Structural change and the dynamics of energy prices: an identification-robust test for time-varying parameters *

Jean-Thomas Bernard † Université Laval

Jean-Marie Dufour ‡ Université de Montréal

Lynda Khalaf § Université Laval

Maral Kichian ¶ Bank of Canada

First version: November 2002
Revised: March 2003, October 2005, December 2005, October 2006
This version: October 29, 2006
Compiled: October 29, 2006, 6:45pm

* This work was supported by the Canada Research Chair Program (Chair in Econometrics, Université de Montréal and Chair in Environment, Université Laval), the Chair on the Economics of Electric Energy (Université Laval), the Institut de Finance mathématique de Montréal (IFM2), the Alexander-von-Humboldt Foundation (Germany), the Canadian Network of Centres of Excellence [program on *Mathematics of Information Technology and Complex Systems* (MITACS)], the Canada Council for the Arts (Killam Fellowship), the Natural Sciences and Engineering Research Council of Canada, the Social Sciences and Humanities Research Council of Canada, the Fonds de recherche sur la société et la culture (Québec), and the Fonds de recherche sur la nature et les technologies (Québec). The authors thank Eva Ortega for several helpful suggestions and comments, and Robert Pindyck for kindly providing the data set. The views expressed in this paper are those of the authors. No responsibility should be attributed to the Bank of Canada.

† Holder of the Chair on the Economics of Electric Energy. Mailing address: GREEN, Université Laval, Pavillon J.-A.-De Sève, St. Foy, Québec, Canada, G1K 7P4. TEL: (418) 656 5123; FAX: (418) 656 7412; e-mail: jtber@ecn.ulaval.ca

‡ Canada Research Chair Holder (Econometrics). Centre interuniversitaire de recherche en analyse des organisations (CIRANO), Centre interuniversitaire de recherche en économie quantitative (CIREQ), and Département de sciences économiques, Université de Montréal. Mailing address: Département de sciences économiques, Université de Montréal, C.P. 6128 succursale Centre-ville, Montréal, Québec, Canada H3C 3J7. TEL: 1 (514) 343 2400; FAX: 1 (514) 343 5831; e-mail: jean.marie.dufour@umontreal.ca . Web page: http://www.fas.umontreal.ca/SCECO/Dufour

§ Canada Research Chair Holder (Environmental and Financial Econometric Analysis). Groupe de recherche en économie de l'énergie, de l'environnement et des ressources naturelles (GREEN) Université Laval, Centre interuniversitaire de recherche en économie quantitative (CIREQ), and Economics Department, Carleton University. Mailing address: Economics Department, Carleton University, Loeb Building 1125 Colonel By Drive, Ottawa, Ontario, K1S 5B6 Canada. Tel (613) 520-2600-8697; FAX: (613)-520-3906. e-mail: Lynda_Khalaf@carleton.ca.

¶ Research Department, Bank of Canada, Ottawa, Canada. TEL: (613) 782 7460; FAX: (613) 782 7163; e-mail: mkichian@bankofcanada.ca

ABSTRACT

We test for the presence of time-varying parameters in long-run real energy prices as suggested in the class of models proposed by Pindyck (1999). These models postulate mean-reverting prices with continuous and random changes in their level and trend, and are estimated using Kalman filtering techniques. Since the latter contain boundary as well as possibly unidentified parameters when there is no parameter variation, the test statistics have non-standard distributions and involve nuisance parameters. To solve the statistical difficulties associated with this problem, we use a simulation-based method, called *maximized Monte Carlo* tests [Dufour (2006)], which can yield provably exact tests for highly irregular problems even in small samples. Significant parameter variation is detected for both oil and natural gas prices, but not for coal price. Contrary to Slade (1982) who used a deterministic trend model, we find no evidence of U-shaped trend for oil or coal prices; there is almost no upward trend either. In contrast, the natural gas price trend has an upward U-shaped form.

Keywords : structural change; time varying parameter; energy prices; coal; gas; crude oil; unidentified nuisance parameter; exact test; Monte Carlo test; Kalman filter.

Journal of Economic Literature classification: C22, C52, C53, Q40.

RÉSUMÉ

Nous testons la présence de paramètres variables dans les équations suggérées par Pindyck (1999) afin de modéliser les prix à long terme de l'énergie. Ces modèles postulent des processus de retour à la moyenne avec des changements continus et aléatoires des coefficients de la tendance, et sont estimés en utilisant des techniques de filtrage de Kalman. Comme ces derniers comprennent des paramètres sur la frontière d'admissibilité ou non-identifiés en l'absence de changement structurel, les distributions des statistiques de test correspondantes sont typiquement non-standards et dépendent de paramètres de nuisance. Pour résoudre les difficultés statistiques associées à ces problèmes, nous appliquons une procédure basée sur des simulations, la méthode des tests de Monte Carlo maximisés [Dufour (2006)], laquelle fournit des tests dont le niveau est contrôlé pour des problèmes hautement irréguliers, même sur de petits échantillons. Nous détectons des paramètres variables dans le cas du pétrole brut et du gaz naturel, mais pas pour le charbon.

Mots clés : changement structurel ; prix de l'énergie; charbon; gaz naturel; pétrole brut; retour à la moyenne; test exact; test de Monte Carlo; test de Monte Carlo maximisé; bootstrap; filtre de Kalman.

Classification du Journal of Economic Literature : C22, C52, C53, Q40.

Contents

List of Tables

List of Figures

1. Introduction

It is widely recognized that fluctuations in energy prices have important and lasting effects on the economies of industrialized countries. As a recent example, Hamilton (2003) found a strongly significant nonlinear relationship between changes in oil prices and GDP growth. Similarly, for small open economies, Amano and van Norden (1998) found long-run links between oil prices and real exchange rates. Enduring price movements in energy commodities can also cause relative price changes among a wide range of products in the economy, which can then influence the rate of inflation over an extended period.[1]

Interpreting and predicting the behavior of energy prices have proved to be quite challenging. In addition to supply and demand conditions – both domestic and international – a complete model also needs to take into account market regulations, technological advances, and geopolitical considerations. These non-market-related aspects present the biggest challenges. For that reason, a fully articulated structural model may be quite difficult to build and unreliable, and Pindyck (1999) has suggested that a more practical approach for long-run forecasting may consist in building simple flexible dynamic models which incorporate implicitly the role played by demand shifts, technological change and depletion. These models are more general than models with deterministic trends as used by Slade (1982) to analyze the long-term behavior of depletable resource prices. Although the latter are special cases of the former, standard statistical tests can be misleading because of boundary nesting and the application of appropriate tests is a contribution of this paper.

