DEUTSCHE
BUNDESBANK

Forecasting the price of crude oil via convenience yield predictions

Thomas A. Knetsch

Discussion Paper
Series 1: Economic Studies
No 12/2006

Deutsche Bundesbank, Wilhelm-Epstein-Strasse 14, 60431 Frankfurt am Main,
Postfach 10 06 02, 60006 Frankfurt am Main

Tel +49 69 9566-1
Telex within Germany 41227, telex from abroad 414431, fax +49 69 5601071

Please address all orders in writing to: Deutsche Bundesbank,
Press and Public Relations Division, at the above address or via fax +49 69 9566-3077

ISBN 3–86558–137–4 (Printversion)
ISBN 3–86558–138–2 (Internetversion)

Abstract

The paper develops an oil price forecasting technique which is based on the present value model of rational commodity pricing. The approach suggests shifting the forecasting problem to the marginal convenience yield which can be derived from the cost-of-carry relationship. In a recursive out-of-sample analysis, forecast accuracy at horizons within one year is checked by the root mean squared error as well as the mean error and the frequency of a correct direction-of-change prediction. For all criteria employed, the proposed forecasting tool outperforms the approach of using futures prices as direct predictors of future spot prices. Vis-à-vis the random-walk model, it does not significantly improve forecast accuracy but provides valuable statements on the direction of change.

Keywords: oil price forecasts, rational commodity pricing, convenience yield, single-equation models.

JEL classification: E37; G12, G13, Q40; C22.

Non-Technical Summary

The paper develops an oil price forecasting technique on the basis of the present value model of rational commodity pricing. The central equation of the theoretical model describes the current spot price of crude oil as the sum of all discounted expected future payoffs received by the owner of one unit of this commodity ("convenience yields"). The discount factor is the sum of the risk-free interest rate and the oil-specific risk premium. The latter compensates for the holder's nondiversifiable risk. Convenience yields are defined as differences between the cost of carry and the futures prices of the commodity.

The forecasting problem is moved to the marginal convenience yield because this entity is clearly more predictable than, say, the oil price percentage change directly. The indirect method, however, requires the oil-specific risk premium to be estimated. This is done by a cointegration approach. Market expectations of the marginal convenience yield can be derived from the term structure of the oil market. Alternatively, the marginal convenience yield can be forecast on the basis of autoregressive (AR) models. Combinations between the two approaches are possible, too. Multi-step AR forecasts can be performed by either the plug-in technique or the direction estimation method. Moreover, several information criteria can be applied for model selection issues.

The forecast accuracy of the proposed technique is evaluated by out-of-sample projection exercises at horizons up to eleven months. The random-walk model and the approach of using futures prices as direct predictors of future spot prices serve as benchmarks. The sample of Brent oil prices used (i.e. starting in April 1991) is split into an estimation and an evaluation period. The latter comprises the post-January 1997 data in the first and the post-July 2000 data in the second experiment. The root mean squared error is the central evaluation criterion. The mean error and the relative frequency of a correct direction-of-change prediction are also considered. Moreover, statistical hypothesis tests à la Diebold and Mariano (1995) and Pesaran and Timmermann (1992) are employed.

The proposed forecasting technique outperforms the direct use of futures prices. The best variants improve the forecast quality significantly. The random walk model, however, is not significantly beaten. Nonetheless, the best variants of the proposed forecasting tool show more advantageous evaluation statistics. With respect to the quality of direction-of-change predictions, the proposed forecasting tool appears to outperform a naive coin flip.

Nicht technische Zusammenfassung

Das Arbeitspapier entwickelt eine Methode zur Prognose von Ölpreisen auf Grundlage des Barwertmodells rationaler Preisbildung auf Rohstoffmärkten. Die zentrale Gleichung des theoretischen Modells beschreibt den gegenwärtigen Kassapreis von Rohöl als Summe aller diskontierter, in Zukunft zu erwartender Zuflüsse aus dem Besitz einer Einheit dieses Rohstoffs ("Vorteilszinsen"). Der Diskontierungssatz ergibt sich rechnerisch aus einem risikofreien Zinssatz sowie einer ölspezifischen Risikoprämie, die den Investor für das nicht-diversifizierbare Risiko eines Engagements auf dem Ölmarkt entschädigt. Vorteilszinsen sind definiert als Abweichungen der Nettofinanzierungskosten von den Terminnotierungen des Rohstoffs.

Das Prognoseproblem wird auf die marginalen Vorteilszinsen mit dem Argument verlagert, dass sich diese deutlich besser prognostizieren lassen als beispielsweise die Veränderungsraten des Ölpreises direkt. Die indirekte Methode macht es jedoch notwendig, die ölspezifische Risikoprämie zu schätzen. Dies geschieht im Rahmen eines Kointegrationsansatzes. Prognosen des marginalen Vorteilszinses ergeben sich einerseits aus der Fristigkeitsstruktur am Ölterminmarkt. Andererseits können die Zeitreiheneigenschaften des marginalen Vorteilzinses für Prognosen auf Basis autoregressiver (AR) Modelle ausgenutzt werden. Schließlich ist auch eine Kombination beider Ansätze denkbar. Mehr-Schritt AR Prognosen lassen sich dynamisch (d.h. durch wiederholtes Einsetzen der Prognosen in das AR Modell) oder mittels direkter Schätzungen der Mehr-Schritt-Prognosegleichungen rechnen. Daneben kann die Modellselektion auf Basis verschiedener Informationskriterien erfolgen.

Die Prognoseeigenschaften des vorgeschlagenen Ansatzes werden im Rahmen rekursiver Out-of-Sample-Projektionen bis zu einem 11-Monats-Horizont überprüft. Als Vergleichsmaßstab dienen die Prognosen auf Basis der Random-Walk Hypothese ebenso wie die der Strategie, die Terminpreise als direkten Schätzer für den zukünftigen Kassapreis zu nehmen. Der Gesamtzeitraum genutzter Daten für Rohöl der Nordseesorte Brent (d.h. ab April 1991) wird in eine (ausschließliche) Schätz- und in eine Evaluationsperiode unterteilt. Letztere umfasst im ersten Experiment den Zeitraum ab Januar 1997 und im zweiten den Zeitraum ab Juli 2000. Die Wurzel der Summe der quadrierten Prognosefehler wird als zentrales Evaluationskriterium ausgewertet. Daneben finden aber auch die Summe der Prognosefehler und die relative Häufigkeit einer korrekten Vorzeichenprognose Beachtung. Ferner werden statistische Hypothesentests à la Diebold und Mariano (1995) sowie Pesaran und Timmermann (1992) angewendet.

Es zeigt sich, dass die vorgeschlagene Prognosetechnik der direkten Terminpreisprognose überlegen ist. Die besten Modellvarianten weisen gegenüber dieser Vergleichstechnik eine statistisch signifikant bessere Prognosequalität auf. Das Random-Walk Modell wird jedoch nicht signifikant geschlagen. Dennoch zeigen die besten Varianten der vorgeschlagenen Prognosetechnik vorteilhaftere Evaluationsstatistiken. Was die Qualität von Vorzeichenprognosen betrifft, scheint die vorgeschlagene Technik den naiven Münzwurf zu schlagen.

Contents

List of Tables

List of Figures

Forecasting the Price of Crude Oil via Convenience Yield Predictions[0]

1 Introduction

The industrialized world depends on crude oil as a central source of energy supply. Since economic activity and inflation are affected by the oil price, macroeconomic forecasts rely on its prospective evolution. The oil price is typically determined outside the core of macroeconomic projection exercises. Formally, it belongs to the set of exogenous assumptions the forecasts are conditioned upon. Against the background that oil prices are difficult to predict, this practice is a reasonable strategy because it allows forecast errors to be excused due to facts for which forecasters cannot be held responsible. But it does not circumvent the problem that the oil price path must be prolonged until the end of the projection horizon.

In general, the future development of the oil price could be inferred from modelling fundamental factors of the oil market such as supply, demand and inventories. However, most economic forecasting institutions only carry out such a structural analysis as complementary information, while the prospective path of the oil price is derived from a rather technical assumption. Two approaches are commonly in use. The first is based on the hypothesis that oil price changes are not predictable at all. Hence, the best guess for the oil price at any date in the future is its current level. Conceptually, the oil price is supposed to follow a random walk. In contrast, the second approach starts from the idea that oil prices would generally be predictable if the universe of information were appropriately processed. This task is difficult to accomplish with a single oil market model because the complex functioning of that market can only be described by a simple, partial and highly aggregated approximation. However, an organized commercial market which efficiently deals with the expectations of many and quite diverse modelling devices might be able to extract the relevant information. This kind of reasoning has led economic forecasters to set the h-step ahead prediction equal to the market price for delivery in h periods. Since the latter are found on futures markets,[1] let the second approach be called the "futures market hypothesis".

