Macroeconomic Dynamics, **15** (Supplement 3), 2011, 337–363. Printed in the United States of America.
doi:10.1017/S1365100511000186

NONLINEARITIES IN THE OIL PRICE–OUTPUT RELATIONSHIP

LUTZ KILIAN
University of Michigan

ROBERT J. VIGFUSSON
Federal Reserve Board

It is customary to suggest that the asymmetry in the transmission of oil price shocks to real output is well established. Much of the empirical work cited as being in support of asymmetry, however, has not directly tested the hypothesis of an asymmetric transmission of oil price innovations. Moreover, many of the papers quantifying these asymmetric responses are based on censored oil price VAR models that have recently been shown to be invalid. Other studies are based on dynamic correlations in the data and do not distinguish between cause and effect. Recently, several new methods of testing and quantifying asymmetric responses of U.S. real economic activity to positive and negative oil price innovations have been developed. We put this literature into perspective, contrast it with more traditional approaches, highlight directions for further research, and reconcile some seemingly conflicting results reported in the literature.

Keywords: Oil Price, Asymmetries, Nonlinearities

1. INTRODUCTION

There is an ongoing debate in the macroeconomic literature about whether unexpected increases in the price of oil cause recessions in oil-importing countries [see, e.g., Hamilton (2009); Kilian (2008)]. Standard theoretical models of the transmission of exogenous oil price shocks, which imply symmetric responses to oil price increases and decreases, cannot explain large declines in aggregate economic activity in response to positive oil price shocks. In contrast, less conventional models that imply asymmetries in the response of aggregate real output to positive and to negative oil price shocks have the ability to explain both larger recessions in response to unexpected oil price increases and smaller economic expansions in response to unexpected declines in the price of oil. Proponents of the view that positive oil price shocks have been the major cause of recessions in the United States therefore inevitably appeal to the presence of asymmetries in

The views in this paper are solely the responsibility of the authors and should not be interpreted as reflecting the views of the Board of Governors of the Federal Reserve System or of any other person associated with the Federal Reserve System. We thank Ron Alquist, Christiane Baumeister, James Hamilton, Ana María Herrera, and four anonymous referees for helpful comments. Address correspondence to: Lutz Kilian, Department of Economics, University of Michigan, 611 Tappan Street, Ann Arbor, MI 48109-1220, USA; e-mail: lkilian@umich.edu.

© 2011 Cambridge University Press 1365-1005/11

https://doi.org/10.1017/S1365100511000186 Published online by Cambridge University Press

the transmission of oil price shocks. This makes the question of how asymmetric the responses of real output are central to the larger question of what lessons to draw from the historical evidence of the 1970s and 1980s. This question also is paramount in assessing the effects of major unexpected declines in the price of oil, as occurred in 1986 and 1998, for example, or, more recently, in late 2008.

It is customary to suggest that the asymmetry in the transmission of oil price shocks to real output is well established. Much of the empirical work cited as being in support of asymmetry, however, has not directly tested the hypothesis of asymmetric transmission of oil price innovations. In fact, many of the papers quantifying these asymmetric responses are based on the censored oil price VAR methodology, which Kilian and Vigfusson (in press) proved to be invalid. Other studies are based on dynamic correlations in the data that do not shed light on the central question of whether the structural responses of real output triggered by positive and negative oil price innovations are asymmetric.

Recently, there has been renewed interest in developing empirical methodologies aimed at establishing and quantifying asymmetries of the response of real output depending on the sign of oil price innovations. One problem is how to detect asymmetric response functions. The problem is not that there have not been earlier attempts to test for asymmetries, but that the simple diagnostic tests commonly used in the literature dating back to the 1990s are not informative about the degree of asymmetry of the response functions of real economic activity. An alternative impulse–response based test of the null of symmetric response functions has been proposed by Kilian and Vigfusson (in press). At the same time, there has been increasing recognition of the importance of using fully specified structural models in constructing estimates of asymmetric responses to oil price innovations [see, e.g., Elder and Serletis (2010); Kilian and Vigfusson (in press)]. The new econometric models proposed in the recent literature differ in how much parametric structure they impose in estimating these response functions. They also tend to produce different empirical results.

Our objective is to put this recent literature into perspective, to contrast it with more traditional approaches, to highlight directions for further research, and to reconcile some seemingly conflicting results reported in the literature. The remainder of the paper is organized as follows. Section 2 reviews the theoretical rationale for asymmetric responses of real economic activity to oil price shocks. In Section 3 we discuss the key modeling choices that affect the strength of the empirical evidence in favor of asymmetry. Section 4 focuses on the encompassing regression approach employed by Kilian and Vigfusson (in press) and Herrera et al. (in press) to quantify potentially asymmetric responses. Section 5 reviews the GARCH-in-mean VAR model designed by Elder and Serletis (2010) to quantify the effect of oil price uncertainty on real economic activity. Section 6 investigates the related conjecture by Hamilton (in press) that incorporating asymmetries into joint forecasting models for real GDP growth and for the price of oil helps reduce the out-of-sample mean squared prediction error (MSPE) of real GDP growth forecasts. We conclude in Section 7.

https://doi.org/10.1017/S1365100511000186 Published online by Cambridge University Press

2. THE THEORETICAL RATIONALE FOR ASYMMETRIC RESPONSES OF REAL OUTPUT

Standard theoretical models of the transmission of oil price shocks have focused on the implications of fluctuations in the price of imported crude oil. The most immediate effect of an unexpected increase in the real price of imported crude oil is a reduction in the purchasing power of domestic households, as income is being transferred abroad. It is important for the argument that it is the price of imported crude oil that increases, because an increase in the price of domestically produced crude oil by itself would merely cause a redistribution of income rather than a reduction of domestic income in the aggregate. This direct effect of an increase in the real price of oil imports is symmetric in oil price increases and decreases. An unexpected increase in the real price of oil will cause aggregate income to fall to the same extent that an unexpected decline in the real price of oil of the same magnitude will cause aggregate income to increase.

The rationale for asymmetric responses of real output to oil price innovations hinges on the existence of additional indirect effects of unexpected changes in the real price of oil. There are three main explanations of such effects in the literature. First, it has been stressed that oil price shocks are relative price shocks, which can be viewed as allocative disturbances that cause sectoral shifts throughout the economy [see, e.g., Hamilton (1988)]. For example, the case has been made that reduced expenditures on energy-intensive durables (such as automobiles) in response to unexpectedly high real oil prices may cause a reallocation of capital and labor away from the automobile sector. As the dollar value of such purchases may be large relative to the value of the energy they use, even relatively small changes in the relative price of oil can have potentially large effects on demand. A similar reallocation may occur within the same sector, as consumers switch toward more energy-efficient durables. If capital and labor are sector-specific or product-specific and cannot be moved easily to new uses, these intersectoral and intrasectoral reallocations will cause labor and capital to be unemployed, resulting in cutbacks in real output and employment that go beyond the changes in households' purchasing power triggered by unexpectedly high oil prices. The same effect may arise if unemployed workers simply wait for conditions in their sector to improve.

The reallocation effect arises every time the relative price of oil changes unexpectedly, regardless of the direction of the oil price change. In the case of an unexpected real oil price increase, the reallocation effect reinforces the recessionary effects of the loss of purchasing power, allowing the model to generate a much larger recession than in standard linear models. In the case of an unexpected real oil price decline, the reallocation effect partially offsets the increased expenditures driven by the gains in purchasing power, causing a smaller economic expansion than implied by a linear model. This means that in the presence of a reallocation effect, the responses of real output are necessarily asymmetric in unanticipated oil price increases and unanticipated oil price decreases.

https://doi.org/10.1017/S1365100511000186 Published online by Cambridge University Press

The quantitative importance of this channel depends on the extent of expenditure switching in response to real oil price shocks and on how pervasive frictions in capital and labor markets are. There is general agreement that the domestic automobile sector is most susceptible to the reallocation effect. For example, Edelstein and Kilian (2009) have suggested that the magnitude of the reallocation effect depends primarily on the size of the domestic automobile industry (as measured by shares in employment and real output), as well as on the extent to which households substitute imported cars for domestic cars. A more extensive study of the U.S. automobile sector is provided in Ramey and Vine (2010). Related evidence for the 2007/2008 recession has also been presented in Hamilton (2009).