We reconsider the class of trend models with time varying parameters (TVP) proposed by Pindyck (1999) for that purpose. Despite their simplicity, these models are quite flexible, allowing both random walks with drift and/or changing trend lines where prices revert to a possibly moving mean. An important feature of the proposed models is the inclusion of time varying trend parameters. Different forms of the model may reflect alternative assumptions on resource depletion and technological change. For example, a form of energy which is produced and sold competitively should have its price reverting to a long-run marginal cost which itself can change over time. Using a simple Hotelling model, Pindyck showed that long-run energy prices should revert to an *unobservable* trending long-run marginal cost, with continuous random changes in the level and slope of the trend. A family of econometric models for these prices which integrate desired features (discrete versions of multivariate Ornstein-Uhlenbeck processes) was also suggested for energy prices. Alternative versions of these models were estimated using Kalman filter techniques, and out-of-sample forecasts were computed. The forecast exercises conducted by Pindyck yield mixed results, but overall the class of models considered appears to be quite promising.

Pindyck (1999) did not, however, provide statistical tests for the proposed class of models. In particular, the time varying parameter specification was not tested statistically. Yet, the decision to

[1] There is no unanimity about the interpretation of the recent history in this respect. For a critical review, see Barsky and Kilian (2004).

use a TVP model (as opposed to a more common autoregressive or fixed coefficient trend model) may have non-negligible statistical consequences.

An important difficulty here comes from the fact that TVP models contain boundary as well as unidentified parameters, especially when there is no parameter variation, so that test statistics have non-standard distributions and involve nuisance parameters. Usual chi-square critical points can easily lead to spurious rejections even with fairly large data sets, because the regularity conditions underlying classical asymptotics fail.[2] Another problem is the fact that likelihood functions tend to be *ill-behaved* for empirically relevant parameter values. Thus, although sophisticated numerical recipes and global maximizers are readily available, it is well known that perfect maximization is not granted in this context.

Here, we complement Pindyck's (1999) study by conducting tests for continuous and random shifts in real energy prices. We are thus able to select, within the suggested family of models, specifications that are statistically justified for crude oil, coal, and natural gas prices. Our methodology relies on exact simulation-based test procedures, applicable – even with small samples – to highly irregular problems for which standard techniques are not valid.

The merit of simulation-based procedures was recently demonstrated in a related framework, namely for the presence of jumps in the context of jump diffusions; see Dufour and Khalaf (2001), Khalaf, Saphores and Bilodeau (2003), and Saphores, Khalaf and Pelletier (2002). Those studies circumvent the unidentified nuisance-parameter problem through bounds Monte Carlo tests. These involve simulation-based cut-off points (or p-values) based on conservative bounds. The approach used in the present paper is related to the latter in the sense that it is also bounds-based. However, here, an *optimal* (*i.e.*, tightest) bound is provided. Specifically, we apply the *maximized Monte Carlo* (MMC) test technique [Dufour (2006)], which is based on comparing the maximal p-value of the test (over the nuisance parameters, obtained by simulation) with the significance level. Consequently, (i) level control is ensured by construction, and (ii) the tightest cut-off point is obtained numerically. For further applications of the MC method in the context of energy markets, see Bernard, Idoudi, Khalaf and Yelou (2006).[3]

Our findings indicate significant TVP effects in two of the three energy-price series examined, which supports Pindyck's proposed class of models. Indeed, we find significant parameter variation in the natural gas and oil series. Our results suggest a slow reverting process of log price to a fast adjusting random mean and a very slow adjusting trend. We find no evidence of U-shaped trend with the logs of oil and coal prices. There is almost no upward trend over 120 years.

In Section 2, we describe the class of proposed models and the test method used. Section 3 documents and discusses our empirical results. We conclude in Section 4.

[2]See Hansen (1996), Dufour (1997, 2003) and Andrews (2000, 2001).

[3]For further discussion of Monte Carlo test method in econometrics, see, for example, Dufour and Kiviet (1996, 1998), Kiviet and Dufour (1997), Dufour, Farhat, Gardiol and Khalaf (1998), Dufour and Khalaf (2001, 2002*a*, 2002*b*), Khalaf and Kichian (2002, 2004, 2005), Dufour, Khalaf, Bernard and Genest (2004).

2. Model and test method

Pindyck (1999) considers a basic Hotelling model for a depletable resource produced in a competitive market. With constant marginal cost of extraction c and an isoelastic demand function with unitary elasticity, the price level is given by

$$P_t = c + [(ce^{rt}/(e^{rcR_0/A} - 1)] \tag{2.1}$$

where R_0 is the initial stock of the depletable resource, A is a demand shifter, and r is the interest rate. This implies that the slope of the price trajectory is given by

$$dP_t/dt = rce^{rt}/(e^{rcR_0/A} - 1)\,, \tag{2.2}$$

so that changes in demand, extraction costs, and reserves all affect this slope. For example, an increase in A causes the slope to increase, while increases in c or R_0 reduce the slope. In addition, increases in c or A cause the price level to increase, whereas an increase in R_0 leads to a decrease in this level. If, as Pindyck (1999) argues, these factors fluctuate in a continuous and unpredictable manner over time, then long-run energy prices should revert to a trend which itself fluctuates in the same fashion.