[0]Deutsche Bundesbank, Economics Department, Wilhelm-Epstein-Str. 14, D-60431 Frankfurt am Main, Germany, email: thomas.knetsch@bundesbank.de. The author thanks Stéphane Dées, Johannes Hoffmann, Karsten Ruth, Gerhard Ziebarth and an anonymous referee of the Bundesbank discussion paper series for useful comments and suggestions. Of course, the author is fully responsible for all remaining shortcomings. The paper expresses the author's personal opinion and does not necessarily reflect the views of the Deutsche Bundesbank.

[1]As futures contracts are marked to market, the futures price might differ from the forward price which would actually be the reference in this setup. In empirical applications, however, the two are treated as equivalent because differences are found to be negligible. Furthermore, futures contracts are more actively traded and price data are more readily available; see Pindyck (2001) for details.

This paper questions the futures market hypothesis. The issue, however, is not to generally deny that futures prices are an important source of information for oil price forecasts. The critical point is rather the assumption that futures prices are *unbiased* predictors of the respective future spot prices. In the empirical literature, there have been several attempts to test this hypothesis with oil market data. For instance, Moosa and Al-Loughani (1994) found that futures prices are neither unbiased nor efficient predictors of spot prices. However, this conclusion was not confirmed by some recent contributions. By running Mincer-Zarnowitz regressions, Chernenko et al. (2004) and Chinn et al. (2005) were not able to reject the unbiasedness hypothesis for crude oil. Coimbra and Esteves (2004) detected a downward bias increasing in the forecast horizon but, at least for short-term predictions, not statistically significant. While empirical evidence is mixed, there are strong theoretical arguments against the view that futures prices are unbiased predictors of future spot prices. The reason is related to the fact that crude oil is a means of production. Even if arbitrage profits arise, risk-averse oil-processing companies will not necessarily sell inventories and go long in the futures market because disposable oil reserves allow cost-minimizing production smoothing, reduce the risk of stock-outs and, occasionally, enable price speculation. As a consequence, there is a wedge between the futures price and the cost of carry. This "flow of services which accrues to the owner of a physical inventory but not to the owner of a contract for future delivery" (Brennan, 1991, p.33) is called convenience yield and is the ultimate source of bias in a scenario with risk-averse agents (see Pindyck, 1993, 2001).

In the case of crude oil, the wedge in the cost-of-carry relationship can rise to considerable levels. Hence, the convenience yield is not only a theoretical construct but also a phenomenon of empirical relevance. As futures contracts are treated for many delivery dates, a term structure of convenience yields can be derived, given money market rates at respective maturities. The yield curve can be understood as the discounted sum of market expectations of future marginal convenience yields. By plugging these predictions into the proposed forecasting model, we obtain a first variant, which can be seen as a tool striving to adjust futures prices for the systematic bias. Moreover, empirical studies show that the time series of the marginal convenience yield can be adequately approximated by a mean-reverting process (see e.g. Gibson and Schwartz, 1990). Therefore, the second variant consists in forecasting the marginal convenience yield by an autoregressive model. The combination of the market expectation and the autoregressive forecast defines a third variant. The forecasting step, however, only produces marginal convenience yield forecasts. The gap to the ultimate goal is bridged by Pindyck's (1993) present value model of rational commodity pricing. To compile oil price forecasts from marginal convenience yield forecasts, it is necessary to estimate a discount rate which is specific to crude oil. This is done by a cointegration approach à la Campbell and Shiller (1987).

Apart from the theoretical background and the econometric methodologies used, the paper presents an evaluation of the proposed forecasting device, in which the above-mentioned technical approaches serve as benchmarks. The analysis is structured as a recursive out-of-sample projection exercise on the basis of futures prices for Brent oil taken from the International Petroleum Exchange (IPE). The data set comprises monthly averages between April 1991 and March 2006. The post-January 1997 sample is reserved for the evaluation of forecasts between one and eleven months ahead. Standard forecast accuracy measures such as the root mean squared error, the mean error and the frequency of a correct direction-of-change prediction are employed. In this context, the paper also presents supplementary information on modelling issues and estimation results.

The forecast evaluation shows that the prospective path of the oil price should not be derived on the basis of the futures market hypothesis. The proposed forecasting tool performs better in terms of root mean squared errors. Diebold-Mariano tests prove that the differences in forecast accuracy are, at least partly, statistically significant. Moreover, its best performers show lower root mean squared errors than the random-walk model but the differences are not statistically significant at conventional levels. Regarding mean errors, the proposed forecasting tool is able to recognizably reduce the bias produced by the futures market hypothesis. Finally, its direction-of-change predictions appear to beat a naive coin flip. This is a particular advantage over the random-walk model which is, by definition, non-informative about the future direction of change.

The remainder of the paper is organized as follows. Section 2 depicts the theoretical and econometric foundations of the convenience yield forecasting model. Section 3 presents the data used. Section 4 elaborates on the central hypothesis that the marginal convenience yield is more predictable than the oil price percentage change. The first part of Section 5 provides the recursive estimates of the oil-specific risk premium as well as material on the specification and estimation of the forecasting models, while the second part comprises the forecast evaluation. Section 6 concludes.

2 The convenience yield forecasting model

Section 2.1 develops the structure of the convenience yield forecasting model. It includes a presentation of the economic theory of rational commodity pricing as well as a detailed description of econometric approaches to multi-step forecasting. Section 2.2 is devoted to parameter estimation.

2.1 Model setup

The forecasting model for the oil price is based on Pindyck's (1993) present value model of rational commodity pricing. Its central equation describes the spot price as the sum of all discounted expected future payoffs received by the owner of one unit of crude oil. Let prices be measured at the beginning of the period. The spot price of crude oil is given by

$$p(t) = \delta \sum_{i=0}^{\infty} \delta^i \mathsf{E}\psi(t+i) \quad \text{with} \quad \delta \equiv (1+\mu)^{-1} \tag{1}$$

where μ is the oil-specific 1-period discount rate, i.e. the expected rate of return an investor would require to hold a unit of crude oil, while $\psi(t)$ is the 1-period (net) marginal convenience yield, i.e. the benefit flow from holding a marginal unit of crude oil from the beginning to the end of period t, net of storage and insurance costs.[2] The operator E represents the expectation conditional on the universe of information available at the beginning of t. For the time being, let μ be constant and defined by $\mu = r + \rho$ where r is the (1-period) risk-free interest rate and ρ the oil-specific risk premium.[3]

Note that (1) is the solution of the difference equation

$$\mathsf{E}p(t+1) = (1+\mu)p(t) - \mathsf{E}\psi(t), \tag{2}$$

representing the one-step ahead forecast of the oil price conditional on an estimate of the marginal convenience yield of the current period. Through (2), the forecasting problem can be transferred from the oil price to the marginal convenience yield, provided that μ is known or taken from elsewhere. This shift would be advantageous if the convenience yield were easier to forecast than the oil price directly. In particular, the finance-related branch of the literature typically models the convenience yield as a mean-reverting process, implying that it has an autoregressive structure. Moreover, as contracts for the future delivery of crude oil are actively traded on organized markets, futures prices may contain market-related information on future oil prices. For reasons which will be shown in due course, it is more appropriate to extract the market information indirectly via the marginal convenience yield process than by following the rule of the futures market hypothesis.

Let us define the T-period convenience yield $\psi(t,T)$ from t to $t+T$, valued at (the beginning of) $t+T$, which is made up of the sequence of marginal convenience yields $\psi(t), \psi(t+1), ..., \psi(t+T-1)$ by the following formula:

$$\psi(t,T) = (1+\mu)^{T-1}\psi(t) + (1+\mu)^{T-2}\psi(t+1) + ... + \psi(t+T-1). \tag{3}$$

This definition is convenient because estimates of T-period convenience yields can be drawn from futures markets by imposing a no-arbitrage condition. The market expectation of $\psi(t,T)$ is determined by the cost-of-carry approach to futures pricing of storable commodities used as inputs for production, i.e.

$$\mathsf{E}_{\mathsf{M}}\psi(t,T) = (1+r_T)p(t) - f(t,T) \tag{4}$$

where $p(t)$ is the spot price, $f(t,T)$ the futures price for delivery at $t+T$ and r_T the risk-free T-period interest rate. For the time being, the latter is assumed to be constant over time. The marginal convenience yield of the current period is calculated through (4) by using the spot price, the futures price for delivery in one period, and the 1-period risk-free interest rate r, i.e. $\psi(t) \equiv \mathsf{E}_{\mathsf{M}}\psi(t,1) = (1+r)p(t) - f(t,1)$.

At this stage, two remarks have to be made. First, the literature sometimes refers to $\psi(t)$ as a measure of weak backwardation. Due to Litzenberger and Rabinowitz's (1995)

[2] Note that $\psi(t)$ is unknown when crude oil is priced in t.