A second explanation of asymmetric response functions focuses on the effects on investment decisions of uncertainty about the future price of oil. If the cash flow from an irreversible investment depends on the real price of oil, real options theory implies that, all else equal, increased uncertainty about the real price of oil prompts firms to delay investments, causing investment expenditures to drop to the extent that unexpected changes in the real price of oil are associated with increased uncertainty about the future real price of oil [see, e.g., Bernanke (1983)]. Uncertainty in practice is measured by the expected volatility of the real price of oil over the relevant investment horizon. Exactly the same reasoning applies to purchases of energy-intensive consumer durables such as cars. Because any unexpected change in the real price of oil may be associated with higher expected volatility, whether the real price of oil goes up or down, this uncertainty effect may amplify the effects of unexpected oil price increases and offset the effects of unexpected oil price declines, much like the reallocation effect, resulting in asymmetric responses of real output. The quantitative importance of this channel depends on how important the real price of oil is for investment and durables purchase decisions and on the share of such expenditures in aggregate spending. For example, it seems intuitive that uncertainty about the real price of oil would be important for decisions about oil drilling in Texas, but less obvious that it would be quite as important for other sectors of the economy, such as textile production or information technology.[1]

A closely related argument has been made in Edelstein and Kilian (2009), who observed that increased uncertainty about the prospects of staying employed in the wake of unexpected changes in the real price of oil could cause an increase in precautionary savings (or equivalently a reduction in consumer expenditures). In this interpretation, uncertainty may affect not merely energy-intensive consumer durables such as cars, but also other consumer expenditures. This argument is logically distinct from the observation that households will smooth their consumption to the extent that unexpectedly higher real oil prices are associated with an increased likelihood of becoming unemployed. The key difference is that consumption-smoothing motives are symmetric in unexpected oil price increases and decreases, whereas precautionary motives are not.

https://doi.org/10.1017/S1365100511000186 Published online by Cambridge University Press

As a third explanation, it has been suggested that the response of the Federal Reserve to oil price shocks is responsible for the depth of the recessions following positive oil price shocks [see, e.g., Bernanke et al. (1997)]. The premise is that the Federal Reserve responds to incipient or actual inflationary pressures associated with unexpected real oil price increases by raising the interest rate, thereby amplifying the economic contraction. The asymmetry arises because the Federal Reserve does not respond as vigorously to unexpected declines in the real price of oil. There is no theoretical model underlying this explanation of asymmetry, and indeed earlier studies have imposed this asymmetry hypothesis in estimation rather than testing it, with the exception of Balke et al. (2002), who concluded that monetary policy alone cannot account for the asymmetry in the responses of real output. More generally, the notion that policy makers should respond to oil price shocks has been shown to be at odds with economic theory, and the empirical evidence in support of such a link has been shown to be fragile and to suffer from identification problems and inconsistent estimates [see, e.g., Carlstrom and Fuerst (2006); Hamilton and Herrera (2004); Herrera and Pesavento (2009); Kilian and Lewis (2011); Nakov and Pescatori (2010)].

We conclude that of the three main explanations for asymmetric responses in real output, only the reallocation effect and the uncertainty effect are firmly grounded in economic theory. Both models imply that an unexpected increase in the real price of oil will cause a negative response of real output that is larger in absolute terms than the positive response of real output to an unexpected decline of the same magnitude in the real price of oil. At issue is not so much whether these theoretical models are correct, but how quantitatively important the asymmetry implied by these models is for the response of U.S. aggregate real economic activity.

Broadly speaking, there have been two strategies for evaluating these economic models. One strategy, exemplified by Kilian and Vigfusson (in press), has been to specify an econometric model that can capture asymmetric responses to positive and negative oil price innovations, building on Mork (1989). This approach is not derived from any specific economic model, but may be used to provide a semiparametric approximation to potentially asymmetric responses generated by a variety of models. The other strategy has been to model one specific source of asymmetric responses. A good example is Elder and Serletis (2010), who focus on the effect of uncertainty on real output at the expense of abstracting from other asymmetric transmission channels. Obviously, the latter approach allows the use of more parametric structure, which helps improve the precision of the estimates and the power of tests of the symmetry null. On the other hand, it is not clear how to interpret these estimates, if there are multiple sources of asymmetric responses, and the results may be sensitive to the specific parameterization of oil price uncertainty.

In addition, a number of authors have suggested alternative asymmetric model specifications that combine asymmetries with additional nonlinearities. The best known example is the net oil price increase specification of Hamilton (1996,

https://doi.org/10.1017/S1365100511000186 Published online by Cambridge University Press

2003, 2009, in press). These models were introduced in recognition of the fact that conventional asymmetric models building on the distinction between oil price increases and oil price decreases do not appear to fit the U.S. data. The net oil price increase specification has been motivated by purely behavioral arguments rather than economic theory. A joint test of the asymmetries and the additional nonlinearities implicit in the approach proposed by Hamilton may be conducted based on an alternative version of the test proposed by Kilian and Vigfusson (in press). The approach of Elder and Serletis (2010), in contrast, is not designed to evaluate the econometric model proposed by Hamilton.

3. MODEL SPECIFICATION ISSUES

A number of model specification issues arise in testing the null hypothesis of symmetric response functions and in quantifying the degree of asymmetry of the estimated response functions, regardless of the modeling approach. Four modeling choices in particular explain most of the differences in the empirical results in the literature.

3.1. Which Price of Oil to Use?

Empirical results, first of all, can be sensitive to the choice of oil price measure. Leading candidates for the oil price series include the price of West Texas Intermediate crude oil, the U.S producer price of crude oil, and the U.S. refiners' acquisition cost available for imported crude oil, for domestic crude oil, and for a composite of domestic and imported crude oil. There is no general consensus on which price of oil to use. Hamilton (in press), for example, makes the case that, for the purpose of testing for asymmetries in the transmission of oil price shocks to U.S. real GDP, the U.S. producer price of crude oil is a better proxy than the refiners' acquisition cost for imported crude oil because the producer price is more highly correlated with the price of gasoline. This argument has some merit, but, taking it to its logical conclusion, we should not be using the price of oil at all, but rather the retail price of gasoline, which itself is a good proxy for the retail price of energy faced by consumers and firms [see Edelstein and Kilian (2009)]. This is indeed the approach taken in the empirical work of Edelstein and Kilian (2009), in Ramey and Vine (2010), and in some of the results shown in Hamilton (2009).

On the other hand, traditionally, oil price shocks have been associated with events in the global oil market. The usual view has been that these foreign shocks were the cause of domestic economic declines. As discussed in Section 2, it is essential for the economic reasoning underlying the baseline linear model of the transmission of oil price shocks that we focus on the price of imported crude oil. After all, typical macroeconomic models of the transmission of oil price shocks are specified in terms of the price of imported crude oil, not the retail price of gasoline. This line of reasoning suggests that taking the price of imported crude oil as the starting point, as in Kilian and Vigfusson (in press), also has merits.[2]

https://doi.org/10.1017/S1365100511000186 Published online by Cambridge University Press

To further complicate matters, Mork (1989) cautions against the use of the U.S. producer price of oil favored by Hamilton for a different reason. His concern is that this oil price has been subject to government regulation and may not be representative of true market prices. Mork (1989, p. 741) observes that "during the price controls of the 1970s, this index is misleading because it reflects only the controlled prices of domestically produced oil. However, since the price control system closely resembled a combined tax/subsidy scheme for domestic and imported crude oil, the marginal cost of crude to U.S. refiners can be approximated by the composite (for domestic and imported) refiner acquisition cost (RAC) for crude oil."

Thus, there is unlikely to be one price of oil that is the right choice for all purposes. We certainly agree with the sentiment in Hamilton (in press) that it is worthwhile to assess the sensitivity of the test results with respect to alternative oil price measures. To the extent that different oil price measures yield similar test results, our confidence in these results would increase. If evidence of asymmetry were to vanish with minor changes in the data, on the other hand, caution would be called for.

3.2. Which Sample Period to Use?

Another important modeling choice is the sample period. Empirical models of the response of U.S. real output to oil price shocks have been fit on a number of different sample periods with different results. Understanding which specifications are appropriate and which are not therefore is important. One possibility is to base the model in question on data for the nominal WTI price of crude oil prior to 1973. This price series is identical to the U.S. producer price index for the same period. It is well known that the nominal price of oil was adjusted only at discrete intervals during that period. This pattern is the result of government regulation. Because the nominal oil price data are generated by a discrete–continuous choice model, conventional dynamic regression models are not appropriate for constructing the responses of real output to oil price shocks during the pre-1973 period.

This problem may be ameliorated by deflating the nominal price of oil, which renders the pre-1973 oil price data continuous and more amenable to VAR analysis. Additional problems arise, however, when oil price data generated by a discrete–continuous choice process are combined with oil price data from the post-1973 era that are fully continuous. This approach is obviously inadvisable in dealing with nominal oil price data because these data cannot be represented as a standard time series process for the pre-1973 subsample and are not homogeneous over time. Perhaps less obviously, the approach of combining pre- and post-1973 data when dealing with dynamic regressions involving the real price of oil is equally unappealing. The problem is not only that the dynamic process governing the real price of oil is not homogeneous over time, but also that the nature and speed of the feedback from the real price of oil to U.S. macroeconomic aggregates differs by construction. For example, it can be shown that the responses to real oil price

https://doi.org/10.1017/S1365100511000186 Published online by Cambridge University Press

innovations are systematically different in pre- and post-1973 data. In particular, the real output responses tend to be systematically larger in pre-1973 data. The same instability arises in the predictive regressions commonly used to test for lagged nonlinear feedback from the real of price of oil to real GDP growth. The *p*-value for the null hypothesis that there is no break in 1973.Q4 in the coefficients of this predictive regression is .001. Given this evidence of instability, combining pre- and post-1973 real oil price data is not a valid option. Regression estimates of the relationship between the real price of oil and domestic macroeconomic aggregates obtained from the entire postwar period are not informative about the strength of these relationships in post-1973 data. It is therefore essential that we restrict the analysis to the post-1973 period in assessing the evidence for asymmetric real output responses. This is one point where our views differ from those of Hamilton (in press), who favors extending the sample back to 1949.