A class of models which integrates the above features is the generalized Ornstein-Uhlenbeck process. Pindyck (1999) proposes a discretized version of this model as a suitable econometric framework for analyzing long-run energy prices. This leads to the following AR(1)-type dynamic model:

$$P_t = c_1 + \phi_{1t} + \phi_{2t}t + c_2 P_{t-1} + \epsilon_t\,, \quad t = 1, \dots, T\,, \tag{2.3}$$

where P_t refers to the logarithm of the real price of an energy product and the coefficients ϕ_{1t} and ϕ_{2t}, follow the stochastic processes

$$\begin{aligned} \phi_{1t} &= c_3\phi_{1,t-1} + v_{1t}\,, && (2.4)\\ \phi_{2t} &= c_4\phi_{2,t-1} + v_{2t}\,. && (2.5) \end{aligned}$$

The processes for ϕ_{1t} and ϕ_{2t} are unobservable, continuously evolving parameters which reflect long-run marginal costs including scarcity rent, in the underlying structural model.

For tractability purposes, we estimate and test this model separately for each price series considered. Clearly, a multivariate analysis (*i.e.* joint estimation) may be preferable, at least in principle, since energy sources are substitutes in the long run. Yet given sample size limitations, the numerical burden (pointed out by Pindyck) may outweigh efficiency gains for this particular application.

Further, modelling jointly all the equations involves additional specification choices that may be questionable.

For similar considerations, we also impose the following distributional assumptions on the univariate process (2.3). The underlying error terms ϵ_t, v_{1t}, and v_{2t}, $t = 1, \ldots, T$, are assumed to be independently and identically normally distributed with zero means and covariances, and variances σ_ϵ^2, $\sigma_{v_1}^2$, and $\sigma_{v_2}^2$, respectively. Stationarity constraints (on c_2, c_3 and c_4) are (at least in principle) not necessary. The lag structure, reliance on linear trends and on uncorrelated unobservable components, is dictated, as argued by Pindyck (1999), by the length of the available sample; see section 3, for a description of the (yearly) data considered for the study. It is worth noting that, despite the latter simplifying hypotheses, the statistical framework remains non-regular, as we will explain below.

Assuming normality of ϵ_t, v_{1t}, and v_{2t}, Pindyck proposes that Kalman filtering be applied to obtain paths for the state variables ϕ_{1t} and ϕ_{2t}. This means that, starting with initial values for model parameters and state variables, the filter computes at each period new values for the state variables to reflect new information on the observable series. Once the full paths of the state variables are determined, the model can be estimated by maximum likelihood. Details of the Kalman filtering procedure are described in Appendix B. In view of assessing the statistical significance of TVP effects, the null hypothesis of interest is a simple mean-reverting model around a fixed trend line [the trending Ornstein-Uhlenbeck process given by equation (24) in Pindyck (1999)], *i.e.*,

$$P_t = c_1 + \phi_1 + \phi_2 t + c_2 P_{t-1} + \epsilon_t\,, \quad t = 1, \ldots, T\,. \tag{2.6}$$

It is clear that the models to be compared statistically are nested at the boundaries of certain parameters; formally,

$$\text{model (2.6)} \subseteq \text{model (2.3) when } \sigma_{v_1}^2 \to 0,\ \sigma_{v_2}^2 \to 0, \text{ and } c_3 = c_4 = 1\,. \tag{2.7}$$

Further, it is easy to see that some parameters may not be identifiable under certain parameter configurations: for example, the constant term is not identified when $c_3 = 1$ and $\sigma_{v_1}^2 = 0$, and it is "poorly" identified when we are close to these values.

In this context, one cannot rely on estimated standard errors and standard limiting distributions, since their use for building tests and confidence sets is not justified even asymptotically. In particular, the distributions of some widely used test statistics, such as t-type and more generally Wald-type statistics, may be difficult (if not impossible) to bound under various null hypotheses, so that controlling the level of such tests may not be feasible. By contrast, the distributions of likelihood-ratio-type statistics appear to be more stable, so such tests provide a more appropriate basis for statistical inference; see Dufour (1997, 2003) and Stock, Wright and Yogo (2002).

Taking into account these observations and the potential computational cost of maximum likelihood (ML) estimation of TVP models, we focus here on a quasi-likelihood-ratio (QLR) statistic of the form:

$$QLR(J) = 2[L_{TVP}(J) - L_{FCM}] \tag{2.8}$$

where $L_{TVP}(J)$ and L_{FCM} are, respectively, the maximum of the log-likelihood functions associated with (2.3) and (2.6), J is the maximum number of iterations allowed in the numerical maximization exercise under (2.3), and the subscript FCM stands for the fixed-coefficient model. J is explicitly spelled out to take into account the fact that the number of iterations has an influence on the finite-sample distribution of the test statistic: depending on the number of iterations (along with a convergence criterion), one really considers different test statistics.[4]

In view of the boundary and identification difficulties mentioned above, it would be wrong to compare the QLR statistic with standard tabulated values from a χ^2 table. Both the finite-sample and asymptotic distributions of the QLR statistic may be affected by the irregularities inherent to the problem studied here.[5] We need a procedure that can take these features into consideration.

To test the hypotheses of interest in this context, we therefore resort to maximized Monte Carlo (MMC) tests [Dufour (2006)], which we apply to the QLR statistic. Let us denote the vector of nuisance parameters that appear under the null hypothesis as:

$$\theta = \left\{\lambda,\, \phi_2,\, c_2,\, \sigma_\epsilon^2\right\},\ \lambda = c_1 + \phi_1\,. \tag{2.9}$$

In Appendix A, we provide a more detailed exposition of the MMC test method for a general test statistic whose null distribution can be simulated given a vector of nuisance parameters θ. Herein, we will summarize the technique as it applies to our specific problem, where θ is given by (2.9). In our test procedure, all that is needed to obtain draws from the null data-generating process is to set a value for θ; the unidentified nuisance parameters (for example, c_3 and c_4 under the constant coefficient model) simply do not appear in the distribution. More precisely, we proceed as follows.