[3] The flow of expected future benefits must compensate for the holder's nondiversifiable risk. In the context of a capital asset pricing model (CAPM), this condition implies $\rho = \phi\sigma\rho_m$ where ϕ is the market price of risk and σ and ρ_m are, respectively, measures of the oil price volatility and the correlation between the returns on crude oil and the market portfolio (see e.g. Dixit and Pindyck, 1994, p.115).

theoretical result, futures markets for crude oil should be weakly backwarded, in general; or formally, $\mathsf{E_M}\psi(t,T) > \mathsf{E_M}\psi(t,T-1) > ... > \mathsf{E_M}\psi(t,1) > 0$ for most t.[4] Owing to (3), weak backwardation requires the marginal convenience yield to possess a positive (unconditional) mean. Second, substituting $\mathsf{E_M}\psi(t,1)$ in (2) leads to $\mathsf{E_M}p(t+1) = f(t,1) + (\mu - r)\,p(t)$, implying that the futures price for delivery in one period would only be an unbiased estimator of the spot price in one period if the oil-specific discount rate were equal to the risk-free interest rate. However, in the relevant case of a strictly positive oil-specific risk premium ($\rho > 0$),[5] the futures price tends to under-predict the future spot price and the bias is proportionate to the current spot price. By generalizing the analysis to multi-step predictions, it is straightforward to show that the proportionality factor increases in the forecast horizon. For instance, Pindyck (2001) guesstimates an under-prediction of around 3 to 4.5 percent (of the current spot price) when the price of a 6-month futures contract is taken as a predictor of the spot price in half a year.

Market expectations of future marginal convenience yields up to $t+T-1$ are obtained on the basis of (3) and (4). With the convention $\psi(t,0) = 0$, a general expression is given by the formula

$$\mathsf{E_M}\psi(t+h) = \mathsf{E_M}\psi(t,h+1) - (1+\mu)\mathsf{E_M}\psi(t,h), \quad h = 0,1,...,T-1. \tag{5}$$

The repeated update of (2), given these estimates, allows us to compute the prospective path of the oil price up to period $t+T$, which is implied by the term structure of the futures market for crude oil at time t. Provided that the oil-specific discount rate is known, the system formed by (1) and (5) constitutes a model which extracts the information drawn from futures prices *and* accounts for a non-zero oil-specific risk premium. In this form, the convenience yield forecasting model can be understood as a tool for bias correction.

The (stylized) fact that the process generating the marginal convenience yield is mean-reverting has not been exploited so far. Under this assumption, the process may be approximated by an autoregressive (AR) model of finite order p,

$$\psi(t) = \alpha_0 + \alpha(L)\psi(t-1) + \varepsilon(t) \tag{6}$$

where α_0 is a positive constant,[6] $\alpha(L)$ a pth order lag polynomial satisfying the stability condition, i.e. $|\,1-\alpha(z)| = 0$ for $|z| > 1$, and $\varepsilon(t)$ a white-noise residual process.[7]

[4] There might be short periods of time when the oil market does not exhibit weak backwardation. In such scenarios, oil companies maintain production nonetheless because a complete cessation of production may be detrimental for some oil wells. For instance, stripper wells, once closed, cannot be reopened (see Litzenberger and Rabinowitz, 1995, p.1522).

[5] In terms of the CAPM, this is valid when the returns on crude oil and the market portfolio co-vary positively. Such a hypothesis seems reasonable, as strong economic growth is associated with above-average demand for oil, leading to high oil prices under normal market conditions (see Pindyck, 2001). Considine and Larson (2001), however, found a negative correlation using prices for Western Texas Intermediate crude oil and Standard and Poor's 500 Stock Price Index during the 1983-1999 period.

[6] This is implied by weak backwardation of the crude oil market.

[7] L is the backshift operator, i.e. $L^i x_t \equiv x_{t-i}$ for non-negative integers i.

Multi-step forecasts of the marginal convenience yield can be performed by iterating forward (6), in which the unknown future values are replaced by their forecasts (plug-in method). Specifically, the h-step ahead AR forecast is given by

$$\mathsf{E}_{\mathsf{A}}\psi(t+h) = [1-\alpha^{h+1}(1)]\psi_0 + \alpha^{h+1}(L)\psi(t-1), \quad h = 0, 1, ..., \tag{7}$$

where $\psi_0 \equiv \alpha_0/[1-\alpha(1)]$ denotes the (positive) unconditional mean. The system formed by (1) and (7) constitutes a model for forecasting oil prices based on the assumption that the marginal convenience yield is mean-reverting. Due to the stability condition, the resulting oil price forecasts converge to the constant level ψ_0/μ which may be interpreted as an equilibrium level because it is dependent on structural parameters such as the risk-free interest rate, the oil-specific risk premium and the mean marginal convenience yield.[8]

Unless marginal convenience yields are truly described by (6) and oil market traders know this data generating process, the two forecasting strategies will end up with different results. On the one hand, one could argue that $\mathsf{E}_{\mathsf{M}}\psi(t+h)$ encompasses $\mathsf{E}_{\mathsf{A}}\psi(t+h)$ at all forecast horizons h because the market mechanism should generally ensure efficient information processing. On the other hand, one cannot exclude the possibility of any kind of irrational pricing behavior on futures markets. It might therefore be advantageous to express the ultimate forecast as an optimal combination of both sources of information. Formally, the combined h-step ahead forecast of the marginal convenience yield is given by

$$\mathsf{E}_{\mathsf{C}}\psi(t+h) = (1-\omega_h)\mathsf{E}_{\mathsf{A}}\psi(t+h) + \omega_h\mathsf{E}_{\mathsf{M}}\psi(t+h), \quad 0 \le \omega_h \le 1, \quad h = 0, 1, ..., \tag{8}$$

where ω_h is the weight attached to the market expectation of the marginal convenience yield h-steps ahead.[9] Using (7) in this equation, we obtain

$$\mathsf{E}_{\mathsf{C}}\psi(t+h) = \pi_{0,h} + \pi_h(L)\psi(t-1) + \omega_h\mathsf{E}_{\mathsf{M}}\psi(t+h), \quad h = 0, 1, ..., \tag{9}$$

where the parameters $\pi_{0,h} \equiv (1-\omega_h)[1-\alpha^h(1)]\psi_0$ and $\pi_h(L) \equiv (1-\omega_h)\alpha^h(L)$ as well as the weight ω_h can be estimated in a single step. The system formed by (1) and (9) perform oil price forecasts combining the mean reversion hypothesis with futures market information.

It is worth commenting on the consequences of the last transformation. First, (9) implies that, in formal terms, it is not forecasts, but information sets which are actually combined. The two sources of information are, on the one hand, past marginal convenience yields and, on the other, market expectations of this quantity. From the theory of economic forecasting, it is well known that pooling information sets is preferable to pooling forecasts unless the former is impossible or results in a highly parametrized specification (see e.g. Diebold, 1989, p.590). Second, the forecasting model (9) has to be estimated both directly and separately for each h. Hence, not only the weight of the market expectation, but also the implied parameters of the AR structure differ from one horizon to another.

[8] Owing to the simplifying assumption of constant structural parameters, the equilibrium oil price level is constant, too. In reality, however, it changes over time as a consequence of moving structural parameters.

[9] The weight can be optimally chosen for the horizons $h = 0, 1, ..., T-1$. When forecasts beyond the period $t+T-1$ need to be performed, $\omega_h = 0$ must be set because market expectations of the marginal convenience yield can be computed up to this period only. The reason is that $f(t,T)$ is the price for the longest futures contract available.

Regarding the AR part only, this may or may not be advantageous a priori. Intuitively, iterating a misspecified autoregression may easily create substantial biases, especially as h increases. Johnston (1974) proved that the direct estimation technique would be less efficient (in terms of mean squared error) than the plug-in method if the estimate of (6) coincided with the true data generating process. However, the opposite is valid when the AR structure is infinite (see e.g. Findley, 1983; Weiss, 1991). Hence, as the relative advantage of either estimation technique depends on characteristics of the true model, it is a reasonable strategy to use both in the empirical application.[10]

All in all, the present value model of rational commodity pricing can be used to forecast oil prices via predictions of the marginal convenience yield. The latter are drawn from either market expectations, AR models, or combinations of the two.

2.2 Parameter estimation

In the previous section, a number of model parameters was assumed to be known even though they are not directly observable. In real projection exercises, these parameters need to be estimated. In the present value model (1), a value is required for the oil-specific discount rate (or risk premium). Furthermore, forecasting equations of the marginal convenience yield have to be specified and estimated.