This point is not just academic. As the analysis in Herrera et al. (in press) illustrates, the evidence for asymmetries using the full sample appears to be driven in large part by the inclusion of pre-1973 data. Herrera et al. show that there is much less evidence for asymmetric responses of aggregate industrial production based on post-1973 data, consistent with the findings for aggregate real GDP in Kilian and Vigfusson (in press) based on post-1973 data. It is, of course, possible that this difference in results merely reflects the reduction in power from working with a shorter sample, but this does not mean that using pre-1973 data is a valid option. It is equally possible that this difference in results reflects the structural change in the underlying time series process in 1973. Either way, the evidence for the post-1973 period is the only evidence we have to go by, and earlier results in the literature based on longer time series have to be viewed with caution.

3.3. Nominal Price of Oil versus Real Price of Oil

Another key difference between existing studies relates to whether the price of oil is specified in real or in nominal terms. One argument sometimes made in support of computing the responses of real output to nominal oil price shocks is that in the pre-1973 period the nominal price of oil often remained frozen for extended periods, so that fluctuations in the real price simply reflect inflation adjustments that are endogenous to the U.S. economy. This means that innovations in the real price of oil during the pre-1973 period are not necessarily indicative of exogenous shocks in the oil market. This argument is correct, but does not justify the use of the nominal price of oil for constructing impulse responses for pre-1973 data, because the standard dynamic regression models from which these responses are computed are not appropriate for discrete–continuous oil price data, as discussed previously. More importantly, it does not justify the use of the nominal price of oil for the post-1973 period. After 1973, one would expect inflation innovations to be immediately reflected in the nominal price of oil to the extent that the nominal price of oil is free to adjust to inflation pressures. As discussed, this conclusion

https://doi.org/10.1017/S1365100511000186 Published online by Cambridge University Press

is more likely to apply to some oil price measures than to others because of the continued regulation of the domestic price of oil in the United States. In the absence of regulation, a positive U.S. monetary disturbance, for example, would be expected to raise the nominal dollar price of oil and U.S. consumer prices to the same extent, leaving the real price of oil unaffected [see Alquist et al. (2010); Gillman and Nakov (2009)]. This argument would hold even if the nominal price of imported crude oil were set by OPEC. Thus, there is good reason to expect innovations in the real price of oil in the post-1973 period to reflect real demand and real supply shocks in the crude oil market.

The focus on real oil price innovations also makes sense because theoretical models that imply asymmetries in the transmission of oil price shocks are expressed in terms of the real price of oil, as discussed in Section 2. This is why Kilian and Vigfusson (in press) specified their model in terms of the real price of oil, as have many other studies including Mork (1989), Lee et al. (1995), Elder and Serletis (2010), and Herrera et al. (in press). Hamilton (in press) makes the counterargument that it is conceivable that consumers of refined oil products choose to respond to changes in the nominal price of oil rather than the real price of oil, perhaps because the nominal price of oil is more visible. There is no direct empirical evidence in favor of this behavioral argument.[3] Rather, the case for this specification, if there is one, has to be based on the fit of models estimated at the aggregate level, as in Herrera et al. (in press), for example, or on the predictive success of such models, which explains the emphasis Hamilton (in press) puts on studying predictive relationships in the data. We will return to this question in Section 6.

3.4. Nonlinearity versus Asymmetry

Even proponents of using the nominal price in empirical models of the transmission of oil price shocks have concluded that there is no stable dynamic relationship between percentage changes in the nominal price of oil and in U.S. macroeconomic aggregates. There is evidence from in-sample fitting exercises, however, of a predictive relationship between suitable nonlinear transformations of the nominal price of oil and U.S. real output. The most successful of these transformations is the net oil price increase measure of Hamilton (2003). Let s_t denote the nominal price of oil in logs and Δ the difference operator. Then the three-year nominal net oil price increase is defined as $\Delta s_t^{+,\text{net,3yr}} \equiv \max[0, s_t - s_t^*]$, where s_t^* is the highest oil price in the preceding three years. This transformation involves three distinct ideas. One is that consumers in oil-importing economies respond to increases in the price of oil only if the increase is large relative to the recent past. If correct, the same logic by construction should apply to decreases in the price of oil, suggesting a net change transformation that is symmetric in increases and decreases. The second idea implicit in Hamilton's definition is that consumers do not respond to net decreases in the price of oil, allowing us to omit the net decreases from the model. In other words, consumers respond asymmetrically to

https://doi.org/10.1017/S1365100511000186 Published online by Cambridge University Press

net oil price increases and net oil price decreases, and they do so in a very specific fashion. The third idea is that what matters for the transmission of oil price shocks is the nominal rather than the real price of oil.

It is important to stress that the net oil price increase is not tightly linked to any of the theoretical models discussed in Section 2, which imply the existence of an asymmetry in the response of real output to real oil price increases and decreases. First, there is general agreement that asymmetric model specifications that do not involve the additional nonlinearity proposed by Hamilton (1996, 2003) such as Mork's (1989) model based on $\Delta r_t^+ \equiv \Delta r_t I(\Delta r_t > 0)$, where $I(\cdot)$ denotes the indicator function and r_t the log of the real price of oil, are not supported by the data.[4] The additional nonlinear structure embodied in Hamilton's (2003) net oil price increase measure, however, is not a feature of the theoretical models discussed in Section 2. It is based on purely behavioral arguments. Second, the theoretical models of Section 2 do not imply that oil price decreases should receive zero weight. They only imply that they should receive less weight than increases of the same magnitude. Third, the use of the nominal price of oil is not consistent with economic theory and requires a behavioral motivation. Nevertheless, Hamilton's nominal net oil price increase variable has become one of the leading specifications in the literature on predictive relationships between the price of oil and the U.S. economy. Hamilton (in press) interprets this relationship as capturing nonlinear changes in consumer sentiment in response to nominal oil price increases.

The behavioral rationale for the net oil price increase measure, of course, applies equally a priori to the nominal price of oil and the real price of oil. Whereas Hamilton (2003) applied this transformation to the nominal price of oil, several other studies have recently explored models that apply the same transformation to the real price of oil [see, e.g., Alquist et al. (2010); Herrera et al. (in press); Kilian and Vigfusson (in press)]. In that case, we define analogously $\Delta r_t^{+,\text{net,3yr}} \equiv \max[0, r_t - r_t^*]$.

4. GENERAL TESTS OF THE HYPOTHESIS OF SYMMETRIC RESPONSE FUNCTIONS

An important question is how to test for the symmetry of the response functions of real output. Kilian and Vigfusson (in press) proposed a new and conceptually simple impulse response–based Wald test for this purpose, which distinguishes between oil price shocks of different magnitudes and encompasses all the channels of transmission discussed in Section 2. This test is built on the observation that under the null of a symmetric response function, the vector of impulse responses to a positive oil price innovation should be equal to the vector of impulse responses to a negative oil price innovation except for its sign, such that the sum of these vectors is equal to a vector of zeros. Here we focus on a version of this test designed for models involving the three-year net oil price increase.

https://doi.org/10.1017/S1365100511000186 Published online by Cambridge University Press

4.1. Slope-Based Tests versus Impulse–Response Based Tests

Kilian and Vigfusson's approach differs from the traditional approach of conducting a Wald test of $H_0 : \delta_i = 0,\ i = 1, \ldots, 4$, based on the regression

$$\Delta y_t = \alpha + \sum_{i=1}^{4} \beta_i \Delta s_{t-i} + \sum_{i=1}^{4} \gamma_i \Delta y_{t-i} + \sum_{i=1}^{4} \delta_i \Delta s_{t-i}^{\text{net},+,3\text{yr}} + u_t, \qquad (1)$$

where y_t denotes U.S. real GDP, or an equivalent regression for the real price of oil. Kilian and Vigfusson's central point was not that this OLS slope-based test is not powerful enough (indeed that test tends to reject the null of symmetric slopes, so lack of power is not an obvious concern). Rather their point was that given the nonlinearity of the impulse response functions in the slope parameters neither a rejection of the null of symmetric slopes nor a nonrejection of the null of symmetric slopes tells us whether there is a nontrivial degree of asymmetry in the implied impulse response functions or not. Kilian and Vigfusson further stressed that the degree of asymmetry of the response functions is in general dependent on the magnitude of the unexpected change in the price of oil. They showed that it is easily possible, in particular, for a linear symmetric model to provide a good approximation to the response functions except for extremely large oil price innovations. Slope-based tests do not distinguish between shocks of different magnitude and hence are inherently unsuitable for evaluating the degree of asymmetry of the response functions.

Motivated by these limitations of slope-based tests, Kilian and Vigfusson (in press) proposed an alternative test of the null hypothesis of symmetric response functions based on the symmetry of the structural impulse response functions rather than the symmetry of the reduced-form slope parameters. They demonstrate that this impulse-response-based test may reject the null of symmetric response functions, when the null of symmetric slopes is not rejected for the same data set; it also may fail to reject the null of symmetric response functions, when the null of symmetric slopes is rejected.