(i) We calculate the likelihood ratio statistic (2.8) using the maximized likelihood value of the TVP model (2.3) – the *alternative model* – and the one of the constant-coefficient model (2.6) – *the null model*. In the process, we save the quasi-maximum-likelihood estimate of θ imposing (2.6) and the observed value of $QLR(J)$. We denote these values $\hat{\theta}_{FCM}$ and $QLR_0(J)$ respectively.

(ii) We generate data from the null model by drawing from the normal distribution and by setting θ equal to its estimated value $\hat{\theta}_{FCM}$, reestimate the restricted and the alternative models from these simulated data, and compute the corresponding test statistic $QLR(J;\,\hat{\theta}_{FCM})$

[4]For further discussion of this issue, see Robinson (1988).
[5]See Andrews (2000, 2001), Dufour (1997, 2003) and Stock et al. (2002).

where the presence of $\hat{\theta}_{FCM}$ indicates that $QLR(J;\,\hat{\theta}_{FCM})$ depends on data simulated after setting $\theta = \hat{\theta}_{FCM}$. This process is repeated N times, yielding N simulated test statistics from a data generating process (DGP) that satisfies the null hypothesis: $QLR_i(J;\,\hat{\theta}_{FCM})$, $i = 1, \ldots , N$.

(iii) $QLR_0(J)$ is compared with this distribution and a p-value is calculated based on the rank of $QLR_0(J)$ relative to its simulated counterparts; see equations (A.2), (A.3), and (A.4) in Appendix A. We will call the number so obtained the *local Monte Carlo* (LMC) p-value.

The MMC technique involves repeating step (ii) above, sweeping over combinations of admissible values of θ instead of $\hat{\theta}_{FCM}$. This can be viewed as a Monte Carlo implementation of the standard definition of the level of a test in the presence of nuisance parameters: when a test is nuisance-parameter dependent, an α level is achieved by comparing the largest *p*-value over all nuisance parameters consistent with the null hypothesis to α [see Lehmann (1986)]. The MMC method, by construction, works exactly in this way. Thus, we obtain a MC p-value for each value of θ. The MMC p-value is then the highest MC p-value so obtained. Since the maximized p-value function is a non-differentiable step function, we use simulated annealing (a global non-gradient-based algorithm) to obtain the maxima [see Goffe, Ferrier and Rogers (1994)]. This provides, by construction, the tightest bound p-value. The MMC test is significant at level α if the MMC p-value (say p_{MMC}) is less than or equal to α :

$$p_{MMC} \leq \alpha\,. \tag{2.10}$$

Of course, if the MC p-value obtained in step (ii) with $\theta = \hat{\theta}_{FCM}$ exceeds α (*e.g.*, 10 per cent), there is no need to proceed with the maximization; this saves execution time. Test results are reported in Section 3.

If the MC p-value is computed using a single (consistent) parameter estimate of the nuisance parameters, we get a *local MC* (LMC) test or, equivalently, a parametric bootstrap test. Bootstrap procedures tend to be considerably more reliable than procedures based on asymptotic critical values. In the context of our problem, however, where the asymptotic distribution may depend in a discontinuous way on nuisance parameters, it is well known that bootstrap procedures may also fail even asymptotically.[6] By contrast, the MMC procedure is immune to such failures [see Dufour (2006)]. In the following section, we report LMC tests along with MMC test results.

[6] See Athreya (1987), Basawa, Mallik, McCormick, Reeves and Taylor (1991), Sriram (1994), Andrews (2000), Inoue and Kilian (2002, 2003).

3. Empirical results

We consider the annual data set analyzed by Pindyck (1999).[7] The series for crude oil and bituminous coal extend from 1870 to 1996, while, for natural gas, the data cover 1919 to 1996. The nominal price series up to 1973 are from Manthy (1978) and the U.S. Bureau of the Census (1975). Pindyck (1999) updated this series through 1995 using data from the U.S. Energy Information Agency and, for 1996, the *Wall Street Journal*. The series are deflated using the U.S. wholesale price index until 1970, and the producers price index thereafter. Estimation is conducted on the logarithm of real prices.

First, we test the constant coefficient model against Pindyck's general TVP specification for each energy product. In other words, we compute the QLR statistic for testing the fixed coefficient model (2.6) against the TVP model (2.3) imposing only stability restrictions:

$$H_{TVP} : |c_2| < 1, 0 < c_3 < 1, \ 0 < c_4 < 1\,. \tag{3.1}$$

Such restrictions as in (3.1) are not necessary for the validity of our test procedure. Yet, for this particular application, we have observed that stability constrains did enhance convergence and avoid corner solutions. Maximization is typically difficult to achieve in TVP contexts, and the numerical burden tends to be relatively heavy. Indeed, to obtain the MC p-values [see step (ii) in the previous section], we run numerical nonlinear optimization algorithms 200 times [once with observed data (to derive the observed test statistic) and 199 times with simulated data (to derive the simulated test statistics)] *for each admissible values of* θ. Furthermore, the underlying simulated samples are generated under the null (constant coefficient) model; so in this case, the maximization problem is by construction, ill-conditioned.

To circumvent these difficulties, we applied the following numerical improvements. First, maxima results obtained by gradient-based algorithms are validated using global maximizers. For that purpose, we used *simulated annealing* [see Goffe et al. (1994)], particularly with the observed samples. Second, following a "normal convergence" output, the associated QLR statistics are checked for positivity. Indeed, if a negative QLR is obtained which signals non-convergence, the maximization algorithm is re-initiated using the "imperfect" solution as a starting value, until positivity is secured.

The results obtained are reported in Table 1, where LMC refers to local MC (parametric bootstrap) p-values and MMC to maximized p-values.[8] The maximal number of iterations allowed

[7] The data were generously provided by Pindyck.

[8] For the regression coefficients, the values of θ explored are determined by taking 5 standard errors on both sides of each component θ_i of θ ($\hat{\theta}_{iFCM} \pm 5SE_i$). This is indeed quite wide (10 standard errors). Given that the estimators of the constant coefficient model converge at a rate at least as large as $T^{1/2}$, the box so obtained can be interpreted as a "consistent set estimator" of the relevant parameters [see Dufour (2006)]. For the error variances, we used the fixed interval $[0.0001,\ 2]$, where the upper bound is well over all the estimates of these parameters.