Let us first address the question how to estimate the oil-specific discount rate. Campbell and Shiller (1987) established the link between the rational expectations present value model and the econometric concept of cointegration. Along this line of reasoning, Pindyck (1993) claimed that spot and futures prices should be cointegrated if these price series followed unit root processes themselves. The theoretical model implies that the relationship $f(t,1)-(1-\rho)p(t)$ is stable in the long run, which means that the oil-specific risk premium results from the freely estimated parameter of the cointegrating vector. By means of Johansen's (1991) multivariate approach, the cointegrating vector could be estimated in a flexible system. However, the present value model also provides information about the adjustment processes towards the long-run equilibrium. Since theory suggests that the error correction term should not Granger-cause the futures price,[11] according to Johansen (1992), it is efficient to run an ordinary least squares (OLS) regression of the conditional error correction model

$$\Delta p(t) = \beta_1 p(t-1) + \beta_2 f(t-1,1) + \beta_3 \Delta f(t,1) + \varepsilon(t) \tag{10}$$

where Δ is the difference operator, i.e. $\Delta \equiv 1-L$, $\beta_1 < 0$, $\beta_2 > 0$ and $\beta_3 > 0$ are parameters to be estimated and $\varepsilon(t)$ is a white-noise residual process.[12]

[10]By studying a large set of macroeconomic time series, Marcellino et al. (2006) found that the plug-in method tends to outperform the direct estimation method in terms of mean squared error.

[11]This implies that $f(t,1)$ is weakly exogenous with respect to the cointegrating vector. A weak exogeneity test in a bivariate vector error correction model validates the theoretical claim.

[12]If $\varepsilon(t)$ were not white noise, lagged differences of both the endogenous and the (weakly) exogenous variable might generally be included in the regression. In the present application, however, the whiteness of residuals is ensured in a specification without additional regressors.

By using the error-correction mechanism test proposed by Banerjee et al. (1998) (BDM), we are able to check whether or not the regression (10) is an appropriate framework for the estimation of the oil-specific risk premium. Essentially, the BDM test computes the t-statistic of the error correction parameter β_1. When this parameter is significantly negative, cointegration between the spot and the futures price is established, provided that the latter is weakly exogenous. The asymptotic distribution is non-standard but critical values are tabulated in Banerjee et al. (1998, Table 1A). In the presence of cointegration, an efficient estimate of the oil-specific risk premium is given by $\hat{\rho} = 1 + (\hat{\beta}_1/\hat{\beta}_2)$, in which the hats indicate the respective OLS estimates.

Let us turn to the issue of specifying and estimating the forecasting equations of the marginal convenience yield. Their structure is given by either (7) or (9)—depending on whether the AR or the combined approach is considered. In general, modelling follows a flexible general-to-specific rule. The lag length of the AR polynomial is chosen by means of information criteria. In practice, a number of different criteria are in use, among which Akaike's (1973) information criterion (AIC) and the Bayesian information criterion (BIC) suggested by Schwarz (1978) are perhaps the most common. If the data generating process is an AR model of finite order, the BIC may be preferable owing to its consistency property.[13] If the order is infinite, the AIC might have advantages. Moreover, the direct estimation of multi-step forecast equations will be asymptotically efficient if the lag order is selected by the AIC anew for each horizon.[14] Since the choice of the information criterion depends on the unknown data generating process, it is recommendable to use both the AIC and the BIC for model selection. When the latter yields a specification in which residuals are serially correlated, model selection is redone on the basis of the consistent but less parsimonious Hannan-Quinn (1979) criterion (HQ).

An AR polynomial, whose maximum lag is chosen by an information criterion, may contain coefficients which are not statistically significant. It is useful to set them equal to zero because the quality of forecasting models suffers from redundant regressors. Concretely, we apply a procedure which successively eliminates regressors whose t-statistics are lowest and do not exceed a threshold in absolute value. According to Brüggemann and Lütkepohl's (2001) result, the threshold can be chosen such that the sequential testing procedure mimics model reduction on the basis of an information criterion.[15] Hence, the model selection exercise as a whole, consisting of the initial lag length selection and the sequential elimination of redundant regressors, can be organized consistently in terms of the information criterion. Moreover, the direct OLS estimation of (9) has to account for autocorrelated residuals. The reason is that error sequences of h-step ahead forecasts are moving averages (MA) of order $h-1$. Correct t-statistics therefore require a heteroskedas-

[13] An information criterion is said to be consistent if it selects the true order asymptotically.

[14] See Bhansali (1999) for an overview of consistency and efficiency results with respect to model selection in the context of multi-step forecasting.

[15] According to Brüggemann and Lütkepohl (2001, Prop.1), the sequential elimination of regressors on the basis of an information criterion is equivalent to the testing procedure in single-equation models if the threshold at step j takes the value $[(e^{c_T/T} - 1)(T - K + j - 1)]^{1/2}$ where T is the number of observations and K the number of regressors, given $c_T = 2$ for the AIC and $c_T = \ln T$ for the BIC.

ticity and autocorrelation consistent (HAC) estimate of the residual covariance matrix. We use Newey and West's (1987) estimator with lag truncation $2(h-1)$.[16]

3 Data

Crude oil is not a uniform commodity. The characteristics of a brand depend upon the geological environment of the wells. The most important quality standards are the gravity and the sulphur content. In Europe, economic research and forecasting institutions typically use the spot price for North Sea Brent as a reference. Although IPE futures contracts with very short maturities have been traded since July 1988, only post-First Gulf War data (i.e. starting in April 1991) are used. Futures prices for the full set of maturities up to twelve months have been available since May 1994.

Although data on actual transaction prices of Brent oil do exist, we abstain from using them in the calculation of convenience yields because cash prices do not pertain the same specification for the commodity as futures prices. Furthermore, they often include discounts and premia resulting from longstanding buyer-seller relationships (see e.g. Pindyck, 2001, p.20). As a consequence, we adopt the usual practice of considering the price of the nearest active futures contract as a proxy for the spot price. Convenience yields are calculated using the cost-of-carry relationship (4). In order to approximate financing costs, risk-free interest rates with corresponding maturities are required. Since Brent oil is invoiced in U.S. dollars, the money market rates ought to refer to this currency as well.[17] Hence, we use the U.S. dollar rates which are offered at the London Interbank Market (LIBOR). Estimates of the T^*-month convenience yield are given by

$$\psi^*(t, T-1) = [1 + r^*(t, T)]f(t, 1) - f(t, T), \quad T = 2, 3, ..., 12, \tag{11}$$

with $r^*(t, T) = [Tr(t, T) - r(t, 1)]/12$ where $r(t, T)$ is the annual T-month money market rate in period t.[18] Owing to $\psi^*(t) = \psi^*(t, 1)$, marginal convenience yields are proxied by the nearest (1-month) and the next-to-nearest (2-months) futures contracts.

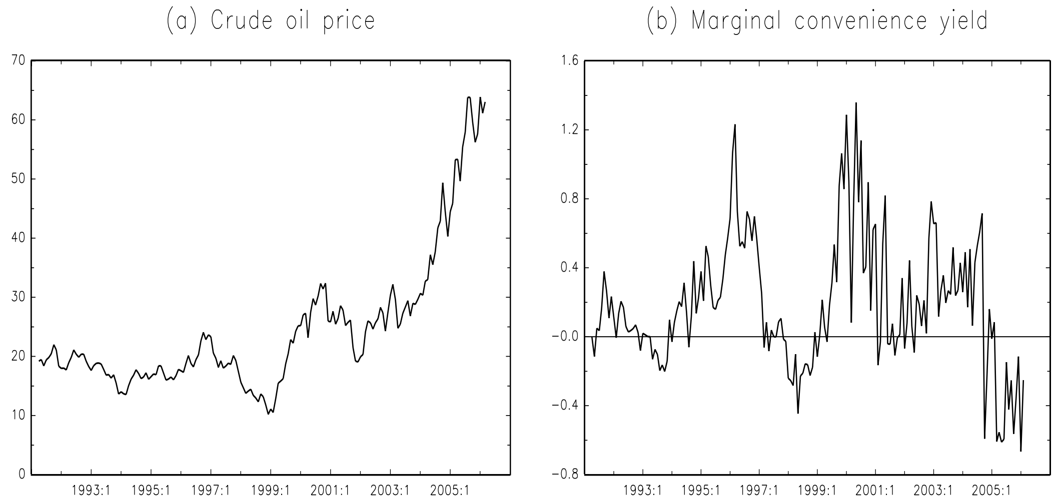
In Figure 1(a), the time series of the spot price of crude oil is plotted. At the beginning of the sample under review, the oil price stood at almost 20 U.S. dollar. After fluctating in the range between 15 and 25 U.S. dollars, the oil price bottomed out by the end of 1998. Then, it rose steadily for almost two years. Before once again taking off at the beginning of 2004, the oil price varied in the range between 20 and 30 U.S. dollars. In August/September 2005 as well as in January 2006, it surged to historical highs, leading to monthly averages of almost 64 U.S. dollars. In the sample at hand, the price of crude oil shows characteristics of a nonstationary series.

