory power to flow demand shocks than the benchmark model at the expense of speculative demand shocks.

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 along with the corresponding pointwise 68% posterior error bands obtained by drawing

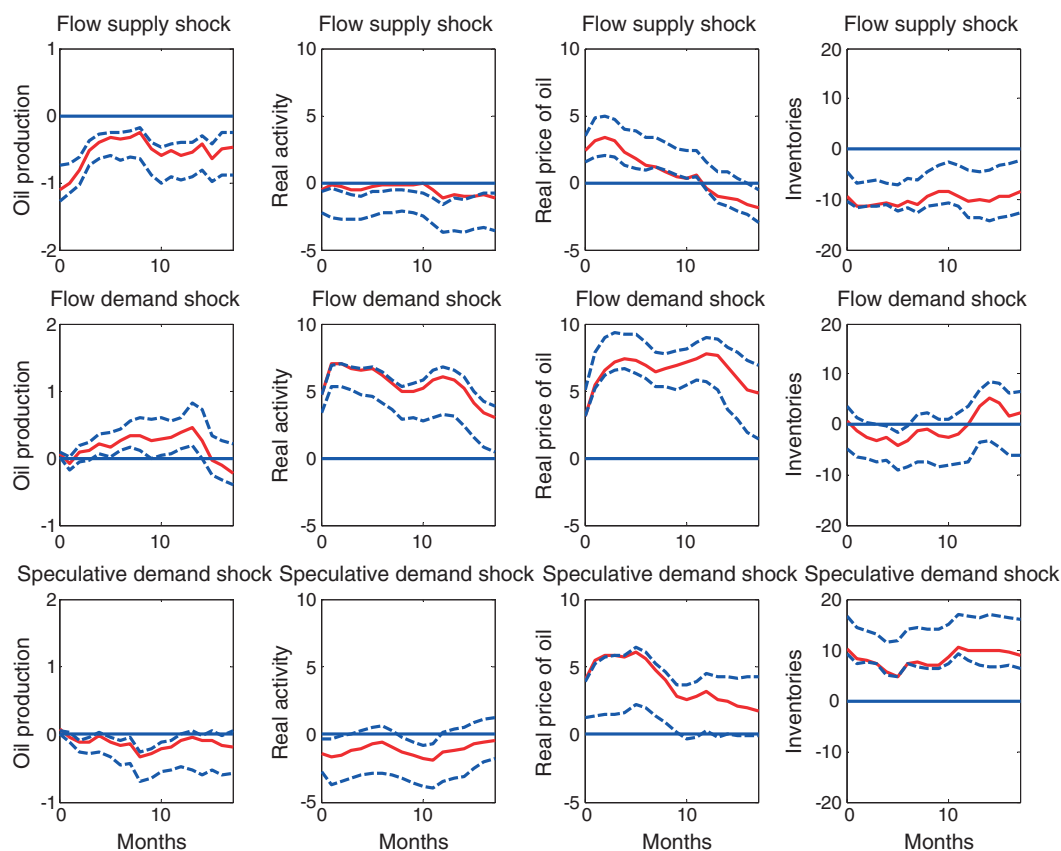


Figure 1. Structural impulse responses: 1973:2–2009:8. Solid lines indicate the impulse response estimates for the model with an impact price elasticity of oil demand in use closest to the posterior median of that elasticity among the admissible structural models obtained conditional on the least-squares estimate of the reduced-form VAR model. Dashed lines indicate the corresponding pointwise 68% posterior error bands. ‘Oil production’ refers to the cumulative percent change in oil production and inventories to cumulative changes in inventories

from the reduced-form posterior distribution. All shocks have been normalized such that they imply an increase in the real price of oil. In particular, the flow supply shock refers to an unanticipated flow supply disruption. Figure 1 illustrates that the role of storage differs depending on the nature of the shock. A flow supply disruption causes inventories to be drawn down in an effort to smooth production of refined products. A positive flow demand shock causes almost no change in oil inventories on impact, followed by a temporary drawdown of oil inventories. After one year, oil inventories reach a level in excess of their starting level. A positive speculative demand shock causes a persistent increase in oil inventories.

A negative flow supply shock is also associated with a reduction in global real activity and a persistent drop in oil production, but much of the initial drop is reversed within the first half year. The real price of oil rises only temporarily. It peaks after three months. 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 shock 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 flow supply shocks than the structural VAR model in Kilian (2009a,b), for example, illustrating the importance of explicitly modeling speculative demand shocks and oil inventories.

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

It can be shown that in the short run 29% of the variation in crude oil inventories is driven by speculative demand shocks, followed by oil supply shocks with 26%. Flow demand shocks have a negligible impact with 2%. At long horizons, in contrast, the explanatory power of speculative demand shocks declines to 27% and that of flow supply shocks to 24%, while the explanatory power of flow demand shocks increases to 15%. This evidence suggests that, on average, fluctuations in oil inventories mainly reflect speculative trading as well as 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, 87% of the variation in the real price of oil can be attributed to flow demand shocks, compared with 9% due to speculative demand shocks and 3% due to flow supply shocks.

Impulse responses and forecast error 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 assess the quantitative importance of speculative demand shocks as opposed to other demand and supply shocks at each point in time since the late 1970s.¹⁷

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. The sharp increase in the real price of oil in 2007/08, especially, has been attributed to speculation. The standard interpretation is (a) that there was an exogenous influx of financial investors into the oil futures market, (b) that this influx drove up oil futures prices and (c) that the increase in oil futures prices was viewed by spot market participants as a signal of an expected increase in the price of oil, shifting inventory demand and hence causing the real spot price to increase.