4.2. Alternative Tests of the Null of Symmetric Slopes

Although slope-based tests are not designed to assess the degree of asymmetry of the responses of real output to oil price innovations, they have been used extensively as tests of the null hypothesis that the data-generating process is linear and symmetric. It is therefore useful to understand the trade-offs between alternative types of slope-based tests that one might implement. The most common test in the literature is due to Balke et al. (2002), who proposed testing $H_0 : \delta_i = 0,\ i = 1, \ldots, 4$, based on model (1), as discussed earlier. Note that this model includes regressors not included in the predictive model proposed by Hamilton

https://doi.org/10.1017/S1365100511000186 Published online by Cambridge University Press

(2003, in press),

$$\Delta y_t = \alpha + \sum_{i=1}^{4} \gamma_i \Delta y_{t-i} + \sum_{i=1}^{4} \delta_i \Delta s_{t-i}^{\text{net},+,3\text{yr}} + u_t, \qquad (1')$$

which imposes the additional restriction $\beta_i = 0,\ \forall i > 0$, in estimation. It may seem that we could alternatively use model (1′) instead of model (1) to test $H_0 : \delta_i = 0,\ i = 1, \ldots, 4$. This test constitutes the second slope-based test to be included in our comparison.

Yet another slope-based test has recently been proposed by Kilian and Vigfusson (in press). Their analysis shows that if we start with a structural representation of the data-generating process for real GDP and the price of oil motivated by exactly the same economic reasoning that led earlier researchers to specify the predictive regression (1), then the dynamic relationship between the price of oil and real output by construction will include contemporaneous oil price variables on the right-hand side. This means that under the null of a symmetric structural model an additional restriction can be imposed in testing. This insight suggests a third slope-based test that involves fitting

$$\Delta y_t = \alpha + \sum_{i=0}^{4} \beta_i \Delta s_{t-i} + \sum_{i=1}^{4} \gamma_i \Delta y_{t-i} + \sum_{i=0}^{4} \delta_i \Delta s_{t-i}^{\text{net},+,3\text{yr}} + u_t \qquad (1'')$$

and testing $H_0 : \delta_i = 0,\ i = 0, \ldots, 4.$ rather than $H_0 : \delta_i = 0,\ i = 1, \ldots, 4$, where we have relied on the nominal price of oil to maintain notational consistency. Note that model (1″) follows directly from the unrestricted fully specified structural model discussed in Kilian and Vigfusson (in press) and that models (1) and (1′) can be derived from special cases of this structural model after imposing additional restrictions.[5]

It can be shown that tests based on model (1′) will suffer from excessive size if the maintained assumption $\beta_i = 0,\ \forall i > 0$, is violated. This size distortion persists asymptotically because censoring the oil price regressor renders the OLS estimator of δ_i inconsistent except in the unlikely case that $\beta_i = 0,\ \forall i > 0$. Hence, if any $\beta_i \neq 0$, the tests based on model (1) or (1″) will have more accurate size than the test based on model (1′). Only if the restriction holds that there is no feedback at all from oil price declines to real GDP in population (and that there is no feedback from oil price increases that do not exceed recent peak levels) does the test based on model (1′) have accurate size and more power than alternative slope-based tests. Hence, it is possible for a test based on model (1′) to generate stronger rejections of the null of symmetric slopes, but such evidence is equally compatible with a size distortion arising from $\beta_i \neq 0$ for some i and with improved power arising from $\beta_i = 0,\ \forall i > 0$. We cannot infer which test is more appropriate from the empirical results. We can say, however, that economic theory does not imply that $\beta_i = 0,\ \forall i > 0$. The models reviewed in Section 2 imply that the feedback from declines in the real price of oil is weaker than the feedback from increases in the

https://doi.org/10.1017/S1365100511000186 Published online by Cambridge University Press

real price of oil, not that it is zero. Virtually all asymmetric models therefore will be characterized by $\beta_i \neq 0,\ i > 0$, and even small departures from $\beta_i = 0,\ i > 0$ will render the OLS estimator of δ_i inconsistent and misleading [see Kilian and Vigfusson (in press)]. Nor can this issue be resolved by statistical testing. The fact that a statistical test of $H_0 : \beta_i = 0,\ i > 0$ typically does not reject the null of zero slopes does not imply that this null is actually true or that it is safe to impose that restriction in estimation. Hence, in cases where a test based on (1′) provides stronger rejections than the other tests on the same data set, it is unclear whether this outcome reflects size distortions or higher power.

As to the choice between tests based on models (1) and (1″), it can be shown that the slope-based test of the joint null of linearity and symmetry based on model (1″) may have higher power against departures from $H_0 : \delta_i = 0,\ i = 0, \ldots, 4$ than the conventional test based on model (1), which omits the contemporaneous regressors. This point has been illustrated in the context of Mork's (1989) asymmetric specification in Kilian and Vigfusson (in press). Depending on the values of the parameters of the omitted contemporaneous regressors, however, it is also possible that the test based on model (1) may have higher power than the test based on model (1″). The power ranking will differ in general, depending on the data set and model specification. For the three-year net oil price increase specification, for example, our results indicate that the test based on model (1) has slightly higher power, given comparable size.

As expected given this evidence, the differences between using the slope-based tests for model (1) and for model (1″) tend to be small for the three-year net increase specification considered. In this context, Hamilton (in press) mistakenly claims that the slope-based test based on model (1″), as implemented in Kilian and Vigfusson (in press) for the real price of oil, fails to reject the null of symmetric slopes for real GDP. Hamilton suggests that this test result is at odds with the rest of the literature and therefore likely to be wrong. Actually, however, Kilian and Vigfusson found results very similar to those in the previous literature. In particular, they found no evidence that the linear model is rejected in favor of the asymmetric model of Mork (1989), consistent with the substantive findings in Hamilton (1996) and Hooker (1996). Moreover, their test rejected the linear model at the 5% level in favor of a model including the three-year net oil price increase, much as Balke et al. (2002) and Hamilton (2003, in press) did. In fact, Kilian and Vigfusson use this example to illustrate the differences between standard slope-based tests (which reject symmetry in the slopes) and impulse response–based tests (which do not reject symmetry in the response functions). Thus, Hamilton's claim that Kilian and Vigfusson presented evidence based on slope-based tests that the predictive relationship is linear (and his conclusion that this alleged difference in results is driven by a number of changes in the model specification relative to previous studies) is not supported by the facts.

This does not mean, of course, that modeling choices such as the choice of sample period or whether the price of oil is expressed in real or in nominal terms cannot in some cases affect the degree of statistical significance. We have already

https://doi.org/10.1017/S1365100511000186 Published online by Cambridge University Press

discussed each of these modeling choices in detail. What Hamilton's (in press) results show is that the rejection of the null of symmetric slopes reported in Kilian and Vigfusson (in press) is robust to a variety of alternative model specifications. We agree with Hamilton that given some alternative modeling choices, one can reject the null of symmetric slopes at significance levels even higher than the 5% level reported in Kilian and Vigfusson. The rejection decision is the same across all specifications, however, making this distinction moot.[6] More importantly, all of these results are tangential to the question at hand, because the results of slope-based tests are not informative about the degree of asymmetry in the response functions of real GDP, which is why Kilian and Vigfusson (in press) caution against the use of any slope-based test, whether the traditional test or the modified test. They are relevant only to the separate question posed by Hamilton (in press) of whether there is nonlinear predictability from the price of oil to real GDP growth. There is no way of inferring "cause" and "effect" from such predictive correlations, which is why Kilian and Vigfusson focus on structural impulse response analysis.

4.3. How to Model and Compute Nonlinear Structural Impulse Responses

Hamilton (in press) does not disagree with the conclusion in Kilian and Vigfusson (in press) that earlier estimates of the responses of real output to oil price shocks were invalid. In fact, he fully agrees with Kilian and Vigfusson's point that fundamental changes are needed in the way that potentially nonlinear models of impulse responses ought to be specified, estimated, and tested.

There are four distinct contributions in Kilian and Vigfusson (in press) that must be viewed in conjunction. First, Kilian and Vigfusson establish that impulse-response estimates from VAR models involving censored oil price variables are inconsistent even when equation (1) is correctly specified. Specifically, they demonstrate that asymmetric models of the transmission of oil price shocks cannot be represented as censored oil price VAR models and are fundamentally misspecified whether the data-generating process is symmetric or asymmetric. This misspecification renders the parameter estimates inconsistent and inference invalid. Second, they show that standard approaches to the construction of structural impulse responses used in this literature are invalid, even when applied to correctly specified models. Instead, Kilian and Vigfusson propose a modification of the procedure discussed in Koop et al. (1996). Third, Kilian and Vigfusson demonstrate that standard slope-based tests for asymmetry based on single-equation models are neither necessary nor sufficient for judging the degree of asymmetry in the structural response functions, which is the question of ultimate interest to users of these models. Kilian and Vigfusson proposed a direct test of the latter hypothesis which requires the model to be correctly specified and the nonlinear responses to be correctly simulated, as discussed in points 1 and 2. Fourth, using this test, they showed empirically that there is no statistically significant evidence of asymmetry in the response functions for U.S. real GDP using data for 1973.Q2–2007.Q4.[7]

https://doi.org/10.1017/S1365100511000186 Published online by Cambridge University Press

TABLE 1. *p*-Values of joint tests of the null of symmetric response functions: real GDP responses to oil price innovations at horizons up to one year

		Impulse response–based test	
		1-std dev shock	2-std dev shock
PPI	Real	0.96	0.04
	Nominal	0.96	0.04
Import RAC	Real	0.98	0.06
	Nominal	0.72	0.04

Notes: Based on model (3) with four lags. PPI stands for the U.S. producer price index for crude oil and Import RAC for the U.S. refiners' acquisition cost for imported crude oil. Using data ending in 2007.Q4, there is no evidence against the null of symmetric responses for any of these tests.