[16] Since the Newey-West estimator applies weights of less than unity to all covariances, it is appropriate to incorporate more lags than implied by the known MA order of the residual process. Our choice is inspired by a suggestion analyzed in Nelson and Kim (1993).

[17] If the interest rates for any other currencies (e.g. euro) were used, the exchange rate risk would have to be taken into account in the cost-of-carry relationship.

[18] The asterisk indicates that convenience yields are computed on the basis of the convention that the 1-month futures price serves as a proxy for the spot price.

Figure 1: Plots



The left-hand graph plots the time series of the price for a barrel of Brent oil. The right-hand graph depicts the time series of the monthly marginal convenience yield. Both entities are measured in U.S. dollars.

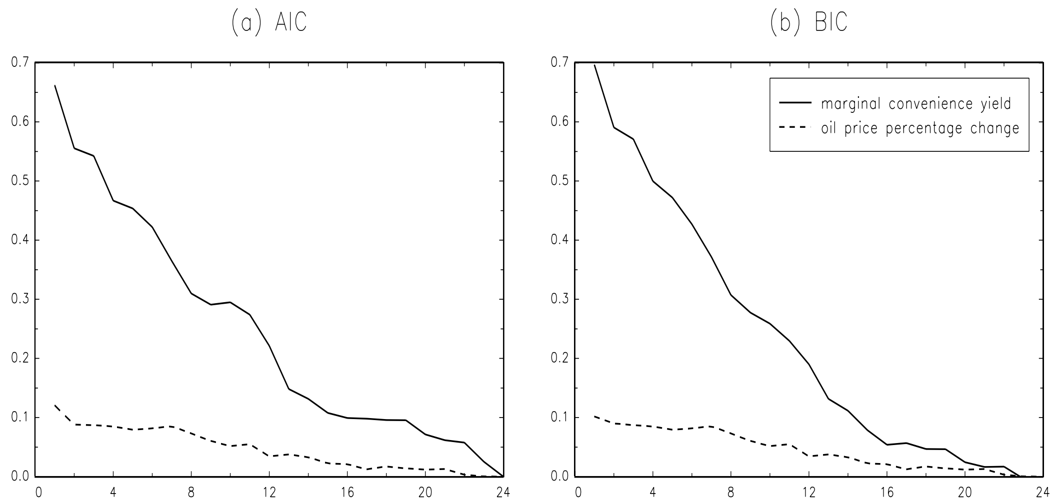
The sequence of marginal convenience yields plotted in Figure 1(b) appears to be stationary, however. Since serial correlation seems substantial, the time series might be well approximated by an AR model with a positive intercept. Hence, the marginal convenience yield exhibits characteristics of a mean-reverting process whose attractor is a positive constant. Furthermore, approximately two-thirds of the realizations lie in the positive range. This observation confirms Litzenberger and Rabinowitz's (1995) claim that the oil market is weakly backwarded in most cases. However, there are two periods where the futures market exhibits contango for some months in a row. The first was around the winter 1998/99 low where oil prices were expected to increase. But also in the second period which started in the summer of 2004 and still persists at the current juncture, market participants have predicted further increases, even though oil prices had surged to historical highs before.

4 Predictability

A prerequisite for the proposed forecasting model to perform better than a direct device is that the marginal convenience yield is more predictable than the oil price. Since the latter is nonstationary, a direct approach would likely be based on the time series of the oil price percentage change.

To measure predictability, we use the concept suggested by Diebold and Kilian (2001). For a quadratic loss function, it is defined by $P = 1 - \mathsf{E}e^2(t+j)/\mathsf{E}e^2(t+k)$ where $e(t+h)$ is the error of the best univariate h-step ahead forecast, $h = j, k$. According to this measure,

Figure 2: Predictability



The graphs depict Diebold and Kilian's (2001) predictability measure for the marginal convenience yield and the oil price percentage change. Results for the AIC and the BIC approximations of the "true" data generating processes are presented. The scale of the horizontal axis is the prediction horizon in months.

a process is the more predictable at horizon j, the smaller the squared prediction error is relative to that of a fixed benchmark horizon k. In principle, predictability is a property of the data generating process. As the latter is unknown for real time series, a parametrized approximation needs to be found and estimated. In finite samples, however, different modelling strategies do not necessarily lead to the same approximation of the true data generating processes. Hence, the predictability measure is calculated on the basis of two approximations. One uses the AIC and the other the BIC as its model selection criterion.[19] The benchmark for the predictability measure is the mean squared error of two-year ahead forecasts ($k = 24$)—a horizon which can be regarded as long in this context.

Figure 2 graphs the predictability measure for the marginal convenience yield and for the oil price percentage change. For both modelling strategies, the former variable clearly outperforms the latter in terms of predictability. Consistent with the view that the oil price is empirically almost indistinguishable from a random walk, the predictability of oil price percentage changes is no higher in the short run than in the long run. For instance, the mean squared error of a 1-step ahead forecast is only 10 per cent lower than for the benchmark. In contrast, the reduction amounts to nearly 70 per cent in the case of the marginal convenience yield. The decline in predictability is slightly stronger for the BIC than for the AIC marginal convenience yield model. This result is caused by the fact that the two criteria select distinct approximations to the data generating process. With the

[19]The specification search comprises lag length selection and sequential elimination of redundant regressors as described in Section 2.2.

BIC, a parsimonious AR(3) model is selected, while the AIC suggests to fit an AR(13) process to the series. As regards the predictability at horizons between 9 and 20 months ahead, it turns out to be advantageous to allow for a longer lag length.

It is worth thinking about the reasons why marginal convenience yields are predictable, while oil price percentage changes are not. Engel and West (2005) provide an argument which may solve this puzzle. Transferred to the present context, they claim that the oil price would manifest near random-walk behavior if the marginal convenience yield followed a unit root process and the discount factor were sufficiently close to unity. From both visual inspection of Figure 1(b) and the evidence of unit root tests,[20] it is clear that the former condition is not met. However, for the implication of their theorem to be valid, it suffices for the convenience yield to be highly persistent. At 0.90 in the BIC case, for instance, the largest root of the estimated AR polynomial takes a value for which Engel and West's simulation experiment reveals that asset price changes will only exhibit a rather weak autocorrelation pattern if the discount factor is not lower than 0.9.[21]

From the fact that Engel and West's claim might apply in this context, one should not conclude that random-walk forecasts might not be outperformed by predictions of the proposed forecasting device, thus making the latter not worth considering. The reverse is very much the case. In finite samples, econometric techniques may fail to detect autocorrelations if they are small. Hence, the likely outcome of an empirical analysis is that the random-walk hypothesis cannot be rejected. Provided that the assumed theoretical structure is true, however, the random-walk model functions as a good—but definitely not the best—approximation. With a limited set of data at hand, it might therefore be preferable to impose the explicit structure of the commodity pricing model because this ensures, at least better than a reduced-form approach, that the autocorrelations existing in theory can empirically be found and exploited in forecasting. Nonetheless, as they are proven to be small, the results of the proposed forecasting device should not be expected to differ much from random-walk forecasts. However, it is still an open issue as to whether these differences affect forecast quality.

5 Recursive modelling and forecasting exercises

The aim of this section is twofold. First, since several approaches to specifying and estimating the forecasting model have been suggested in Section 2, it remains to filter out the best performer within this set of variants. Second, it needs to be shown whether the indirect strategy is, in fact, superior to oil price forecasts performed on the basis of techniques commonly used in macroeconomic projection exercises.

[20] At the 5% level, the augmented Dickey-Fuller (ADF) test rejects the presence of a unit root in the case of the long lag order suggested by the AIC. For the parsimonious BIC choice, however, the ADF test is only able to reject the unit root hypothesis at the 10% level. With lag truncation parameters automatically selected by Newey and West's (1994) automatic procedure, the results of both the Phillips-Perron and the KPSS test clearly point to the fact that the marginal convenience yield is stationary.

[21] Taking the estimates of the oil-specific risk premium documented in the next section as given, we consider this condition to be met. For instance, discount factors used in the long evaluation period (see Section 5) lie in the range between 0.866 and 0.996 with a median of 0.904.

Seven variants of the convenience yield forecasting model [henceforth CY($\cdot$)] are considered. While the estimated oil-specific risk premium is the same, these variants differ with respect to the way the marginal convenience yield is predicted. As discussed in Section 2.1, the first simply uses market expectations. The AR and the combined variants require the specification of forecasting equations. The AIC and the BIC/HQ are applied as alternative approaches to model selection. AR forecasts can be performed by either the plug-in technique or the direct estimation method.[22] Future oil price paths, which are derived from the random-walk assumption (RWA) or the futures market hypothesis (FMH), are considered as benchmarks in the forecast evaluation.