This explanation implies that speculative demand shocks in the structural model should explain the bulk of the surge in the real price of oil after 2003. Figure 2 shows that there is no evidence to support this view. There has been no systematic upward movement in the real price of oil after 2003 associated with speculative demand shocks. This result has far-reaching implications. First, in the policy debate, it is common to distinguish between normal speculation in oil markets that reflects expected fundamentals and purely financial speculation that is viewed as excessive. While there is no operational definition of ‘excessive speculation’, as discussed in Fattouh *et al.* (2013), the evidence in Figure 2 suggests that the distinction between normal speculation and speculation that is excessive is moot, for if there is no speculation in the physical market at all there cannot have been excessive speculation either under any definition. Second, this result tells us that an exogenous influx of financial speculators cannot have driven up the oil futures price, because – under the standard assumption of arbitrage between

¹⁷ We do not include the contribution of the residual demand shock because that shock makes no large systematic contribution to the evolution of the real price of oil.

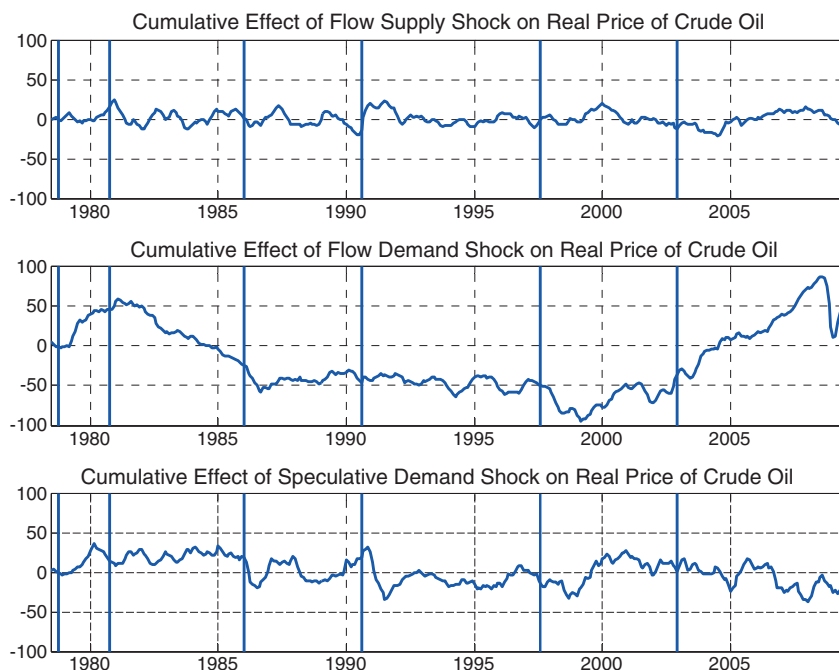


Figure 2. Historical decompositions for 1978:6–2009:8. Based on benchmark estimate as in Figure 1. The vertical bars indicate major exogenous events in oil markets, notably the outbreak of the Iranian Revolution in 1978:9 and of the Iran–Iraq War in 1980:9, the collapse of OPEC in 1985:12, the outbreak 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

the futures and spot markets for oil maintained by proponents of the financial speculation hypothesis – the absence of speculation in the spot market also rules speculation in oil futures markets.¹⁸

A competing view of speculation is that OPEC in anticipation of even higher oil prices held back its production after 2001, using oil below ground effectively using oil below the ground as inventories (see, for example, Hamilton (2009a, p. 239)). One way of testing this hypothesis is through the lens of our structural model. In our model, OPEC holding back oil production in anticipation of rising oil prices would be observationally equivalent to a negative flow supply shock. Figure 2 provides no indication that negative flow supply shocks were an important determinant of the real price of oil between 2003 and mid 2008. 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. Hence we can reject the speculative supply shock hypothesis.

An alternative explanation of the evolution of the real price of oil is the peak oil hypothesis, which predicted that around 2006 world oil production should have peaked. If so, one would have expected to

¹⁸ Our analysis of the spot market would remain valid if this arbitrage were impeded or broke down completely, but the oil futures price would become disconnected from the spot price in that case. In the limiting case of no arbitrage we would be unable to infer from our model whether there is speculation in oil futures markets, although we could still infer whether there is speculation taking place in the spot market. Clearly, a situation in which arbitrage breaks down is not consistent with the scenario envisioned by researchers who attribute rising spot prices to speculation by financial investors, because in that case speculation-driven increases in the oil futures price could not possibly be transmitted to the spot price of oil.

see a sequence of negative flow supply shocks drive up the real price of oil after 2005, but we already showed that flow supply shocks have very limited explanatory power. The peak oil hypothesis could also affect the real price of oil if traders, rightly or wrongly, believed in this hypothesis and stocked up on oil in anticipation of a shortage of oil. That explanation would be observationally equivalent to the financial speculation hypothesis, however, which we already rejected on the basis of the results in Figure 2. Hence 'peak oil' may be safely ruled out as an explanation of the surge in the real price of oil after 2003, along with financial speculation and speculation by oil producers.

Instead our model provides a different explanation. It supports the substantive conclusion in Kilian (2009a,b) 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 associated with the global business cycle. It is important to stress that this result does not arise by construction. Indeed, the identification of our model is quite different from that in Kilian (2009a) and, as we will show below, the two models may produce substantively different empirical results. Our model shows that the run-up in the real price of oil occurred because of the cumulative effects of many positive flow demand shocks over the course of several years. It may seem unlikely *ex ante* that a model would generate many more positive than negative demand shocks between 2003 and mid 2008, but Kilian and Hicks (2013) show that this feature is consistent with the errors in professional real GDP forecasts during this period. Even professional forecasters persistently underestimated global growth during 2003–2008, especially growth in emerging Asia, lending credence to our model results.