4.4. The Evidence from Impulse Response–Based Tests

With these clarifications in mind, it is time to focus on the empirical evidence. Table 1 updates the results in Kilian and Vigfusson (in press) for U.S. real GDP growth. There are a number of differences in the specification. First, the sample period is set to 1974.Q1–2009.Q4. This takes account of the structural break in the predictive relationship in 1973.IV and facilitates a clean comparison of alternative oil price series. Second, for expository purposes we focus on the U.S. producer price index favored by Hamilton and the U.S. refiners' acquisition cost for imported crude oil used in Kilian and Vigfusson (in press). Third, we show results for both the nominal and real prices of oil. Fourth, we focus on a model with four lags rather than the six lags used in Kilian and Vigfusson (in press). The latter two changes are intended to facilitate the comparison with Hamilton's preferred model specification.

Table 1 shows that there is no statistically significant evidence against the null of symmetric responses to unexpected oil price increases and unexpected oil price decreases for shocks of typical magnitude. For the real PPI, for example, the *p*-value of the impulse response–based test is 0.90. Qualitatively similar results are obtained whether the price of oil is specified in real or in nominal terms. The benchmark of one–standard deviation shocks is representative of about two-thirds of all oil price innovations that occurred historically. Only when we focus on much larger two–standard deviation shocks is there statistically significant evidence of asymmetry with a *p*-value of 0.04 for both the real and the nominal PPI. Shocks of this magnitude or larger have historically occurred with a probability of only 5%.[8]

The results in Table 1 for two–standard deviation shocks differ from those reported in Kilian and Vigfusson (in press) for a sample period ending in 2007.Q4. That study found no evidence against the null of symmetric responses for one– as well as two–standard deviation shocks. It can be shown that the difference in

https://doi.org/10.1017/S1365100511000186 Published online by Cambridge University Press

results is driven by the extended sample period. As will be discussed in more detail later, there is reason to be cautious in interpreting the post-2007.Q4 results, however. If we exclude data after 2007.Q4 from the estimation period, all evidence of asymmetric responses vanishes, even in response to two–standard deviation shocks, suggesting that the model may be confounding the effects of the financial crisis with those of net oil price increases. We will return to this point.

The analysis for U.S. real GDP growth in Kilian and Vigfusson (in press) has recently been complemented by additional evidence based on the same impulse response–based test applied to U.S. industrial production. Herrera et al. (in press) investigate not only aggregate data, but industrial production data disaggregated by sector. Their sample period roughly corresponds to that underlying Table 1. Their baseline model utilizes the real price of oil.[9] Herrera et al.'s analysis shows considerable evidence of asymmetric responses even after accounting for data mining when the model is estimated on the full sample including pre-1973 data. That evidence, however, weakens considerably when the pre-1973 data are discarded, as we did in Table 1. The post-1973 results in Herrera et al. for aggregate monthly industrial production are remarkably similar to those in Table 1. In both cases, the impulse response–based test suggests that there is statistically significant evidence of asymmetries only in response to very large oil price innovations, but not in response to oil price innovations of more typical magnitude. Their *p*-values for responses up to one year are 0.80 and 0.09, compared with 0.90 and 0.04 in Table 1. Thus, the additional evidence for aggregate industrial production in Herrera et al. (in press) is broadly consistent with the results in Table 1 for aggregate real GDP growth.

One important difference is that Herrera et al.'s analysis shows in which sectors the asymmetry originates. Their results illustrate that the impulse response–based test is powerful enough to detect departures from the null of symmetric responses to one-standard deviation shocks within a year. In particular, for oil price innovations of typical magnitude, there is no evidence of asymmetric responses in aggregate industrial production based on $\Delta r_t^{+,\text{net},3\text{yr}}$, but there is statistically significant evidence (even after accounting for data mining) against the symmetry null at the disaggregate level for sectors such as *chemicals*, *transit equipment*, *petroleum and coal*, *plastics and rubber*, *primary metal*, and *machinery*, although interestingly not for *motor vehicles*. This evidence confirms that there can be considerable heterogeneity in the degree of asymmetry across sectors. It suggests that the failure to reject the null of symmetric responses at the aggregate level may simply mean that asymmetries at the sectoral level are not strong enough to be detected in the aggregate data.

For oil price innovations corresponding to two–standard deviation shocks, Herrera et al. (in press) find somewhat stronger evidence against the null of symmetric responses, but at the aggregate level the evidence remains mixed, with at best a marginal rejection at the 10% level at a horizon of one year (not accounting for data mining). At the disaggregate level, significant rejections are found in sectors such as *plastics and rubber*, *chemicals*, and *transit equipment*,

https://doi.org/10.1017/S1365100511000186 Published online by Cambridge University Press

which is to be expected, as these industries are known to be energy-intensive, but again not for *motor vehicles*. None of these rejections for two–standard deviation shocks is significant after controlling for data mining, however, which may reflect a loss of power, as large oil price innovations are rare in the data and the resulting responses are less precisely estimated.

Overall, we conclude that the evidence in favor of asymmetries is mixed. For a one–standard deviation shock, the response functions of U.S. real economic activity appear to be well approximated by those of a linear symmetric VAR model. This means that the linear symmetric model can be expected to provide a good approximation in modeling the responses of real output to innovations in the real price of oil in most situations. For very large innovations in the price of oil, there is evidence that the aggregate responses are asymmetric in that the expansion triggered by negative oil price innovations is smaller than is predicted by a linear model. An interesting question for future research will be to determine to what extent that asymmetry is associated with specific expenditure components, building on Edelstein and Kilian (2007, 2009).

It should also be stressed that the nonlinearity that the impulse response–based test detects in response to very large oil price innovations is based on the net increase measure developed by Hamilton (1996, 2003) using purely behavioral arguments. It therefore does not provide any support for the reallocation effect or the uncertainty effect discussed in Section 2. In fact, if we replaced the net increase measure with a more conventional measure of oil price increases—as defined in Mork (1989)—all rejections of the symmetry null hypothesis would vanish. Thus, the economic nature of the apparently asymmetric response to very large oil price innovations remains to be investigated. One way of approaching this question would be to focus on consumer sentiment data, as in Edelstein and Kilian (2009).

Finally, we have to keep in mind that these test results are tentative only. There are indications that the evidence in favor of asymmetries in response to large oil price innovations may be spurious. The sensitivity of the test results to the inclusion of data from the 2008 financial crisis suggests that the evidence in favor of asymmetric responses could reflect overfitting resulting from the use of a quadratic loss function in conjunction with a relatively short sample. The concern is that the coincidence of a financial crisis following a large net oil price increase may cause the model to attribute the effects of the financial crisis on real GDP to the earlier net oil price increases. Given the unusual decline in real GDP during this episode, the ability to fit this one episode greatly improves the overall fit of the net increase model. Longer samples will be needed to resolve this question.

5. TESTS OF THE UNCERTAINTY EFFECT

Although the uncertainty effect has played a prominent role in discussions of asymmetric responses of real output for two decades, the first study to provide a fully specified model of this transmission mechanism is Elder and Serletis (2010).

https://doi.org/10.1017/S1365100511000186 Published online by Cambridge University Press

Elder and Serletis's baseline model is a VAR(4) model for real GDP growth and the percent change in the real price of oil with GARCH-in-mean. The oil price measure is the composite U.S. refiners' acquisition cost. The model is estimated on post-1974 data. Oil prices are treated as predetermined with respect to real economic activity.[10] Rather than testing the symmetry of the response functions directly, as proposed in Kilian and Vigfusson (in press), Elder and Serletis test the null of no feedback from the one quarter–ahead conditional variance of the real price of oil in the conditional mean equation. They report that information criteria favor the model including this term. The presence of GARCH-in-mean effects implies that the response of real output in this model is asymmetric in unexpected oil price increases and decreases. It also implies that the degree of asymmetry will in general depend on the magnitude of the shock. Elder and Serletis report shocks normalized to one standard deviation of the unconditional distribution of the percent change of the real price of oil (rather than one standard deviation of the unconditional regression residual). They provide evidence that increased oil price volatility exacerbates the negative response of real economic activity to an unexpected increase in the real price of oil, whereas it dampens the positive response to an unexpected decline in the real price of oil. It is difficult to compare the impulse response estimates in Elder and Serletis with those in Kilian and Vigfusson because the magnitude of the oil price shock differs and because Elder and Serletis do not provide a formal test of the symmetry of the response functions, but if we take their point estimates at face value, there appears to be strong evidence of asymmetric responses.

One possibility is that their results for U.S. real GDP and for U.S. industrial production are stronger than the results in Kilian and Vigfusson (in press) and Herrera et al. (in press) for similar sample periods because their approach is more parametric and therefore has greater power to detect departures from linearity.[11] Even allowing for that possibility, there is reason to be skeptical of the large uncertainty effect Elder and Serletis estimate. The only study to date to provide formal empirical evidence of an uncertainty effect at the firm level is Kellogg (2010), which analyzed the investment decisions of oil companies in Texas. Kellogg showed that competitive oil companies, all else equal, significantly reduce their drilling of oil wells in response to increased expected oil price volatility. This finding makes sense given the overriding importance of the price of oil for the cash flow of oil producers. A similar argument may be plausible for purchases of automobiles, but the share of the automobile sector is relatively small in the U.S. economy, and more generally energy prices constitute a small determinant of the cash flow of investment projects. This is clearly an area that deserves further study.