Table 1. Tests of constant parameter model and Pindyck's specifications

Model tested $(H_0) : P_t = c_1 + \phi_1 + \phi_2 t + c_2 P_{t-1} + \epsilon_t$				
Alternative model (H_1)	Energy type	LR	LMC	MMC
$P_t = c_1 + c_2 P_{t-1} + \phi_{1t} + \phi_{2t} t + \epsilon_t$	Oil	14.7973	.005	.010
$\phi_{1t} = c_3 \phi_{1t-1} + v_{1t}$	Coal	2.5973	.195	-
$\phi_{2t} = c_4 \phi_{2t-1} + v_{2t}$	Gas	35.0205	.005	.005

Note – The table reports p-values. LMC refers to the local MC (or parametric bootstrap) p-value and MMC to the maximized Monte Carlo p-value. The number of replications used for the MC tests is $N = 199$. Since the MMC p-value must be larger than the LMC p-value, it is not necessary to report the MMC p-value if the former is larger than the level of the test (in this case, 0.05).

before the estimation process stops is $J = 120$.[9] All the MC tests are implemented using $N = 199$ replications and the algorithm that maximizes the p-value function (in terms of θ) is initialized at the value used for the LMC test.

We see from the results in Table 1 that the constant coefficient model is rejected (by MMC tests) against the general TVP model for gas and oil prices at conventional significance levels, but not for coal prices (by the LMC procedure, hence also by the MMC procedure). Parameter estimates are reported in Table 2. We observe a slow adjustment process for all price series (estimates of c_2 exceed .5) which reverts to a fast adjusting random mean (estimates of c_3 are less than .5) and a random trend which has a very slow adjustment coefficients particularly for gas (estimates of c_4 exceed .8). These results indicate good overall statistical support for Pindyck's proposed class of models: there are significant TVP effects in two of the three price series examined. Our test conclusions may also explain why Pindyck's estimation results signal a poor fit for the coal price series.

Figure 1 shows our estimated price trends [formally, the estimated $\phi_{2t} t$ term from (2.3)] over the sample period. The trend functions spike at the wake of the late seventies fuel-price shocks; yet, following their impact, all three revert (slowly) to their respective long-run mean. Interestingly, the latter is flat for both oil and coal prices, and slopes upward for natural gas. These results may be contrasted with Slade (1982)'s U-shaped curves, except for gas where in our case a U-shaped form is visible albeit less markedly than in Slade (1982). Recall however that the series analyzed by Slade ends in 1978; differing results may thus be driven by the post 1985 data.

One may note that estimates reported by Pindyck (1999, Table 2) signal "corner solution" values (duly discussed in the paper): $\hat{c}_3$ = 1.0009 for oil, $\hat{c}_4$ is set to 1 for coal, and the time-varying drift is omitted for gas in which case the estimated value of c_4 is close to unity. Since we focus on (2.3) [with (3.1)] for all price series, our estimation results differ numerically from the ones reported by Pindyck. Differences may also reflect the sensitivity of Kalman filter estimates to filter initializations

[9]Typically, in the examples considered here, convergence was achieved with much less than 80 iterations.

Table 2. Parameter estimates for the considered energy price models

Model	Type	c_1	c_3	c_2	c_4	σ_{v_1}	σ_{v_2}	σ_ε
$P_t = c_1 + c_2 P_{t-1} + \phi_{1t} + \phi_{2t} t + \epsilon_t$	Oil	.3508	.4799	.7415	.8898	6.3E-5	.0206	.4031
$\phi_{1t} = c_3 \phi_{1t-1} + v_{1t}$	Coal	.3437	.1150	.8004	.9297	.2814	.0140	.0193
$\phi_{2t} = c_4 \phi_{2t-1} + v_{2t}$	Gas	2.7144	.0833	.5247	.9613	.1766	.0383	.0080

Model	Type	λ	c_2	ϕ_2	σ_ε
$P_t = \lambda + \phi_2 t + c_2 P_{t-1} + \epsilon_t$	Oil	.1894	.7869	.0016	.1944
$\lambda = c_1 + \phi_1$	Coal	.1288	.8976	.0006	.0861
	Gas	.0035	.9717	.0011	.1194

(a point raised by Pindyck).

Our statistical analysis underscores the merit of the TVP model, and motivates further improvements, for example via multivariate approaches. Such specifications may also prove to be quite useful to model long-run components of two-factor short-run long-run models for resource prices.[10]

4. Conclusion

This paper has tested the statistical significance of Pindyck's (1999) suggested class of econometric models for the behavior of long-run real energy prices. These postulate mean-reverting prices with continuous and random changes in their level and trend, using Kalman filtering for the estimation. In such contexts, the distributions of the test statistics are typically non-standard and depend on nuisance parameters. Exploiting simulation-based procedures to address this issue, we have reported results for both a standard Monte Carlo test and a maximized Monte Carlo test. Our findings lend support to the proposed TVP class of energy models against the null hypothesis of fixed-coefficient mean-reverting equations. That is, we have found statistically significant instabilities for oil and natural gas prices, but not for coal prices.

Our main finding is that there is little evidence in favour of an increasing trend in the price of oil from 1890 to 1996, although OPEC decisions did leave an imprint in the late seventies. The same observations apply to the price of coal. However, the natural gas price series which started in 1919 displays a U-shaped trend which clearly slopes upward. Our results regarding the oil price trend agrees with the long held view of Adelman (2004): the estimated oil price trends shows no indication of impending scarcity but bears the marks of the OPEC decisions in the post 1973 era.

[10] See $e.g.$ Schwartz and Smith (2000), Schwartz (1997), Cortazar and Schwartz (2003), and Bernard, Khalaf, Kichian and McMahon (2005).