This situation only changed with the financial crisis of late 2008. The V-shaped dip in the real price of oil in 2008/09 coincided with a similar dip in the global real activity measure and is largely driven by flow demand shocks. A similar, if much less pronounced, dip had followed the Asian crisis in 1997. Whereas the recovery from the all-time low in the real price of oil in 1999–2000 resulted from a combination of coordinated OPEC oil supply cuts, a gradual increase in flow demand (often associated with the US productivity boom) and increased speculative demand perhaps in anticipation of increased future real activity and/or further oil supply reductions, the resurgence of the real price of oil starting in early 2009 reflected primarily increased flow demand (see Figure 2).

We conclude that economic fundamentals on the demand side of the oil market are capable of explaining the evolution of the real price of oil during the last decade. No non-standard explanations are required. This finding is important because it implies that further regulation of oil markets would have done nothing to stem the increase in the real price of oil. Indeed, it shows that there is no basis for the premise that such regulation is required to lower the real price of oil or that it would be helpful. Our structural model also implies that even dramatic increases in US oil production would not lower the real price of oil substantially at the global level, while a full recovery of the global economy would raise the real price of oil by as much as \$50 in real terms (see Baumeister and Kilian, 2012b).

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, suggesting that our model has the ability to detect speculative demand shocks when they exist. One particularly interesting example is the oil price spike associated with the Persian Gulf War of 1990/91. In related work, Kilian (2009a) presented indirect evidence – based on a model without oil inventories – that the 1990 oil price increase was driven mainly by a shift in speculative demand (reflecting concerns 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. This observation prompted Hamilton to reject the hypothesis that shifts in speculative demand were behind the sharp increase in the real price of oil in mid 1990 and

its fall after late 1990. Given the consensus that flow demand did not move up sharply in mid 1990, Hamilton suggested that perhaps this price increase must be attributed to flow 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. Moreover, there is general agreement that flow supply shocks cannot explain the collapse of the real price of oil in late 1990. 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 modest decline in oil inventories. At the same time, the observed large 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 about one third of the price increase from July to August of 1990 was caused by speculative demand shocks and two thirds by flow supply shocks. This result is in sharp contrast to the estimates in Kilian (2009a), 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 November of 1990, when the threat of Saudi oil fields being captured by Iraq had been removed by the presence of US 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

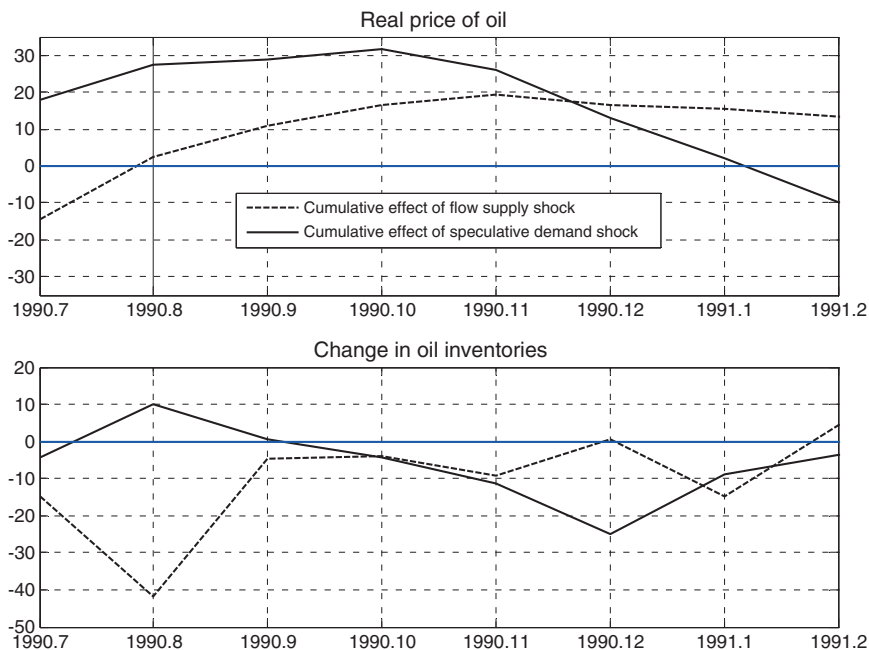


Figure 3. Historical decompositions for the Persian Gulf War episode of 1990/91

sharp decline in the real price of oil starting in late 1990 with data on oil production. The evidence of a sharp decline in speculative demand in late 1990 in turn raises the obvious question of when and why speculative demand had surged in the first place. The bottom panel of Figure 2 reveals there was in fact a substantial increase in speculative demand already in the months leading up to the invasion. 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.¹⁹ The reason that this increase in the speculative component of the real price of oil went unnoticed by the general public was a simultaneous substantial unexpected increase in oil production with offsetting effects on the real price of oil in early 1990, as shown in the top panel of Figure 2. In fact, that expansion of oil production served as the motivation for Iraq's increasing hostility to neighboring countries such as Kuwait, which it accused of undermining the price of oil, making it more difficult for Iraq to service its foreign debt. Taken in conjunction our evidence implies a much larger role for speculative demand in 1990/91 than the data for the month of August alone would suggest.