An alternative explanation is that the parametric GARCH-in-mean VAR model is misspecified. One limitation of this approach, acknowledged by the authors, is that it is not clear that the conditional variance implied by the GARCH model is the appropriate measure of oil price uncertainty. One problem is that GARCH volatility is inherently backward-looking, whereas investors' expectations tend to be forward-looking. In addition, a good case can be made that the market's

https://doi.org/10.1017/S1365100511000186 Published online by Cambridge University Press

uncertainty about the future price of oil may actually fall during sharp oil price declines, making price volatility a poor measure of the uncertainty about future oil prices.[12] Finally, there is reason to believe that investors in the oil industry respond to the risk that the real price of oil will fall below a prespecified level corresponding to the marginal cost of production, rather than to oil price volatility, because upside risk in the form of higher oil prices should not prevent the investment from going ahead. Similarly, investors outside the oil sector will be primarily concerned with the risk of energy prices rising above a prespecified level. The construction of such oil price risk measures is discussed in Alquist et al. (2010). Clearly, different forms of risk will be relevant for different sectors of the economy, suggesting that generic oil price volatility measures may not be appropriate.

A second and more fundamental concern is that the relevant measure of oil price volatility in the theoretical models described by Bernanke (1983) and Pindyck (1991) is not the one quarter–ahead volatility of the real price of oil used by Elder and Serletis. Rather, the relevant volatility measure is the volatility of the real price of oil at horizons relevant to purchase and investment decisions, which is typically measured in years or even decades rather than days or months, making all standard measures of short-term price volatility inappropriate. Measuring the volatility of the real price of oil at longer forecast horizons is inherently difficult, given how short the available time series are, and indeed researchers in practice have typically asserted rather than measured these shifts in longer-horizon real price volatility. Elder and Serletis (2010) effectively treat measures of the short-horizon oil price volatility as a proxy for longer-horizon volatility. One concern with using standard monthly or quarterly GARCH models to quantify changes in the longer-term expected volatility of the real price of oil is that GARCH forecasts of the conditional variance quickly revert to their time-invariant unconditional expectation, as the forecasting horizon increases. If the economically relevant longer-term volatility is constant by construction, it cannot explain variation in real activity over time. If investment projects take time to implement before they start generating a cash flow, this weakens and possibly eliminates the feedback from current uncertainty to investment decisions.[13] This reasoning suggests that additional survey data on the distribution of long-run forecasts and beliefs about future oil prices at long horizons would be helpful in refining tests of the uncertainty hypothesis.[14]

One could indeed make the case that the GARCH-in-mean model specified by Elder and Serletis (2010) is more likely to capture the reallocation effect of Hamilton (1988) than the uncertainty effect of Bernanke (1983) in that it effectively measures volatility in the current real price of oil (indicative of relative price changes) rather than the expected price of oil at horizons relevant to investment decisions. If so, we should be able to show in more detail how oil price uncertainty affects employment flows at the sectoral, firm, and plant levels. This question has already been investigated in the literature, but only based on censored oil price VAR models of the type shown to be invalid in Kilian and Vigfusson (in press). Further studies using updated time series data and state-of-the-art methods of estimation and inference appear promising avenues for research.

https://doi.org/10.1017/S1365100511000186 Published online by Cambridge University Press

A third concern with the model specification of Elder and Serletis is that it leaves no room for other forms of nonlinearity such as the nonlinear adjustment implied by Hamilton's net increase measure. If we believe that such nonlinearities matter (or for that matter that there are other sources of asymmetries not formally modeled), then the GARCH-in-mean VAR model will be misspecified, making it difficult to determine what the estimated responses of real economic activity represent.

A final concern is how to reconcile the disaggregate results in Elder and Serletis for U.S. real consumption with those in Edelstein and Kilian (2009), who showed that a linear symmetric model does quite well in explaining the historical evolution of U.S. consumption in both 1979 and 1986. One difference is that Edelstein and Kilian (2009) allow for nonlinearities related to the evolution of the share of energy in consumption, whereas Elder and Serletis do not. Another difference is that Edelstein and Kilian focused on retail energy prices rather than the price of oil. An interesting question, considered in ongoing research, is to what extent the apparent asymmetry in the consumption responses in Elder and Serletis (2010) may be caused by omitted nonlinearities linked to the time-varying expenditure share on energy or by the choice of the energy price series.[15]

6. OUT-OF-SAMPLE PREDICTION OF REAL GDP GROWTH WITH ASYMMETRIC MODELS

Hamilton (in press) agrees with Kilian and Vigfusson as far as impulse response analysis from censored oil price VAR models is concerned, but suggests that nonlinear models such as model (1) may still be useful for out-of-sample forecasting. This is a legitimate conjecture, which we explore in some detail in this section. The fundamental difference between impulse response analysis and out-of-sample forecasting is that in the latter context parsimony may be more important than specifying the model correctly. Thus, models that are suitable for impulse-response analysis need not be suitable for out-of-sample forecasting. It is important to keep in mind that the question of whether nonlinear models improve forecast accuracy also is distinct from the question of whether there is a nonlinear predictive relationship in population, which is the question that Hamilton (2003) focused on. As is well known, the in-sample fit of nonlinear predictive models selected using tests such as the slope-based tests discussed earlier does not guarantee gains in out-of-sample forecast accuracy.

A comprehensive analysis of the forecasting accuracy of alternative nonlinear forecasting models of this type can be found in Alquist et al. (2010). Table 2 shows recursive MSPE ratios for one quarter–ahead forecasts from model (1′) relative to the AR(4) benchmark model for U.S. real GDP growth.[16] A robust finding across all specifications is that model (1′) is less accurate out of sample than the AR model for real GDP growth, except during the recent financial crisis period. This conclusion is robust to the choice of lag order, to the choice of oil price series, and to the use of nominal or real oil prices. To the extent that some of the 20 model

https://doi.org/10.1017/S1365100511000186 Published online by Cambridge University Press

TABLE 2. One quarter–ahead MSPE ratios for U.S. real GDP growth based on model (1′)

	Forecast evaluation period					
	1990.Q2–2007.Q4		1990.Q2–2010.Q2		2008.Q1–2010.Q2	
Specification	Real (1)	Nominal (2)	Real (3)	Nominal (4)	Real (5)	Nominal (6)
	MSPE relative to AR(4) benchmark					
Four lags:						
Import RAC	1.11	1.22	**0.91**	1.02	**0.50**	**0.60**
Composite RAC	1.50	1.58	1.16	1.26	**0.48**	**0.60**
Domestic RAC	1.55	1.50	1.23	1.23	**0.57**	**0.69**
PPI	1.63	1.58	1.23	1.23	**0.42**	**0.53**
WTI	1.24	1.01	1.04	**0.92**	**0.63**	**0.74**
	MSPE relative to AR(6) benchmark					
Six lags:						
Import RAC	1.40	1.49	1.10	1.18	**0.45**	**0.52**
Composite RAC	2.80	3.30	2.04	2.41	**0.47**	**0.55**
Domestic RAC	2.18	2.09	1.66	1.63	**0.57**	**0.66**
PPI	2.05	2.11	1.53	1.59	**0.43**	**0.51**
WTI	1.32	1.12	1.08	**0.97**	**0.60**	**0.67**

Notes: Reductions in the recursive MSPE relative to the benchmark are shown in boldface. The estimation period starts in 1975.Q2 and 1975.Q4, respectively, owing to the data constraints discussed earlier. All MSPE ratios are expressed relative to the corresponding AR model for real GDP growth. Model (1′) is the nonlinear forecasting model favored by Hamilton (in press). In contrast, model (1), estimated using the real import RAC, would be the one step–ahead forecasting model implied by the structural model employed in Kilian and Vigfusson (in press). For completeness, we note that the recursive MSPE ratio for the latter model with six lags is 1.10 in the evaluation period 1990.Q2–2010.Q2. Hamilton (in press) reports an MSPE ratio of 0.87 for the same model for the evaluation period 1990.Q2–2010.Q2. The difference in results is due to the use of additional pre-1974 data in Hamilton's regression. It can be shown that after the financial crisis period is excluded Hamilton's regression estimated on his data yields an MSPE ratio of 1.08 and the same regression estimated on our data yields an MSPE ratio of 1.41.

specifications considered beat the AR benchmark on the full evaluation sample, this result as well is driven entirely by the financial crisis episode. Moreover, there is no consistent pattern as to which specifications work well. This finding reinforces the concerns with overfitting we already expressed in Section 4 when discussing the in-sample fit of the model.

Further analysis in Alquist et al. (2010) shows that nonlinear models tend to produce systematically less accurate forecasts than the AR(4) benchmark model for real GDP growth through most of the evaluation period. In fact, nonlinear models have a tendency to predict major economic declines any time the net price of oil increases substantially, and in many cases those predictions proved incorrect. The ability to forecast the extreme decline of 2008/2009, however, under a quadratic loss function more than compensates for these earlier forecasting errors and accounts for the higher average forecast accuracy of the four-quarter ahead model. Another telltale sign of overfitting is that nonlinear models attribute most

https://doi.org/10.1017/S1365100511000186 Published online by Cambridge University Press

of the economic decline of 2008/2009 to the preceding oil price increase rather than to the financial crisis. Even leaving aside these obvious concerns, however, it is useful to keep in mind that even robust evidence of nonlinear forecasting success would tell us nothing about the direction of causality or the about transmission of oil price innovations.