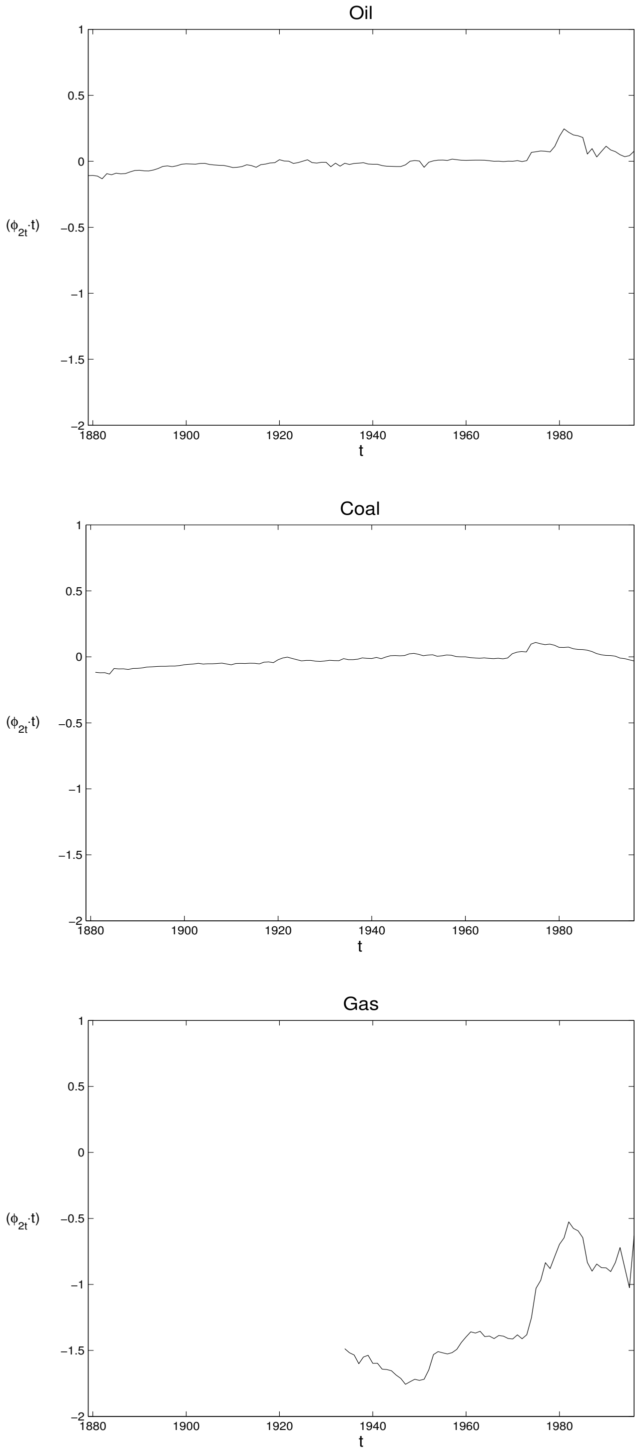

Figure 1. Estimated price trends

Appendix

A. Maximized Monte Carlo tests

Consider a (continuous) test statistic, S, whose distribution under the null hypothesis can be simulated once a finite set of nuisance parameters are specified. Conforming with the notational framework of Section 2, let us denote the nuisance parameter vector $\theta \in \Omega$, and let Ω_0 refer to the nuisance parameter subspace compatible with the null hypothesis H_0 under test.

Denote by S_0 the observed value of S and let $S_j,\ j = 1\ ,\ \dots\ ,\ N$ refer to N i.i.d. random draws from the null distribution of the statistic (given the parameter value θ). A *maximized Monte Carlo* (MMC) test is defined by the critical region

$$\sup_{\theta \in \Omega_0} [\hat{p}_N(S_0 \,|\, \theta)] \leq \alpha \tag{A.1}$$

where

$$\begin{aligned} \hat{p}_N(S_0 \,|\, \theta) &= \frac{N\hat{G}_N(S_0 \,|\, \theta) + 1}{N+1}\,, &\text{(A.2)}\\ \hat{G}_N(S_0 \,|\, \theta) &= \frac{1}{N}\sum_{i=1}^{N} \mathbf{1}_{[0,\infty)}(S_i - S_0)\,, &\text{(A.3)}\\ \mathbf{1}_A(x) &= \begin{cases} 1, & \text{if } x \in A\,,\\ 0, & \text{if } x \notin A\,. \end{cases} \end{aligned}$$

Note that $N\hat{G}_N(S_0 \,|\, \theta)$ is the number of simulated values of the test statistic not smaller than S_0. The formula for $\hat{p}_N(S_0 \,|\, \theta)$ gives a an *empirical p-value* for a given parameter vector θ. Then we can show the MMC test based on (A.1) has level α [see Dufour (2006)]:

$$\mathsf{P}\Big[\sup_{\theta \in \Omega_0} [\hat{p}_N(S_0 \,|\, \theta)] \leq \alpha\Big] \leq \alpha \quad \text{under } H_0.$$

The only condition needed to implement this procedure is the possibility of simulating the relevant test statistic under the null hypothesis.[11] The values of N and T (neither the number of replications nor the sample size) are taken as given, and no asymptotic argument is needed.

[11] For example, for $\alpha = 0.05$, N can be as low as 19. Although, in principle, raising N will typically increase the test power and decrease its sensitivity to the underlying randomization, the simulation results reported in Dufour and Kiviet (1996, 1998), Dufour et al. (1998), Dufour and Khalaf (2001, 2002*a*, 2002*b*), and Dufour et al. (2004), suggest that increasing N beyond 99 has only a small effect on power.