4.5. What Caused the 1979 and 1980 Oil Price Shocks?

Speculative demand also played an important role in 1979. The traditional interpretation of this episode is that this oil price increase was caused 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 later in 1979 after Iranian oil production had resumed. 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, coupled with expectations of strong flow demand.²⁰

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 after May of 1979 by a sharp increase in speculative demand, exactly as described by Yergin (1992). Whereas flow demand pressures on the real price of oil gradually receded starting in 1981, speculative demand pressures on average remained relatively 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, in part 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 historical decomposition in Figure 4 confirms that negative 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

¹⁹ In this regard, Gause (2002) notes a shift in Iraqi foreign policy toward a more aggressive stance in early 1990.

²⁰ 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.' Also see Yergin (1992, p. 687).

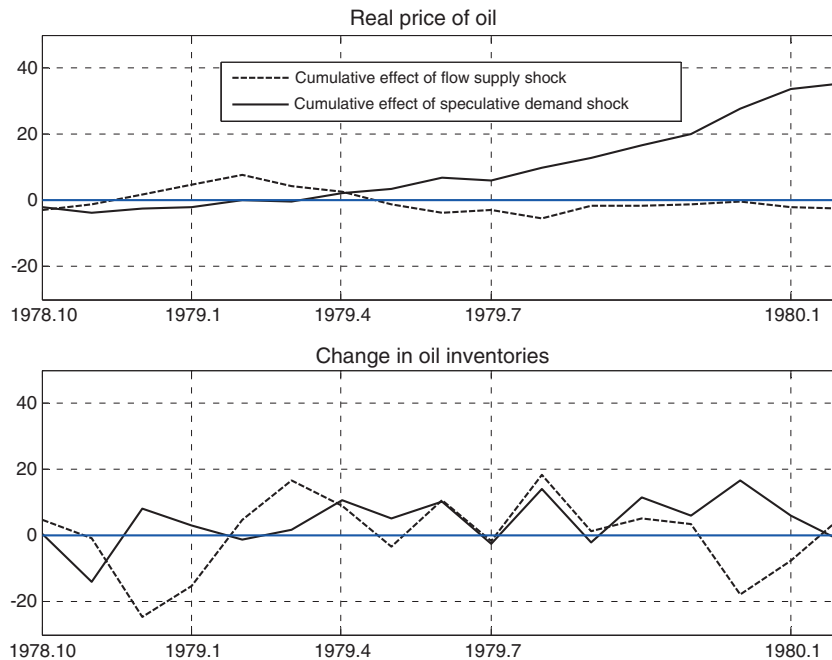


Figure 4. Historical decompositions for the Iranian Revolution of 1978/79

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 continued into 1980. Thus there is no indication that flow supply shocks played an important role in 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 disruption was slightly larger than the price response to the 1979 supply disruption. There is also evidence of a small resurgence of speculative demand 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. This pattern is similar to the pattern in Figure 3, except in reverse. Although OPEC attempted to reunite and control production in 1987, amidst increased speculation, these attempts proved unsuccessful in the long run.

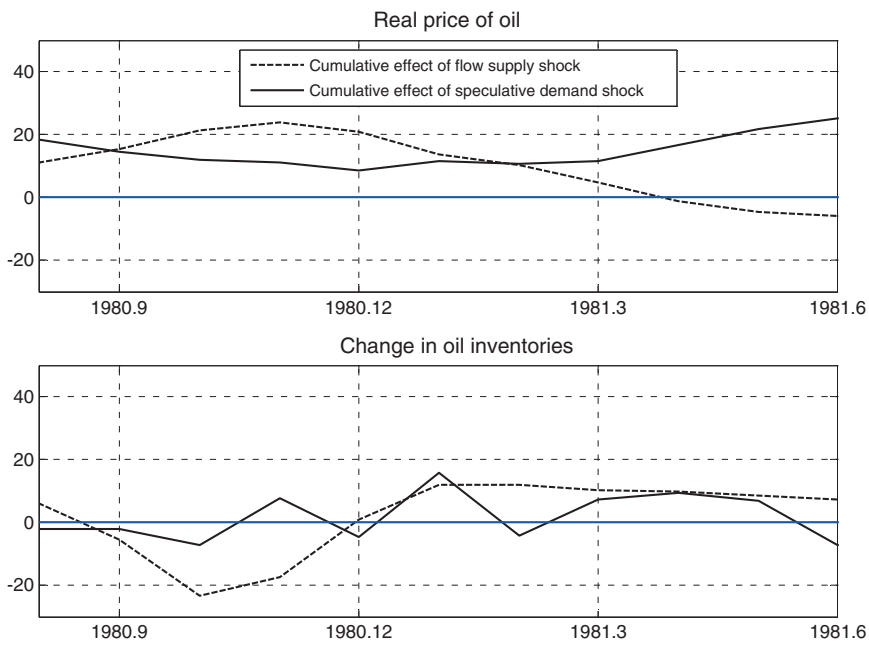


Figure 5. Historical decompositions for the outbreak of the Iran–Iraq War in 1980