7. CONCLUSION

The problem of how to test the null of symmetric responses of U.S. real economic activity to oil price innovations has been solved only very recently. Kilian and Vigfusson (in press) proposed a statistical test of this hypothesis based on a class of econometric models that encompasses all standard theoretical explanations of asymmetric responses or, alternatively, can be modified to test alternative explanations of asymmetries not based on economic theory. Preliminary evidence based on these models revealed no evidence against the null of symmetric response functions in U.S. real GDP data.

Additional results in this paper based on an extended data set that includes observations until 2009.Q4 showed some evidence of asymmetry in the response of real GDP to very large shocks, but none in response to shocks of typical magnitude. Herrera et al. (in press) provided related evidence for U.S. aggregate industrial production based on the same econometric approach. Their findings are broadly consistent with the results for real GDP. They found stronger evidence of asymmetric responses at the sectoral level, however, than in the aggregate data. In short, the econometric methodology developed in Kilian and Vigfusson indicates that most of the time the linear symmetric model provides a good approximation in modeling the responses of real output to innovations in the real price of oil. What evidence exists for an asymmetric response appears limited to uncommonly large innovations in the real price of oil. There are indications, however, that even this limited evidence may be spurious. If we exclude data after 2007.Q4 from the estimation period, all evidence of asymmetric responses vanishes. The sensitivity of the test result to one very unusual episode in the data may reflect the use of a quadratic loss function and a relatively short sample. Longer samples will be needed to resolve this question.

In related work, Elder and Serletis (2010) proposed an alternative econometric model that imposes additional parametric structure in an effort to isolate one of the nonlinear transmission mechanisms discussed in the literature. As we discussed, this approach has both advantages and disadvantages compared with the semiparametric approach of Kilian and Vigfusson (in press). Although the estimates in Elder and Serletis are suggestive of strong asymmetries in the transmission of oil price innovations, formal tests of the degree of asymmetry of their response-function estimates are not yet available. A natural approach would be to adapt the impulse response–based test of Kilian and Vigfusson (in press) for this purpose. For now, a fair summary of this empirical literature is that the evidence on the degree of

https://doi.org/10.1017/S1365100511000186 Published online by Cambridge University Press

asymmetry of the responses of U.S. real output may be highly dependent on what parametric structure we impose.

We also investigated the view recently expressed in Hamilton (in press) that we should focus on the problem of jointly forecasting real GDP growth and the price of oil out of sample rather than evaluating the underlying structural economic models. We showed that the forecast accuracy of nonlinear asymmetric models depends in general on the model specification, horizon, and evaluation period. We found, in particular, that the one–step ahead nonlinear forecasting model favored by Hamilton (in press) based on extensive in-sample model specification searches tends to be less accurate out of sample than a linear AR model for real GDP growth, except during the recent financial crisis period. We stressed reasons to be skeptical of the apparent forecasting success of this nonlinear model during 2008–2010.

We also noted the difficulty of rationalizing these nonlinear forecasting models based on existing economic models of the transmission of oil price shocks, and we cautioned against confusing predictive correlations with evidence of "cause" and "effect" [also see Cooley and LeRoy (1985)]. The predictive correlations studied by Hamilton (2003, in press) in particular do not shed light on the macroeconomic effects of oil price shocks. As such, we disagree with Hamilton's (in press) view that structural impulse responses are simply conditional forecasts plotted as a function of the horizon.[17] Similarly, we need to distinguish between overall variation in the price of oil and unpredictable variation in the price of oil. A common misconception in the literature is to equate a large cumulative increase in the price of oil with a large oil price innovation. We emphasized that the consequences of extraordinary events in oil markets can only be studied with the help of historical decompositions based on structural models, not by inspecting dynamic comovement in the data. There is no substitute for structural econometric modeling in quantifying possibly nonlinear effects of unexpected oil price shifts.

NOTES

1. Bernanke (1983) also suggests that the magnitude of the uncertainty effect will depend on the how certain consumers are about the persistence of changes in the real price of oil. Recent evidence in Anderson et al. (2010) based on Michigan consumer survey data, however, suggests that U.S. consumers have consistently employed a no-change forecast of the real price of oil that treats all unexpected changes in the real price of oil as permanent.

2. To illustrate this point, consider the limiting case of an autarchic economy in which all oil is produced domestically. In that case, there would be no reason to expect aggregate purchasing power to change in response to an unexpected increase in the price of oil, with the implication that a change in the price of oil of given magnitude would be equally recessionary, regardless of sign. It would no longer be the case that unexpected oil price increases are more recessionary than unexpected oil price declines of the same magnitude.

3. Direct evidence in this context refers to microeconomic evidence from structural models at the firm, plant or household level. One example would be evidence of nonlinear adjustment in consumer sentiment in response to oil price innovations. Although Edelstein and Kilian (2009) and Ramey and Vine (2010) have documented adjustments in U.S. consumer sentiment in response to retail energy

https://doi.org/10.1017/S1365100511000186 Published online by Cambridge University Press

price innovations, their analysis is based on linear regression models and cannot be used to motivate the use of nonlinear models based on net oil price increases. The same caveat applies to the evidence in Ramey and Vine (2010) on the responses of key indicators for the U.S. automobile industry to various measures of energy price shocks. Both papers transform retail energy price data to allow for time variation, but the type of nonlinearity considered in these two papers is fundamentally different from that in Mork (1989) or Hamilton (1996, 2003), the focus is on retail energy prices rather than the price of oil, and no allowance is made for asymmetric responses.

4. Although the empirical results in Mork (1989) continue to be cited as evidence of asymmetry, the substance of his results was overturned in Hamilton (1996) and Hooker (1996). Moreover, Kilian and Vigfusson (in press) show that even the original statistical test used by Mork (1989) does not reject the null of symmetric slopes when using data for 1973.II–2007.IV. Qualitatively similar results are obtained using impulse response–based tests of symmetry.

5. OLS estimation of model (1″) is based on the premise that the error u_t is uncorrelated with Δs_t. This is not an additional assumption, but a direct implication of the existence of a structural model in which the price of energy is predetermined with respect to real GDP. Model (1) does not require that assumption because it is not structural.

6. It should be noted that the three-year net oil price increase specification was originally selected by Hamilton (2003) on the basis of Bayesian model selection methods applied to data that overlap substantially with the sample on which the slope-based test of this specification is based. This introduces an element of data mining that is not reflected in standard asymptotic critical values [see, e.g., Inoue and Kilian (2004)]. Following the related literature, our analysis ignores this caveat.

7. Hamilton's (in press) discussion may give the impression that the central idea of Kilian and Vigfusson is already contained in Balke et al. (2002). As discussed in detail in the working paper version of this article, this is not the case. Balke et al. certainly deserve credit for being the first researchers to have recognized that censored oil price VAR models are inherently misspecified, but their solution to this problem is different from that in Kilian and Vigfusson (in press) in several dimensions, and the full extent of their analysis of the problems with censored oil price VAR models is a statement that such models "are not completely suitable for an examination of asymmetry" and that "it is not at all clear how to interpret a negative Hamilton innovation."

8. Raising the lag order to six increases the p-values somewhat. For the real PPI, for example, the p-values increase to .96 for the one–standard deviation oil price innovation and to 0.15 for the two–standard deviation oil price innovation.

9. There is no a priori reason for the results for industrial production to match the results for real GDP. One reason is that the share of the service sector in U.S. real GDP has greatly increased in recent decades. To the extent that the economic models implying asymmetric responses of real output are designed for the industrial sector, one would expect weaker evidence of asymmetric responses for aggregate real GDP than for aggregate industrial production. A second reason is that asymmetric responses may be important at the sectoral level without necessarily dominating aggregate real economic activity. That is why disaggregate analysis is valuable in assessing the empirical content of economic models of the transmission of oil price shocks. At the same time, disaggregate analysis involving a large number of sectors necessitates the use of critical values that are robust against data mining.

10. In contrast, earlier studies such as Lee et al. (1995) and Ferderer (1996) focused on single-equation models of the effect of oil price uncertainty on real output in which the price of oil is treated as exogenous. That approach is consistent with theoretical models of the uncertainty effect such as Bernanke (1983), which treat the price of oil as exogenous with respect to the U.S. economy, but inconsistent with the modern view that the real price of oil contains an important endogenous component [see Kilian (2008)]. Moreover, earlier empirical studies of the uncertainty effect used data from the pre-1973 period, making the empirical results difficult to interpret.

11. A useful exercise for future research would be to evaluate data generated from the asymmetric model estimated in Elder and Serletis using the methodology of Kilian and Vigfusson (in press) to determine whether that procedure has the power to detect the underlying departures from the null of symmetric responses. If it did, this would cast doubt on the findings in Elder and Serletis. If it did not, the power argument would seem compelling.

https://doi.org/10.1017/S1365100511000186 Published online by Cambridge University Press

12. For example, as the OPEC regime collapsed in 1986, the likelihood of future supply disruptions caused by OPEC evaporated, removing a major source of uncertainty about the future price of oil. There was no reason to extrapolate from the observed price volatility in 1986 to future volatility if the collapse of OPEC was viewed as a one-time event.