In this context, given any consistent estimate of θ that satisfies H_0 (denoted $\hat{\theta}$), a parametric bootstrap-type critical region can be obtained as:

$$\hat{p}_N(S_0 \mid \hat{\theta}) \leq \alpha \,. \tag{A.4}$$

In general, however, nothing guarantees that the level property

$$\mathsf{P}\left[\hat{p}_N(S_0|\theta) \leq \alpha\right] \leq \alpha \text{ under } H_0$$

holds. Under specific regularity conditions, the bootstrap p-value may be valid asymptotically in the sense that

$$\lim_{T\to\infty} \left\{ \mathsf{P}[\hat{p}_N(S_0|\hat{\theta}) \leq \alpha] - \mathsf{P}[\hat{p}_N(S_0|\theta_0) \leq \alpha] \right\} = 0 \tag{A.5}$$

where $\hat{p}_N(S_0|\theta_0)$ is the empirical p-value that one would obtain for the "true" (unknown) nuisance parameters values. Generic conditions for (A.5) to apply are given in Dufour (2006).

Unfortunately, in the context of the TVP models, (A.5) may easily not hold, for example because of identification problems [see Dufour (1997, 2003)]. In practice, this means that, if a test based on (A.4) rejects, this result may be spurious *even in large samples*. Yet bootstrap non-rejections are not subject to the same limitations: if the bootstrap-type test is not significant, then we can be sure that the exact MMC test is not significant at level α. Indeed,

$$\hat{p}_N(S_0|\hat{\theta}) > \alpha \Rightarrow \sup_{\theta \,\in\, \Omega_0} \left[\hat{p}_N(S_0|\theta)\right] > \alpha \,.$$

It is thus a good strategy to start the MMC sup-p-value step using a commonly used (*e.g.*, a constrained QMLE) estimate of θ.

B. Kalman Filtering and time varying parameters

This appendix draws heavily on Kim and Nelson (1999, Chapter 3). Consider the TVP model (2.3), which we rewrite for convenience in matrix notation as:

$$y_t = H_t\beta_t + Az_t + \epsilon_t \,, \tag{B.1}$$

$$\beta_t = F\beta_{t-1} + \eta_t,\; t = 1,\, \ldots\, , T \,, \tag{B.2}$$

$$\epsilon_t \overset{i.i.d.}{\sim} \mathrm{N}[0, R] \,, \tag{B.3}$$

$$\eta_t \overset{i.i.d.}{\sim} \mathrm{N}[0, Q] \,, \tag{B.4}$$

where

$$y_t = P_t\,, \quad H_t = \left[\begin{array}{cc} 1 & t \end{array}\right], \quad A = \left[\begin{array}{cc} c_1 & c_2 \end{array}\right], \tag{B.5}$$

$$\beta_t = \left[\begin{array}{c} \phi_{1t} \\ \phi_{2t} \end{array}\right], \quad z_t = \left[\begin{array}{c} 1 \\ P_{t-1} \end{array}\right], \quad z_t = \left[\begin{array}{c} 1 \\ P_{t-1} \end{array}\right], \quad \eta_t = \left[\begin{array}{c} v_{1t} \\ v_{2t} \end{array}\right], \tag{B.6}$$

$$F = \left[\begin{array}{cc} c_3 & 0 \\ 0 & c_4 \end{array}\right], \quad Q = \left[\begin{array}{cc} \sigma^2_{v_1} & 0 \\ 0 & \sigma^2_{v_2} \end{array}\right]. \tag{B.7}$$

The prediction equations in the Kalman filter algorithm are given by:

$$\beta_{t|t-1} = F\beta_{t-1|t-1}\,, \tag{B.8}$$

$$S_{t|t-1} = FS_{t-1|t-1}F' + Q\,, \tag{B.9}$$

where $\beta_{t|t-1}$ is the forecast value of β_t on the basis of information available through date $t-1$, and $S_{t|t-1}$ is its conditional variance. The conditional forecast error and its conditional variance can be obtained as:

$$e_{t|t-1} = y_t - H_t\beta_{t|t-1} - Az_t\,, \tag{B.10}$$

$$f_{t|t-1} = H_tS_{t|t-1}H'_t + R\,. \tag{B.11}$$

These expressions can be used in the updating equations of the algorithm according to

$$\beta_{t|t} = \beta_{t|t-1} + K_te_{t|t-1}, \tag{B.12}$$

$$S_{t|t} = S_{t|t-1} - K_tH_tS_{t|t-1}\,, \tag{B.13}$$

where the Kalman gain term is $K_t = S_{t-1|t-1}H'_tf^{-1}_{t|t-1}$.

If, in addition to the error terms ϵ_t and η_{it}, the initial value of β is Gaussian, then the distribution of y_t conditional on information available through time $t-1$ is also Gaussian, and its log-likelihood function is:

$$\ln(L) = -(1/2)\sum_{t=1}^{T} ln(2\pi f_{t|t-1}) - (1/2)\sum_{t=1}^{T} e'_{t|t-1}f_{t|t-1}e_{t|t-1}. \tag{B.14}$$

Therefore, given initial values for model parameters and state variables, the log-likelihood function can be maximized over the sample to yield maximum-likelihood parameter estimates. See Kim and Nelson (1999) for additional details.

References

Adelman, M. A. (2004), 'The real oil problem', *Regulation* **Spring**, 16–21.

Amano, R. and van Norden, S. (1998), 'Oil prices and exchange rates', *Review of International Economics* **6**, 683–694.

Andrews, D. W. K. (2000), 'Inconsistency of the bootstrap when a parameter is on the boundary of the parameter space', *Econometrica* **68**, 399–405.

Andrews, D. W. K. (2001), 'Testing when a parameter is on the boundary of the maintained hypothesis', *Econometrica* **69**, 683–733.

Athreya, K. B. (1987), 'Bootstrap of the mean in the infinite variance case', *The Annals of Statistics* **15**, 724–731.

Barsky, R. B. and Kilian, L. (2004), 'Oil and the macroeconomy since the 1970s', *Journal of Economic Perspectives* **18**, 115–134.

Basawa, I. V., Mallik, A. K., McCormick, W. P., Reeves, J. H. and Taylor, R. L. (1991), 'Bootstrapping unstable first-order autoregressive processes', *The Annals of Statistics* **19**, 1098–1101.

Bernard, J.-T., Idoudi, N., Khalaf, L. and Yelou, C. (2006), 'Finite sample inference methods for dynamic energy demand models', *Journal of Applied Econometrics* **forthcoming**.