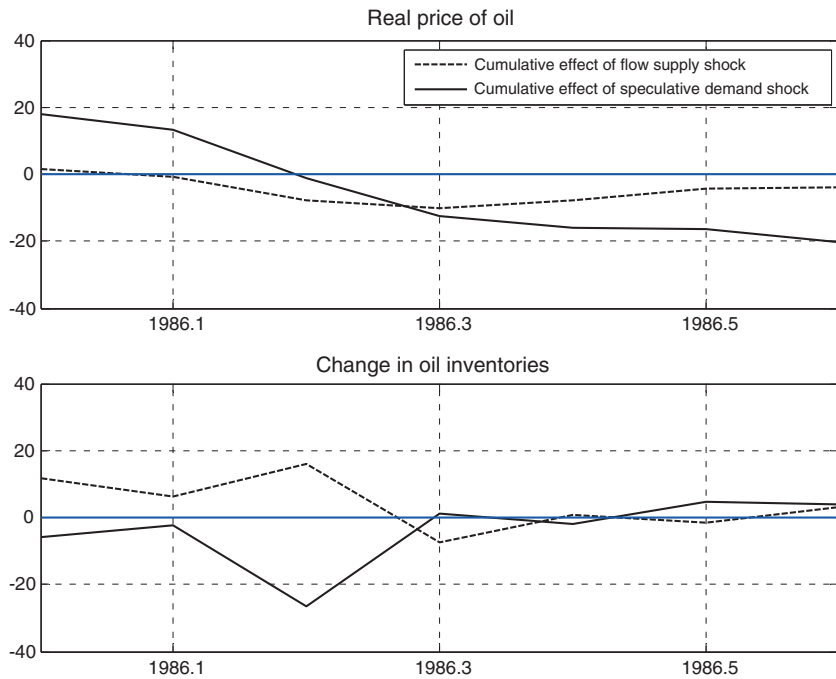


Figure 6. Historical decompositions for the collapse of OPEC in 1986

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

Figure 7 focuses on the flow 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 a magnitude similar to the oil supply disruptions of the 1970s (see Kilian, 2008a). Figure 7 shows that this event reflected a combination of negative flow supply shocks and positive 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 this period also coincided with an increase in speculative demand in anticipation of the 2003 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. IMPLICATIONS OF THE MODEL FOR THE SHORT-RUN PRICE ELASTICITY OF OIL DEMAND

The short-run price elasticity of oil demand has important implications for theoretical models of speculative demand. For example, it is a key parameter in the theoretical models of speculation

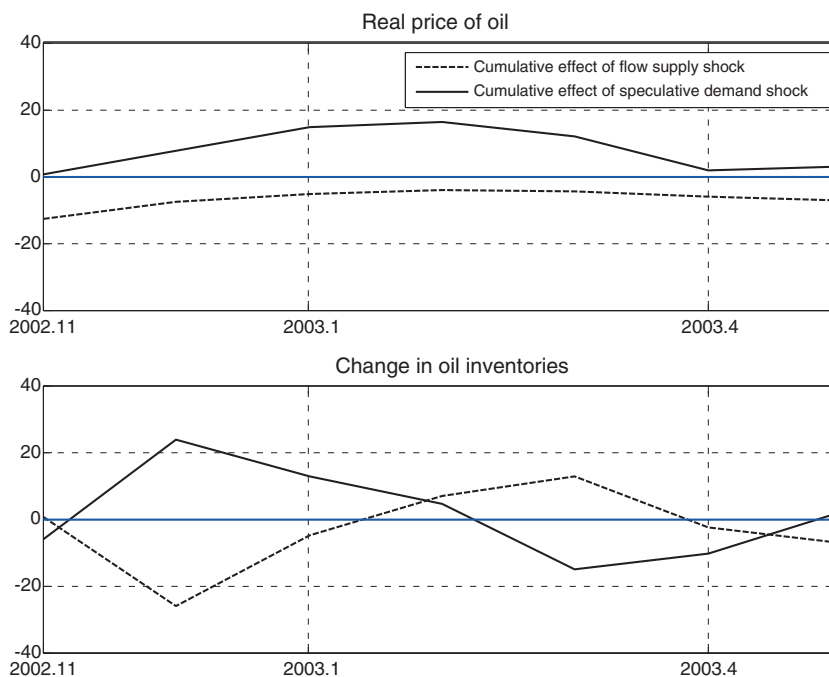


Figure 7. Historical decompositions for Venezuelan crisis and Iraq War in 2002/03

discussed in Caballero *et al.* (2008) and Hamilton (2009a). All else equal, standard models of speculation imply that oil inventories must increase to enable the real price of oil to increase. Recently, Hamilton (2009a) observed that speculation in oil futures markets may drive up the real price of oil even without any increase in oil inventories, if refiners are able to pass on fully to gasoline consumers an exogenous increase in the real price of oil driven by speculation. This result requires the demand for gasoline to be completely price-inelastic. As shown in the online Appendix (supporting information), the short-run price elasticity of gasoline demand is approximately of the same magnitude as the short-run price elasticity of oil demand in use. Whether the limiting case discussed in Hamilton (2009a) is empirically relevant thus depends on the magnitude of the short-run price elasticity of oil demand.

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 magnitude of the short-run price elasticity of oil demand. It is widely believed that this elasticity is close to zero, making it difficult to rule out the limiting case described by Hamilton. Although there is no shortage of elasticity estimates in the literature that seem to confirm this impression, these estimates suffer from two limitations.

First, much of the existing literature has attempted to estimate the oil demand elasticity from dynamic reduced-form models that do not distinguish between oil demand and oil supply shocks (see, for example, Dahl, 1993; Cooper, 2003). This is not possible because the identification of the demand elasticity requires an exogenous shift of the contemporaneous oil supply curve along the contemporaneous oil demand curve within the context of a structural model. Reduced-form estimates of the oil demand elasticity fail to account for the endogeneity of the price of crude oil and hence are biased toward zero. This fact helps explain the low elasticity estimates typically reported in the literature.²¹ Second, typical estimates of the oil demand elasticity in the literature have been based on models that equate the percent change in quantity with the percent change in the production of crude oil, ignoring the existence of oil inventories. In this paper, we refer to this conventional elasticity measure as the oil demand elasticity in production, denoted by $\eta^{O, \text{Production}}$. A more appropriate definition of the price elasticity of oil demand for policy questions is the elasticity in use. The latter demand elasticity is based on the change in the use of oil, defined as the sum of the change in oil production and of the depletion of oil inventories, which more accurately captures the response of oil consumers.