In order to assess the forecast accuracy of the competing models, it is appropriate to carry out a recursive out-of-sample forecast exercise. This means that the forecasting equations are modelled and estimated anew at each forecast origin using all information available up to this point. The forecast accuracy of the h-step ahead oil price forecasts, $h = 1, 2, ..., 11$, is evaluated by a number of statistics. The root mean squared error (RMSE) is perhaps the most common metric. As unbiasedness is an important topic in this context, we consider the mean error (ME), too. With respect to these criteria, the forecast quality of the CY models in comparison with the benchmark devices is formally checked by means of statistical tests à la Diebold and Mariano (1995) (DM).[23] Although the proposed forecasting tool nests the benchmarks, the standard test versions appear to be applicable.[24] Moreover, it might be interesting to evaluate the probability that the direction of the oil price change is predicted correctly. Such a qualitative criterion becomes relevant when forecasters doubt the usefulness of quantitative statements on future oil prices. Pesaran and Timmermann (1992) suggest a procedure for testing whether a direction-of-change forecast is valuable in the sense of Henriksson and Merton (1981), i.e. it beats a naive coin flip. Stekler (1994) argue that the same hypothesis can be tested by using Fisher's exact test of independence in a 2×2 contingency table.

Data limitations require to separate the evaluation in two experiments. The first considers a long evaluation period starting in January 1997. The advantage of this experiment is that the forecast accuracy is checked on the basis of 100 through 110 predictions. The

[22] In order to distinguish between the seven CY variants, the upper case letter indicates the type of the forecasting model, i.e. "M" for market expectation forecasts, "P" for plug-in AR forecasts, "D" for direct AR forecasts, and "C" for combined forecasts. A lower case letter stands for the information criterion used, i.e. "a" for AIC and "b" for BIC/HQ. As an example, the plug-in AR model specified by means of the AIC is abbreviated by the acronym CY(Pa).

[23] Owing to the small samples at hand, the DM test is used in modified form as suggested by Harvey et al. (1997). In the case of the ME, the test is equivalent to a t-test of the null hypothesis that two forecast error sequences possess the same mean.

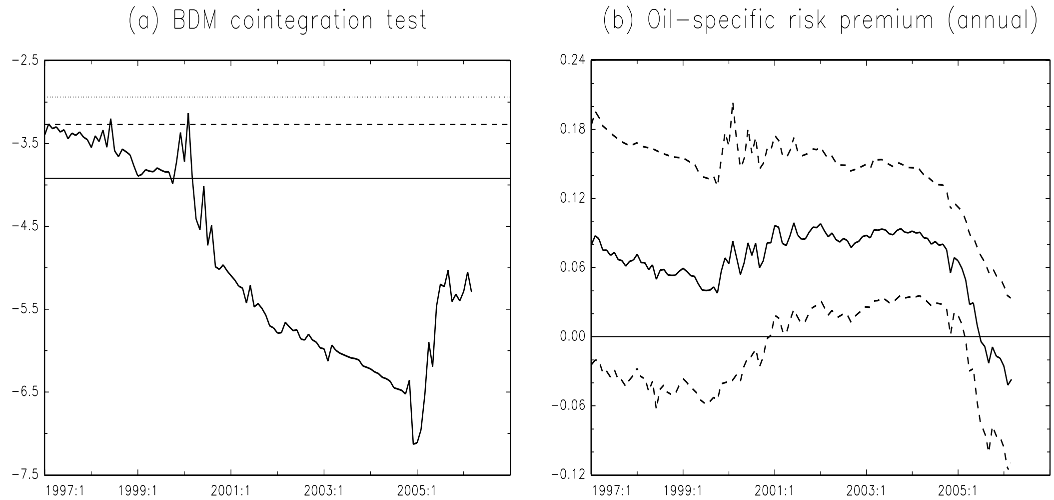
[24] Analytically, the CY model will reduce to the RWA or the FMH if $\psi(t)$ is white noise and $\mu = 0$ or $\mu = r$ respectively. Clark and McCracken (2001) and Clark and West (2006) argue that the limiting distribution of the DM statistic is degenerate in the case of nested benchmarks. The theorems presented therein, however, require (direct) least squares estimation of the unrestricted forecasting model. In the present framework, this condition may be seen as fulfilled for the coefficients of the CY models but certainly not for the oil-specific risk premium, which is superconsistently estimated by a cointegration approach. At the rate of convergence, at which the DM statistic is asymptotically normal, the risk premium can therefore be treated as fixed and known.

disadvantage is, however, that the estimation period is too short to specify combined CY models because futures prices of longer maturities have only been available since 1994. In the long evaluation period, the market expectations and the AR forecasts—performed by either the plug-in technique or the direct estimation method—are checked against the two benchmarks. In the second experiment, the post-July 2000 data is reserved for the forecast evaluation. At the cost of 42 lost predictions, this experiment allows us to examine the performance of the combined approach.

5.1 Recursive model selection and parameter estimation

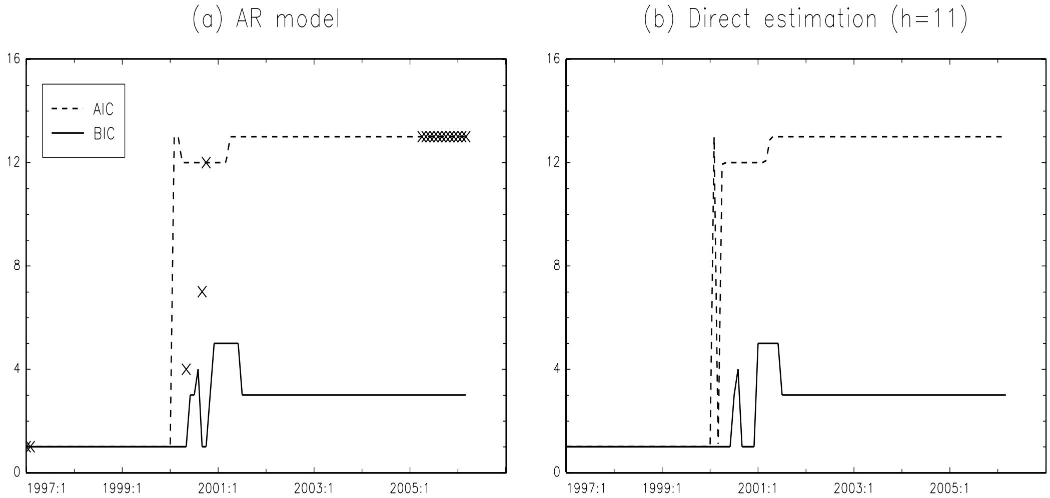
In Section 2.2, we proposed a cointegration approach to estimate the oil-specific risk premium. The idea is to set up a single-equation error correction model from which the oil-specific risk premium can be derived. Although there are convincing theoretical reasons to believe in cointegration between the spot and the futures prices of crude oil, an empirical examination of this hypothesis is useful nonetheless. As graphically depicted in Figure 3(a), the results of the BDM test show that the absence of cointegration is rejected at the 5% level in almost all samples under review. Hence, it is possible to estimate the oil-specific risk premium from the single-equation error correction model because cointegration may be well established even if the samples considered are comparably short.

Figure 3: Estimation of the oil-specific risk premium



In the graph on the left-hand side, the (non-horizontal) solid line depicts the series of the BDM cointegration test statistic carried out in samples starting in April 1991 and ending in the month indicated on the horizontal axis. The horizontal lines display the significance levels of the test; the 1% level is indicated by the solid line, the 5% level by the dashed line and the 10% level by the dotted line. In the graph on the right-hand side, the solid line depicts the recursive estimates of the oil-specific risk premium. The dashed lines indicate ±2 standard errors around the point estimates.

Figure 4: Lag order selection



The graphs plot the results of the lag selection analysis carried out in samples starting in April 1991 and ending in the month indicated on the horizontal axis. The proposed lag orders of the AR model are depicted on the left-hand side. The respective values for the multi-step forecasting equation, which is specified to perform forecasts eleven months ahead, are presented on the right-hand side. When the model selection is redone using the HQ, the chosen lag order is indicated by a "×".