Bernard, J.-T., Khalaf, L., Kichian, M. and McMahon, S. (2005), Conditional heteroskedasticity, jumps, and mean reversion in commodity prices, Technical report, GREEN, Université Laval, and Bank of Canada, Ottawa.

Cortazar, G. and Schwartz, E. S. (2003), 'Implementing a stochastic model for oil future prices', *Energy Economics* **25**, 215–238.

Dufour, J.-M. (1997), 'Some impossibility theorems in econometrics, with applications to structural and dynamic models', *Econometrica* **65**, 1365–1389.

Dufour, J.-M. (2003), 'Identification, weak instruments and statistical inference in econometrics', *Canadian Journal of Economics* **36**(4), 767–808.

Dufour, J.-M. (2006), 'Monte Carlo tests with nuisance parameters: A general approach to finite-sample inference and nonstandard asymptotics in econometrics', *Journal of Econometrics* **133**(2), 443–477.

Dufour, J.-M., Farhat, A., Gardiol, L. and Khalaf, L. (1998), 'Simulation-based finite sample normality tests in linear regressions', *The Econometrics Journal* **1**, 154–173.

Dufour, J.-M. and Khalaf, L. (2001), Monte Carlo test methods in econometrics, *in* B. Baltagi, ed., 'Companion to Theoretical Econometrics', Blackwell Companions to Contemporary Economics, Basil Blackwell, Oxford, U.K., chapter 23, pp. 494–519.

Dufour, J.-M. and Khalaf, L. (2002*a*), 'Exact tests for contemporaneous correlation of disturbances in seemingly unrelated regressions', *Journal of Econometrics* **106**(1), 143–170.

Dufour, J.-M. and Khalaf, L. (2002*b*), 'Simulation based finite and large sample tests in multivariate regressions', *Journal of Econometrics* **111**(2), 303–322.

Dufour, J.-M., Khalaf, L., Bernard, J.-T. and Genest, I. (2004), 'Simulation-based finite-sample tests for heteroskedasticity and ARCH effects', *Journal of Econometrics* **122**(2), 317–347.

Dufour, J.-M. and Kiviet, J. F. (1996), 'Exact tests for structural change in first-order dynamic models', *Journal of Econometrics* **70**, 39–68.

Dufour, J.-M. and Kiviet, J. F. (1998), 'Exact inference methods for first-order autoregressive distributed lag models', *Econometrica* **66**, 79–104.

Goffe, W. L., Ferrier, G. D. and Rogers, J. (1994), 'Global optimization of statistical functions with simulated annealing', *Journal of Econometrics* **60**, 65–99.

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

Hansen, B. E. (1996), 'Inference when a nuisance parameter is not identified under the null hypothesis', *Econometrica* **64**, 413–430.

Inoue, A. and Kilian, L. (2002), 'Bootstrapping autoregressive processs with possible unit roots', *Econometrica* **70**(1), 377–391.

Inoue, A. and Kilian, L. (2003), 'The continuity of the limit distribution in the parameter of interest is not essential for the validity of the bootstrap', *Econometric Theory* **19**(6), 944–961.

Khalaf, L. and Kichian, M. (2002), Simulation-based tests of pricing-to-market, *in* E. Kontoghiorghes, B. Rustem and S. Siokos, eds, 'Computational Methods in Decision-Making, Economics and Finance', Kluwer Academic Publishers, The Netherlands, pp. 583–603.

Khalaf, L. and Kichian, M. (2004), 'Pricing-to-market tests in instrumental regressions: Case of the transportation equipment industry', *Empirical Economics* **29**, 293–309.

Khalaf, L. and Kichian, M. (2005), 'Exact tests of the stability of the Phillips curve: The Canadian case.', *Computational Statistics and Data Analysis* **49**, 445–460.

Khalaf, L., Saphores, J.-D. and Bilodeau, J.-F. (2003), ‘Simulation-based exact jump tests in models with conditional heteroskedasticity’, *Journal of Economic Dynamic and Control* **28**, 531–553.

Kim, C.-J. and Nelson, C. R. (1999), *State-Space Models with Regime Switching: Classical and Gibbs-Sampling Approaches with Applications*, The MIT Press, Cambridge, Massachusetts.

Kiviet, J. F. and Dufour, J.-M. (1997), ‘Exact tests in single equation autoregressive distributed lag models’, *Journal of Econometrics* **80**, 325–353.

Lehmann, E. L. (1986), *Testing Statistical Hypotheses, 2nd edition*, John Wiley & Sons, New York.

Manthy, R. S. (1978), *Natural Resource Commodities: A Century of Statistics*, John Hopkins Press, Baltimore.

Pindyck, R. S. (1999), ‘The long run evolution of energy prices’, *The Energy Journal* **20**, 1–27.

Robinson, P. M. (1988), ‘The stochastic difference between econometric statistics’, *Econometrica* **56**(3), 531–548.

Saphores, J.-D., Khalaf, L. and Pelletier, D. (2002), ‘On jumps and arch effects in natural resource prices: An application to stumpage prices from pacific northwest national forests’, *American Journal of Agriculture Economics* **84**, 387–400.

Schwartz, E. S. (1997), ‘The stochastic behavior of commodity prices : Implications for valuation and hedging’, *The Journal of Finance* **52**, 923–973.

Schwartz, E. S. and Smith, J. E. (2000), ‘Conditional skewness in asset pricing tests’, *Short-Term Variations and Long-Term Dynamics in Commodity Prices* **46**, 893–911.

Slade, M. E. (1982), ‘Trends in natural resource commodity prices: An analysis of the time domain’, *Journal of Environmental Economics and Management* **9**, 132–137.

Sriram, T. N. (1994), ‘Invalidity of bootstrap for critical branching processes with immigration’, *The Annals of Statistics* **22**, 1013–1023.

Stock, J. H., Wright, J. H. and Yogo, M. (2002), ‘A survey of weak instruments and weak identification in generalized method of moments’, *Journal of Business and Economic Statistics* **20**(4), 518–529.