5.1. The Short-Run Price Elasticity of Oil Demand in Production

Our structural model of the oil market may be used to obtain direct estimates of the short-run price elasticity of oil demand in production and in use, allowing us to assess the empirical relevance of models relying on a zero price elasticity of oil demand. The elasticity in production can be estimated from model (1) as the ratio of the impact response of oil production to a flow supply shock relative to the impact response of the real price of oil. Our posterior median estimate of this elasticity, as shown in the first column of the upper panel of Table II, is -0.44 . This estimate is seven times higher than typical conjectures in the recent literature. It is also much higher in magnitude than conventional reduced-form regression estimates of this elasticity. For example, surveys by Dahl (1993) and Cooper (2003) report estimates between -0.05 and -0.07 . The difference in results can be attributed to the difference between estimating a structural and a reduced-form model.²² The first column in the upper

²¹ Producers of these estimates sometimes acknowledge the need for instrumental variable estimation methods, but, having acknowledged this point, tend to revert to using OLS, given the absence of suitable instruments.

²² One way of demonstrating this point is to note that fitting the conventional reduced-form log-level specification used in some of the earlier literature to our data would yield an elasticity estimate of only -0.02 , in line with the existing consensus. Another way of putting these results into perspective is to observe that other recent studies relying on alternative structural models have obtained similarly large oil demand elasticity estimates ranging from -0.35 to -0.41 (see, for example, Serletis *et al.*, 2010; Bodenstern and Guerrieri, 2012; Baumeister and Peersman, 2013).

Table II. Posterior distribution of the short-run price elasticity of demand for crude oil

		$\eta^{O, \text{Production}}$	$\eta^{O, \text{Use}}$
$\eta_{\text{Supply}}^{\text{Supply}} \leq 0.025$	16th percentile	-0.80	-0.54
	50th percentile	-0.44	-0.26
	84th percentile	-0.23	-0.09
$\eta_{\text{Supply}}^{\text{Supply}} \leq 0.050$	16th percentile	-0.80	-0.57
	50th percentile	-0.45	-0.27
	84th percentile	-0.29	-0.09
$\eta_{\text{Supply}}^{\text{Supply}} \leq 0.100$	16th percentile	-0.76	-0.61
	50th percentile	-0.47	-0.30
	84th percentile	-0.24	-0.10

Note: $\eta^{O, \text{Production}}$ refers to the impact price elasticity of oil demand in production and $\eta^{O, \text{Use}}$ to the average impact price elasticity of oil demand in use. The latter definition accounts for the role of inventories in smoothing oil consumption. $\eta_{\text{Supply}}^{\text{Supply}}$ refers to the impact price elasticity of oil supply.

panel of Table II also shows the 68% posterior credible set for this elasticity. The model assigns substantial probability mass to values between -0.80 and -0.23 and very little probability mass to values close to zero.

5.2. The Short-Run Price Elasticity of Oil Demand in Use

The posterior median estimate of $\eta^{O, \text{Production}}$ in the first column of Table II, while instructive when compared to conventional estimates, is misleading in that it ignores the role of inventories. Our model also permits the estimation of the price elasticity of oil demand in use, allowing us to assess the role of changes in inventories (see Appendix). By construction, allowing for inventory responses will tend to lower the magnitude of the price elasticity of oil demand. The second column in the upper panel of Table II shows that posterior median estimate of $\eta^{O, \text{Use}}$ is only -0.26 , compared with the estimate of -0.44 for the demand elasticity in oil production. While this large reduction in the magnitude of the elasticity highlights the importance of accounting for changes in oil inventories, the revised median estimate is still four times larger than conventional elasticity estimates. Indeed, this result illustrates that relatively high short-run price elasticities of oil demand are fully compatible with the view that economic fundamentals are responsible for the surge in the real price of oil after 2003.²³ Moreover, the second column of Table II again shows that there is little probability mass on elasticity values close to zero, casting doubt on models of speculation that do not involve a change in oil inventories.²⁴

6. 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

²³ The lower panels of Table II show that these estimates are quite robust to relaxing the upper bound on the impact price elasticity of oil supply. Relaxing this bound to 0.050 or even to 0.100 raises the median oil demand elasticity in use slightly without affecting the substance of the conclusions.

²⁴ It is worth noting that our estimate of the price elasticity of gasoline demand is larger than some estimates in the literature. For example, Hughes *et al.* (2008) report an average elasticity estimate of -0.18 based on US data for 1975–1980 and for 2001–2006, similar to estimates in Dahl and Sterner (1991). On the other hand, Burger and Kaffine (2009) report estimates as high as -0.29 . Our estimate is also smaller than the instrumental variable regression estimate of the gasoline tax elasticity of gasoline demand of -0.47 reported in Davis and Kilian (2011), with a standard error of 0.23. An unresolved question is to what extent the price elasticity of gasoline demand may have declined in magnitude in very recent years and at what time (see, for example, Hughes *et al.*, 2008). It would take a substantial decline, however, to make the limiting case of a zero elasticity discussed in Hamilton (2009a) empirically relevant.

to the flow supply of oil and the flow demand for oil. In this paper we augmented the structural model to include shocks to inventory demand reflecting shifts in expectations about future oil supply and future oil demand that cannot be captured by flow demand or flow supply shocks. Such speculative demand 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 rather than an increase in future oil demand. 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. We proposed a dynamic simultaneous equation model including crude oil inventories that allows the simultaneous identification of all three types of shocks conditional on past data.