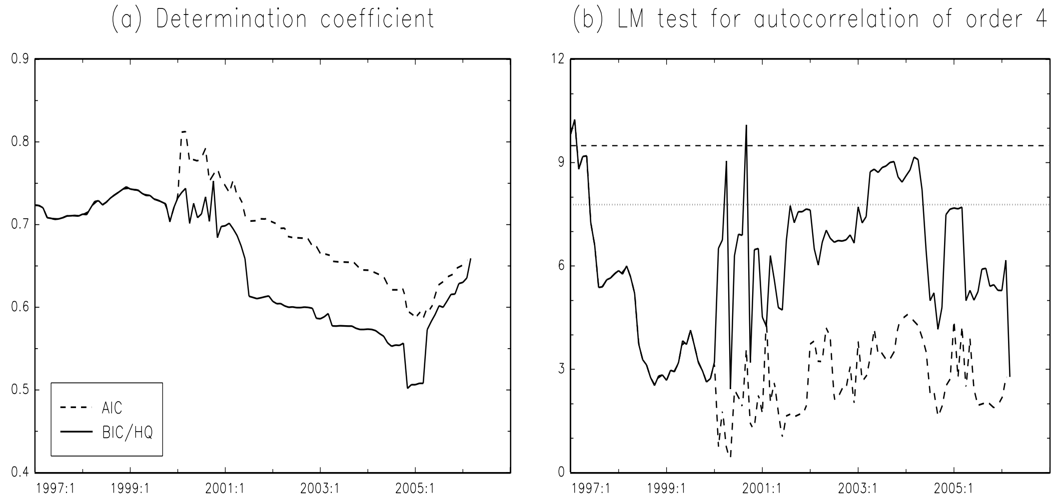
Figure 3(b) presents the recursive estimates of the annual oil-specific risk premium, suggesting that, at least until 2004, investors were indifferent between oil-market activities and holding, say, a money market deposit when the expected future earnings out of the former investment imply a rate of return which is between 6 and 10 percentage points higher than the risk-free rate. At the beginning of 2005, the risk premium started declining sharply, reaching even negative values half a year later.[25] Note that the region of lower risk premia widely corresponds with the second phase of persistently negative marginal convenience yields documented in Figure 1(b). These have obviously induced investors to downgrade the excess return required to hold crude oil instead of a risk-free asset. At the current juncture, it is not possible to judge whether this is a temporary collapse or the beginning of a new regime of low (and negative) risk premia.

The point estimates are surrounded by confidence sets of ± 2 standard errors.[26] It is worth noting that the asymptotic distribution of the estimate is not standard because it is constructed on the basis of the parameters of the cointegrating vector. Although the

[25]The negative estimates do not cause conceptual problems because they are lower than the relevant LIBOR rates in absolute values. Hence, the oil-specific discount rate is still strictly positive in the period after mid-2005.

[26]The standard errors are computed by means of the delta method on the basis of the variance-covariance matrix characterizing the joint distribution of the parameters of regression model (10).

Figure 5: Diagnostics on the AR model



The graphs plot the results of two diagnostic checks on the AR models estimated in the sequence of samples starting in April 1991 and ending in the month indicated on the horizontal axis. In the right-hand graph, the horizontal lines display χ^2 critical values of the LM test statistic; the 5% level is indicated by the dashed line and the 10% level by the dotted line.

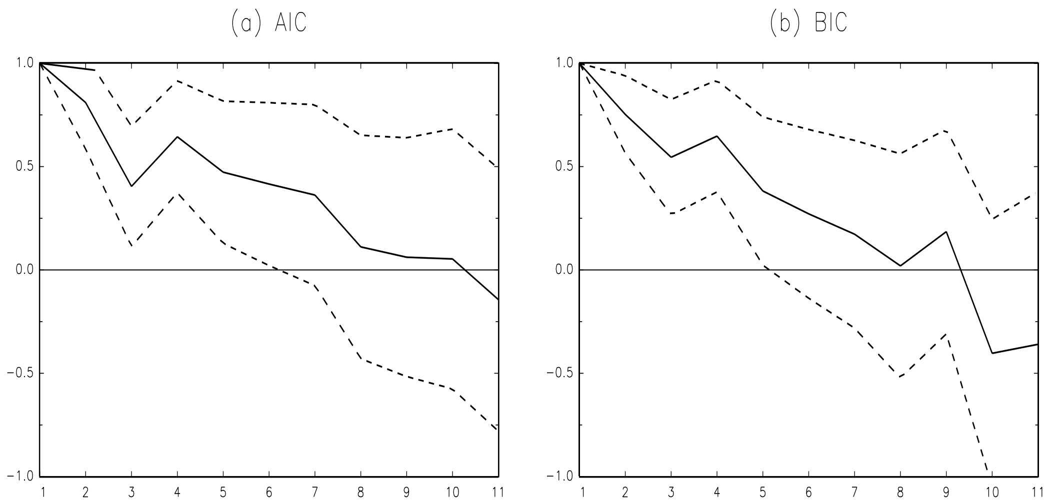
standard errors cannot be used for a formal t-test, this evidence can be rated as support for the view that the oil-specific risk premium was positive during the 1990s and in the first half of this decade.

As regards the specification of the AR model, Figure 4(a) shows that both the AIC and the BIC choose lag order 1 in samples ending in the 1997-1999 period. In the longer samples under review, however, the lag lengths differ greatly between criteria. While the AIC mostly opt for an AR(13), the BIC tends to choose an AR(3). For the latter, LM tests sometimes detect significant residual correlation. In these instances, model selection is redone using the HQ.[27] In samples with endpoints since mid-2005, the HQ mimics the AIC choice in terms of lag order selection. The final specifications may nonetheless be different because, compared with the AIC, the HQ implies a higher threshold in the following elimination procedure.

The adequacy of the final specifications is checked by some diagnostics. In Figure 5(a), we observe that both the AIC and the BIC/HQ models are able to explain a large portion of the variability of the marginal convenience yield. In the AIC case, the R^2 lies between 0.6 and 0.8. In the BIC/HQ case, the corresponding value is somewhat lower. Since the (unadjusted) determination coefficient is nondecreasing in the number of parameters, the AIC appears to be advantageous only in cases where the alternative specification is selected by the BIC (instead of the HQ). As shown in Figure 5(b), the final AR models, i.e. after a potential switch to the HQ, do not exhibit problems with residual correlation up to order 4.

[27]In Figure 4(a), the switch to the HQ is indicated by a small "×".

Figure 6: Weights in the combined forecasting model



The graphs plot the weights of the market expectation in the sequence of combined forecasting models. The weights depend on the forecast horizon which is indicated on the horizontal axis. The dashed lines show ± 2 HAC standard errors around the point estimates.

The direct estimation method requires the forecasting equation to be estimated anew at each horizon. The pattern of the lag orders selected by either information criterion does not differ markedly over forecast horizons and is quite similar to what has been found in the case of the AR models used for the plug-in technique. As an example, Figure 4(b) depicts the outcomes for the longest horizon $h = 11$. In contrast to the AR models, switches from the BIC to the HQ because of residual correlation are extremely rare. In the documented case, there is no instance at all.

In the context of the combined forecasting model, it is interesting to analyze the weight attributed to the market expectation, which may vary over the forecast horizon. The medians of the horizon-dependent weight profiles estimated out of the samples of the shorter evaluation experiment are plotted in Figure 6. At $h = 1$, the weight is unity by construction. The weight profiles do not differ recognizably between the two criteria. In general, the longer the horizon, the lower the weight of the market expectation. At $h = 11$, the weight is negative in the median of the point estimates. But already for horizons longer than half a year, the confidence sets of ± 2 HAC standard errors contain the zero line. In other words, the combined forecasts attach a significantly positive weight to the market expectation only at short horizons. In the modelling exercise, the weighting parameter is therefore included in the set of regression coefficients which are restricted to zero if their t-statistics are lower than the threshold in absolute values. Not seldom at horizons longer than six months, the combined forecasting model has a pure AR structure. In this case, the combined forecasts "degenerate" to AR forecasts performed on the basis of the direct estimation method.