13. It may seem that if there is increased uncertainty about the real price of oil that we expect to be resolved next quarter, then firms affected by this price should necessarily postpone their investment decisions, even if there is no change in the longer-term oil price uncertainty. Upon reflection, that conclusion is not self-evident. For example, consider a company deciding whether to build a plant to produce a new SUV. The company's profits depend on the price of oil, in that demand for this vehicle will decline if the price of oil increases. Suppose that if the company decides to build the plant today, it will take four quarters to start production and for the cash flow from selling the SUVs to start. Also take as given that the investment is supposed to amortize itself within five years. To the extent that our forecast of the conditional variance converges back to the unconditional variance within four quarters, an unexpected increase in the conditional variance of the price of oil today will not affect this investment decision (because the cash flow from the investment that depends on the price of oil does not start within that horizon). More generally, even if the predictable variation in the conditional variance is more persistent and will affect expected oil price volatility for a few additional quarters, the overall effect is likely to be small, considering the long horizon for which the cash flow arises (quarters 5 through 24 in the example) and the tendency for GARCH forecasts of the conditional variance to revert to the unconditional variance at short horizons.

14. In rare cases, the relevant forecast horizon may be short enough for empirical analysis. For example, Kellogg (2010) makes the case that for the purpose of drilling oil wells in Texas, as opposed to Saudi Arabia, a forecast horizon of only 18 months is adequate. Even at that horizon, however, there are no oil-futures options price data that would allow the construction of implied volatility measures. Kellogg (2010) therefore converts the one-month volatility to 18-month volatilities based on the term structure of oil futures.

15. A closely related point has recently been made by Ramey and Vine (2010), who propose to adjust the real price of gasoline for the cost of gasoline shortages induced by regulation. They also consider defining energy price shocks in terms of unanticipated changes in Michigan consumer sentiment regarding car buying conditions. As in Edelstein and Kilian's work, these adjustments allow us to convert a nonlinear regression relationship into a linear regression model.

16. The raw data start in January 1974 (which is when the EIA started collecting RAC data) both to avoid the well-documented structural change in the predictive relationship in late 1973 and to avoid having to make ad hoc assumptions about backdating the oil price data. This facilitates a clean comparison of alternative oil price measures and implies that the estimation period starts in 1975.Q2. The initial estimation period ends in 1990.Q1 (right before the invasion of Kuwait in August of 1990) and the forecast evaluation ends in 2010.Q2.

17. In this regard, Hamilton (in press) provides additional evidence based on nonlinear local projections (LP) in the spirit of Jordà (2005, 2009), but his analysis ignores precisely the concerns articulated in Kilian and Vigfusson (in press) that such local projections depend on recent history and on the magnitude of the innovation, necessitating the use of Monte Carlo integration methods [also see Jordà (2005, p. 168)]. Moreover, Hamilton's local projections are not designed to capture structural responses. This would require a consistent estimate of the structural impact multiplier matrix from an external structural model such as the model proposed by Kilian and Vigfusson [see, e.g., Jordà (2009) and Kilian and Kim (2011), equations (5)–(7)]. Nor can the slope-based test that Hamilton reports for the LP model be interpreted as a test of the symmetry of the response to oil price innovations.

REFERENCES

Alquist, Ron, Lutz Kilian, and Robert J. Vigfusson (2010) Forecasting the price of oil. In G. Elliott and A. Timmermann (eds.), *Handbook of Economic Forecasting* 2. Amsterdam: North-Holland.

https://doi.org/10.1017/S1365100511000186 Published online by Cambridge University Press

Anderson, Soren, Ryan Kellogg, and James Sallee (2010) What Do Consumers Know (or Think They Know) About the Price of Gasoline? Mimeo, University of Michigan.

Balke, Nathan S., Stephen P.A. Brown, and Mine K. Yücel (2002) Oil price shocks and the U.S. economy: Where does the asymmetry originate? *Energy Journal* 23, 27–52.

Bernanke, Ben S. (1983) Irreversibility, uncertainty, and cyclical investment. *Quarterly Journal of Economics* 98, 85–106.

Bernanke, Ben S., Mark Gertler, and Mark W. Watson (1997) Systematic monetary policy and the effects of oil price shocks. *Brookings Papers on Economic Activity* 1, 91–142.

Carlstrom, Charles T. and Timothy S. Fuerst (2006) Oil prices, monetary policy and counterfactual experiments. *Journal of Money, Credit and Banking* 38, 1945–1958.

Cooley, Thomas F. and Stephen F. LeRoy (1985) Atheoretical macroeconometrics: A critique. *Journal of Monetary Economics* 16, 283–308.

Edelstein, Paul and Lutz Kilian (2007) The response of business fixed investment to energy price changes: A test of some hypotheses about the transmission of energy price shocks. *B.E. Journal of Macroeconomics* 7(1). Available at http://www.bepress.com/bejm/vol7/iss1/art35.

Edelstein, Paul and Lutz Kilian (2009) How sensitive are consumer expenditures to retail energy prices? *Journal of Monetary Economics* 56, 766–779.

Elder, John and Apostolos Serletis (2010). Oil price uncertainty. *Journal of Money, Credit and Banking* 42, 1138–1159.

Ferderer, J. Peter (1996) Oil price volatility and the macroeconomy: A solution to the asymmetry puzzle. *Journal of Macroeconomics* 18, 1–16.

Gillman, Max and Anton Nakov (2009) Monetary effects on nominal oil prices. *North American Journal of Economics and Finance* 20, 239–254.

Hamilton, James D. (1988) A neoclassical model of unemployment and the business cycle. *Journal of Political Economy* 96, 593–617.

Hamilton, James D. (1996) This is what happened to the oil price–macroeconomy relationship. *Journal of Monetary Economics* 38, 215–220.

Hamilton, James D. (2003) What is an oil shock? *Journal of Econometrics* 113, 363–398.

Hamilton, James D. (2009) Causes and consequences of the oil shock of 2007–08. *Brookings Papers on Economic Activity* 1 (Spring), 215–261.

Hamilton, James D. (in press) Nonlinearities and the macroeconomic effects of oil prices. *Macroeconomic Dynamics*.

Hamilton, James D. and Ana María Herrera (2004) Oil shocks and aggregate economic behavior: The role of monetary policy: Comment. *Journal of Money, Credit and Banking* 36, 265–286.

Herrera, Ana María., Latika G. Lagalo, and Tatsuma Wada (in press) Oil price shocks and industrial production: Is the relationship linear? *Macroeconomic Dynamics*.

Herrera, Ana María and Elena Pesavento (2009) Oil price shocks, systematic monetary policy, and the "Great Moderation." *Macroeconomic Dynamics* 13, 107–137.

Hooker, Mark A. (1996) What happened to the oil price–macroeconomy relationship? *Journal of Monetary Economics* 38, 195–213.

Inoue, Atsushi and Lutz Kilian (2004) In-sample or out-of-sample tests of predictability: Which one should we use? *Econometric Reviews* 23, 371–402.

Jordà, Òscar (2005) Estimation and inference of impulse responses by local projections. *American Economic Review* 95, 161–182.

Jordà, Òscar (2009) Simultaneous confidence regions for impulse responses. *Review of Economics and Statistics* 91, 629–647.

Kellogg, Ryan (2010) The Effect of Uncertainty on Investment: Evidence from Texas Oil Drilling. Mimeo, University of Michigan.

Kilian, Lutz (2008) The economic effects of energy price shocks. *Journal of Economic Literature* 46, 871–909.

Kilian, Lutz and Yun Jung Kim (2011) How reliable are local projection estimators of impulse responses? *Review of Economics and Statistics* 93(4), 1460–1466.

https://doi.org/10.1017/S1365100511000186 Published online by Cambridge University Press

Kilian, Lutz and Logan T. Lewis (2011) Does the Fed respond to oil price shocks? *The Economic Journal* 121, 1047–1072.

Kilian, Lutz and Clara Vega (2011) Do energy prices respond to U.S. macroeconomic news? A test of the hypothesis of predetermined energy prices. *Review of Economics and Statistics* 93(2), 660–671.

Kilian, Lutz and Robert J. Vigfusson (in press) Are the responses of the U.S. economy asymmetric in energy price increases and decreases? Quantitative Economics.

Koop, Gary, Pesaran M. Hashem, and Simon M. Potter (1996) Impulse response analysis in nonlinear multivariate models. *Journal of Econometrics* 74, 119–147.

Lee, Kiseok, Shawn Ni, and Ronald A. Ratti (1995) Oil shocks and the macroeconomy: The role of price variability. *Energy Journal* 16, 39–56.

Mork, Knut A. (1989) Oil and the macroeconomy. When prices go up and down: An extension of Hamilton's results. *Journal of Political Econ*omy 97, 740–744.

Nakov, Anton and Andrea Pescatori (2010) Monetary policy trade-offs with a dominant oil producer. *Journal of Money, Credit and Banking* 42, 1–32.

Pindyck, R.S. (1991) Irreversibility, uncertainty and investment. *Journal of Economic Literature* 29, 1110–1148.

Ramey, Valerie A. and Daniel J. Vine (2010) Oil, automobiles, and the U.S. economy: How much have things really changed? *NBER Macroeconomics Annual* 25, 333–368.

https://doi.org/10.1017/S1365100511000186 Published online by Cambridge University Press