The inclusion of oil inventories matters. For example, our structural model implies a larger role for flow supply shocks in explaining fluctuations in the real price of oil than previous estimates. The added explanatory power of flow supply shocks for the real price of oil, especially in 1990, comes at the expense of the explanatory power of speculative demand shocks. We showed that, nevertheless, the largest and most persistent fluctuations in the real price of oil since the 1970s have been associated primarily with business cycle fluctuations affecting the flow demand for crude oil. Of particular interest for recent policy debates is the sustained increase in the real price of oil from 2003 until mid 2008. In the context of a model that nests all leading explanations of how this oil price surge came about, we were able to provide direct evidence against the popular view that this increase was driven by speculators. This conclusion holds even for the 2007/08 period. Shifts in speculative demand played a more important role during several earlier oil price shock episodes, however, notably in 1979, 1986, 1990 and 2002. 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 'peak oil' (or for that matter deliberate production cutbacks by oil producers) had much bearing on the recent oil price fluctuations. Rather, our results support recent findings in the literature that the sustained run-up in the real price of oil between 2003 and mid 2008 was caused primarily by shifts in the global flow demand for oil. Likewise, the collapse of the real price of oil in late 2008 and its partial recovery since early 2009 have been primarily driven by flow demand. The model implies that the real price of oil is expected to rise further 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. By contrast, additional regulation of oil derivatives markets is not likely to lower the real price of oil, nor can increased domestic oil production in the USA be expected to have much of an effect on the global real price of oil (also see Baumeister and Kilian, 2012b).

Our model also reconciles some seemingly puzzling observations related to earlier oil price surges. For example, it has been noted 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 rather than reduced-form evidence in interpreting the price and quantity data.

The use of a structural regression model also is important for the estimation of the short-run price elasticity of oil demand. For example, Hamilton (2009a,b) suggested that 1978–1981 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 1978–1981. Oil supply shocks played a small role only, violating the premise of Hamilton's calculations.

We also observed that traditional estimates of the short-run price elasticity of oil demand are not credible. One problem is that conventional estimates of this elasticity from dynamic reduced-form regressions, as in Dahl (1993) and Cooper (2003), have ignored the endogeneity of the real price of oil, causing the elasticity estimate to be downward biased. Moreover, existing estimates, including the structural estimates recently provided by Baumeister and Peersman (2013), have ignored the role of inventories in smoothing oil consumption in response to oil supply shocks. We provided a model that allows the estimation of both the traditional oil demand elasticity in production and of the more relevant oil demand elasticity in use which incorporates changes in oil inventories. Our short-run elasticity estimates are substantially higher than standard estimates cited in the literature, casting doubt on models of speculative trading based on perfectly price-inelastic short-run demand for oil and for gasoline.

ACKNOWLEDGEMENTS

We thank the editor, two anonymous referees, Christiane Baumeister, Domenico Giannone, Ana María Herrera, Ryan Kellogg, Chris Knittel, Harald Uhlig, Whitney Newey and Wei Xiong for helpful discussions.

REFERENCES

- Alquist R, Kilian L. 2010. What do we learn from the price of crude oil futures? *Journal of Applied Econometrics* **25**: 539–573.
- Barsky RB, Kilian L. 2002. Do we really know that oil caused the great stagflation? A monetary alternative. In *NBER Macroeconomics Annual 2001*, Bernanke BS, Rogoff K (eds.); MIT Press: Cambridge, MA; 137–183.
- Baumeister C, Kilian L. 2012a. Real-time forecasts of the real price of oil. *Journal of Business and Economic Statistics* **30**: 326–336.
- Baumeister C, Kilian L. 2012b. Real-time analysis of oil price risks using forecast scenarios. Mimeo, Department of Economics, University of Michigan.
- Baumeister C, Peersman G. 2013. The role of time-varying price elasticities in accounting for volatility changes in the crude oil market. *Journal of Applied Econometrics* (forthcoming).
- Bodenstein M, Guerrieri L. 2012. On the sources of oil price fluctuations. Mimeo, Federal Reserve Board.
- Burger NE, Kaffine DT. 2009. Gas prices, traffic, and freeway speeds in Los Angeles. *The Review of Economics and Statistics* **91**: 652–657.
- Caballero RJ, Farhi E, Gourinchas PO. 2008. Financial crash, commodity prices, and global imbalances. *Brookings Papers on Economic Activity* **2 Fall**: 1–55.
- Canova F, De Nicolò D. 2002. Monetary disturbances matter for business cycle fluctuations in the G-7. *Journal of Monetary Economics* **49**: 1131–1159.
- Considine TJ. 1997. Inventories under joint production: an empirical analysis of petroleum refining. *The Review of Economics and Statistics* **79**: 493–502.
- Cooper JCB. 2003. Price elasticity of demand for crude oil: estimates for 23 countries. *OPEC Review* **27**: 1–8.
- Dahl CA. 1993. A survey of oil demand elasticities for developing countries. *OPEC Review* **17**: 399–419.
- Dahl CA, Sterner T. 1991. Analyzing gasoline demand elasticities: a survey. *Energy Economics* **13**: 203–210.
- Davis LW, Kilian L. 2011. Estimating the effect of a gasoline tax on carbon emissions. *Journal of Applied Econometrics* **16**: 1187–1214.
- Dvir E, Rogoff K. 2010. Three epochs of oil. Mimeo, Department of Economics, Harvard University.
- Fattouh B, Kilian L, Mahadeva L. 2013. The role of speculation in oil markets: what have we learned so far? *Energy Journal* (forthcoming).
- Fry R, Pagan AR. 2011. Sign restrictions in structural vector autoregression: a critical review. *Journal of Economic Literature* **49**: 938–960.

- Gause GF. 2002. Iraq's decision to go to war, 1980 and 1990. *Middle East Journal* **56**: 47–71.
- Giannone D, Reichlin L. 2006. Does information help recover structural shocks from past observations? *Journal of the European Economic Association* **4**: 455–465.
- Hamilton JD. 2009a. Causes and consequences of the oil shock of 2007–08. *Brookings Papers on Economic Activity* **1**(Spring): 215–261.
- Hamilton JD. 2009b. Understanding crude oil prices. *Energy Journal* **30**: 179–206.
- Hamilton JD, Herrera AM. 2004. Oil shocks and aggregate economic behavior: the role of monetary policy. *Journal of Money, Credit, and Banking* **36**: 265–286.
- Hausman JA, Newey WK. 1995. Nonparametric estimation of exact consumers' surplus and deadweight loss. *Econometrica* **63**: 1445–1476.
- Hughes JE, Knittel CR, Sperling D. 2008. Evidence of a shift in the short-run price elasticity of gasoline demand. *Energy Journal* **29**: 113–134.
- Inoue A, Kilian L. 2012. Inference on impulse response functions in structural VAR models. *Journal of Econometrics* (forthcoming).
- Kellogg R. 2011. Short-run responses of oil production to spot and future oil prices. Mimeo, Department of Economics, University of Michigan.
- Kilian L. 2008a. Exogenous oil supply shocks: how big are they and how much do they matter for the U.S. economy? *The Review of Economics and Statistics* **90**: 216–240.
- Kilian L. 2008b. The economic effects of energy price shocks. *Journal of Economic Literature* **46**: 871–909.
- Kilian L. 2009a. Not all oil price shocks are alike: disentangling demand and supply shocks in the crude oil market. *American Economic Review* **99**: 1053–1069.
- Kilian L. 2009b. Comment on 'Causes and consequences of the oil shock of 2007–08' by James D. Hamilton. *Brookings Papers on Economic Activity* **1**(Spring): 267–278.
- Kilian L, Hicks B. 2013. Did unexpectedly strong economic growth cause the oil price shock of 2003–2008? *Journal of Forecasting* (forthcoming).
- Kilian L, Murphy DP. 2012. Why agnostic sign restrictions are not enough: understanding the dynamics of oil market VAR models. *Journal of the European Economic Association* **10**: 1166–1188.
- Rubio-Ramirez J, Waggoner D, Zha T. 2010. Markov-switching structural vector autoregressions: theory and application. *Review of Economic Studies* **77**: 665–696.
- Serletis A, Timilsina G, Vasetsky O. 2010. International evidence on sectoral interfuel substitution. *Energy Journal* **31**: 1–29.
- Sockin M, Xiong W. 2012. Feedback effects of commodity futures prices. Mimeo, Department of Economics, Princeton University.
- Sweeney JL. 1984. The response of energy demand to higher prices: what have we learned? *American Economic Review* **74**: 31–37.
- Terzian P. 1985. *OPEC. The Inside Story*. Zed Books: London.
- Uhlig H. 2005. What are the effects of monetary policy on output? Results from an agnostic identification procedure. *Journal of Monetary Economics* **52**: 381–419.
- Yatchew A, No JA. 2001. Household gasoline demand in Canada. *Econometrica* **69**: 1697–1709.
- Yergin D. 1992. *The Prize: The Epic Quest for Oil, Money, and Power*. Simon & Schuster: New York.

APPENDIX: CONSTRUCTION OF THE PRICE ELASTICITY OF OIL DEMAND IN USE

In this Appendix we show how the oil demand elasticity in use can be approximated with the help of our structural model of the oil market. The amount of oil used in period t , denoted by U_t , equals the quantity of oil produced in that period (Q_t) minus the oil that is added to the stock of inventories (ΔS_t):

$$U_t = Q_t - \Delta S_t$$

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

$$\eta_t^{\text{Use}} \equiv \frac{\% \Delta U_t}{\% \Delta P_t} = \frac{\Delta Q_t - \Delta^2 S_t}{Q_{t-1} - \Delta S_{t-1}} \frac{\% \Delta P_t}{\% \Delta P_t}$$

where Δ represents changes, $\% \Delta$ indicates percent changes in response to an oil supply shock in period t , and P_t denotes the real price of oil. Denote by \tilde{B}_{11} the impact response of the percent change in oil production to an oil supply shock, where \tilde{B}_{ij} refers to the ij th element of \tilde{B} . Then the implied change in oil production is $\Delta Q_t = Q_{t-1} \times (1 + \tilde{B}_{11}/100) - Q_{t-1} = Q_{t-1} \times \tilde{B}_{11}/100$. Moreover, $\Delta^2 S_t = \Delta S_t - \Delta S_{t-1} = \bar{\Delta S} + \tilde{B}_{41} - \bar{\Delta S} = \tilde{B}_{41}$, where the change in oil inventories in response to the oil supply shock equals the impact response \tilde{B}_{41} and, prior to the shock, the change in crude oil inventories is equal to its mean $\bar{\Delta S}$, which is observable. Finally, the impact percent change in the real price of oil in response to an oil supply shock is \tilde{B}_{31} . Hence the demand elasticity in use can be expressed equivalently as

$$\eta_t^{\text{Use}} = \frac{(Q_{t-1} \times \tilde{B}_{11}/100) - \tilde{B}_{41}}{Q_{t-1} - \bar{\Delta S}} \frac{\tilde{B}_{31}/100}{\tilde{B}_{31}/100}$$

Note that by construction η_t^{Use} depends on Q_{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 $\eta^{O, \text{Use}}$.