Table 1: Root mean squared error

Evaluation	h	CY(M)	CY(Da)	CY(Db)	CY(Pa)	CY(Pb)	RWA	FMH
starting in Jan 97	1	2.479 [0.44 ; 0.13]	2.479 [0.44 ; 0.13]	2.479 [0.44 ; 0.13]	2.479 [0.44 ; 0.13]	2.479 [0.44 ; 0.13]	2.513 [0.89]	2.506
	2	3.508 [0.25 ; 0.09]	3.544 [0.40 ; 0.46]	3.546 [0.39 ; 0.47]	3.557 [0.53 ; 0.59]	3.537 [0.40 ; 0.35]	3.618 [0.78]	3.587
	3	4.178 [0.41 ; 0.04]	4.250 [0.74 ; 0.39]	4.234 [0.62 ; 0.25]	4.264 [0.85 ; 0.47]	4.238 [0.71 ; 0.31]	4.293 [0.74]	4.345
	4	4.683 [0.53 ; 0.05]	4.727 [0.73 ; 0.24]	4.704 [0.61 ; 0.13]	4.738 [0.79 ; 0.28]	4.682 [0.56 ; 0.14]	4.790 [0.41]	4.952
	5	5.150 [0.94 ; 0.06]	5.101 [0.87 ; 0.14]	5.060 [0.70 ; 0.07]	5.155 [0.94 ; 0.24]	5.070 [0.79 ; 0.10]	5.134 [0.16]	5.557
	6	5.961 [0.79 ; 0.08]	5.834 [0.79 ; 0.11]	5.784 [0.58 ; 0.06]	5.971 [0.81 ; 0.27]	5.857 [0.91 ; 0.13]	5.892 [0.15]	6.481
	7	6.745 [0.81 ; 0.10]	6.545 [0.59 ; 0.10]	6.486 [0.36 ; 0.06]	6.724 [0.90 ; 0.29]	6.596 [0.82 ; 0.15]	6.678 [0.16]	7.372
	8	7.457 [0.75 ; 0.12]	7.184 [0.54 ; 0.10]	7.080 [0.24 ; 0.06]	7.367 [0.98 ; 0.28]	7.258 [0.82 ; 0.17]	7.356 [0.17]	8.202
	9	8.060 [0.63 ; 0.14]	7.660 [0.46 ; 0.09]	7.540 [0.17 ; 0.06]	7.922 [0.97 ; 0.30]	7.807 [0.83 ; 0.19]	7.905 [0.17]	8.935
	10	8.630 [0.50 ; 0.16]	8.089 [0.38 ; 0.09]	7.980 [0.12 ; 0.07]	8.486 [0.88 ; 0.33]	8.357 [0.89 ; 0.21]	8.413 [0.18]	9.650
	11	9.277 [0.53 ; 0.18]	8.646 [0.28 ; 0.09]	8.555 [0.07 ; 0.09]	9.216 [0.73 ; 0.37]	9.032 [0.92 ; 0.23]	9.064 [0.22]	10.420

Evaluation	h	CY(M)	CY(Da)	CY(Db)	CY(Ca)	CY(Cb)	RWA	FMH
starting in Jul 00	1	2.862 [0.28 ; 0.18]	2.862 [0.28 ; 0.18]	2.862 [0.28 ; 0.18]	2.862 [0.28 ; 0.18]	2.862 [0.28 ; 0.18]	2.923 [0.75]	2.903
	2	3.995 [0.07 ; 0.15]	4.068 [0.18 ; 0.61]	4.070 [0.15 ; 0.62]	4.015 [0.10 ; 0.22]	4.013 [0.08 ; 0.19]	4.214 [0.50]	4.116
	3	4.619 [0.04 ; 0.07]	4.780 [0.25 ; 0.60]	4.762 [0.12 ; 0.45]	4.685 [0.10 ; 0.23]	4.674 [0.06 ; 0.12]	4.956 [0.66]	4.872
	4	4.984 [0.03 ; 0.08]	5.136 [0.22 ; 0.38]	5.111 [0.09 ; 0.23]	5.027 [0.09 ; 0.18]	5.005 [0.03 ; 0.07]	5.407 [0.99]	5.409
	5	5.247 [0.07 ; 0.09]	5.323 [0.25 ; 0.25]	5.295 [0.12 ; 0.16]	5.288 [0.16 ; 0.18]	5.254 [0.08 ; 0.09]	5.607 [0.43]	5.918
	6	6.138 [0.07 ; 0.10]	6.193 [0.25 ; 0.23]	6.184 [0.13 ; 0.17]	6.163 [0.15 ; 0.17]	6.152 [0.08 ; 0.11]	6.481 [0.32]	7.010
	7	7.008 [0.07 ; 0.12]	7.075 [0.35 ; 0.24]	7.064 [0.16 ; 0.18]	7.045 [0.26 ; 0.19]	7.025 [0.13 ; 0.13]	7.378 [0.31]	8.074
	8	7.824 [0.08 ; 0.15]	7.936 [0.47 ; 0.26]	7.859 [0.18 ; 0.20]	7.931 [0.42 ; 0.23]	7.836 [0.16 ; 0.16]	8.212 [0.31]	9.091
	9	8.463 [0.07 ; 0.15]	8.530 [0.58 ; 0.25]	8.435 [0.23 ; 0.19]	8.593 [0.64 ; 0.24]	8.466 [0.25 ; 0.17]	8.798 [0.26]	9.981
	10	9.080 [0.05 ; 0.16]	9.085 [0.68 ; 0.23]	8.940 [0.26 ; 0.19]	9.236 [0.87 ; 0.23]	9.027 [0.32 ; 0.18]	9.319 [0.23]	10.886
	11	9.864 [0.00 ; 0.17]	9.776 [0.59 ; 0.22]	9.711 [0.26 ; 0.21]	9.947 [0.79 ; 0.21]	9.777 [0.27 ; 0.20]	10.097 [0.26]	11.889

The tables report the RMSEs in U.S. dollars. The figures in brackets are p-values of modified DM tests of equal mean squared errors against a benchmark. In the CY($\cdot$) columns, the first figure in the brackets refers to the RWA and the second to the FMH chosen as the benchmark. In the RWA column, the brackets contain the p-value of a modified DM test against the FMH.

5.2 Recursive out-of-sample forecast evaluation

The recursive out-of-sample forecasts are evaluated by standard statistics of forecast accuracy. A common criterion of forecast performance is the RMSE. For model i performing h-step ahead forecasts at the sequence of origins $\tau = t_0, t_0 + 1, ..., t_1$, it is defined by

$$\text{RMSE}_i(h) = \sqrt{(t_1 - t_0 - 1)^{-1} \sum_{\tau=t_0}^{t_1} \left[p(\tau + h) - p_i^f(\tau + h)\right]^2} \tag{12}$$

where $p(\cdot)$ denotes the true oil price and $p_i^f(\cdot)$ model i's forecast.

According to the results of the first experiment reported in the upper panel of Table 1, the CY models outperform the FMH at all horizons. At very short horizons, the RWA is beaten by the FMH. Although its performance improves relative to the other models when the horizon increases, the RWA is inferior to the best CY performers over the whole range of forecast horizons. Amongst the CY models, the simple bias correction device is the best variant at short horizons,[28] while the CY(Db) shows the lowest RMSEs for predictions five months ahead and longer. The results of the modified DM tests indicate that differences in forecast accuracy between the CY models and the RWA are not statistically significant at any horizon. The only exception is the CY(Db) which significantly outperforms the RWA at 11-months ahead forecasts. At least at the 10% level, the CY(M) is significantly better than the FMH at horizons up to seven months. At longer horizons, the CY(Db) exhibits a significantly better forecast accuracy. Although the same is true for the CY(Da), it turns out that, in general, the AIC models perform worse than their BIC/HQ competitors. This may be due to the fact that the AIC models are highly parametrized, inducing a loss of predictive ability owing to the comparably large number of coefficients to be estimated.

The results of the short evaluation period documented in the lower panel of Table 1 serve as a cross-check but, predominantly, they are analyzed with a special focus on the relative performance of the combined CY models. First, it is reassuring that the general pattern found in the upper panel is not considerably changed by the fact that the evaluation period is shortened. Second, the combined variants do not perform best within the class of CY models at any horizon. Hence, the RMSE loss caused by the need to estimate the weighting parameter outweighs the benefit of using market expectations. Against the background that the CY(M) performs best at short horizons while the direct AR forecasts have advantages at longer horizons, the combined models may be regarded as compromise devices—not in the sense that pooling creates an extra benefit but owing to their ability to incorporate all approaches in a flexible manner.

The ME as a measure for the bias is of special interest in this context. It is defined by

$$\text{ME}_i(h) = (t_1 - t_0 - 1)^{-1} \sum_{\tau=t_0}^{t_1} \left[p(\tau + h) - p_i^f(\tau + h)\right]. \tag{13}$$

[28] It is worth mentioning that, owing to the construction of the marginal convenience yield, all CY variants "degenerate" to the CY(M) at the 1-step ahead forecast.

Table 2: Mean error

Evaluation	h	CY(M)	CY(Da)	CY(Db)	CY(Pa)	CY(Pb)	RWA	FMH
starting in Jan 97	1	0.310 [0.15 ; 0.00]	0.310 [0.15 ; 0.00]	0.310 [0.15 ; 0.00]	0.310 [0.15 ; 0.00]	0.310 [0.15 ; 0.00]	0.361 [0.04]	0.447
	2	0.653 [0.41 ; 0.00]	0.656 [0.39 ; 0.00]	0.662 [0.40 ; 0.00]	0.651 [0.39 ; 0.00]	0.655 [0.41 ; 0.00]	0.737 [0.09]	0.934
	3	1.087 [0.70 ; 0.00]	1.043 [0.50 ; 0.00]	1.049 [0.50 ; 0.00]	1.024 [0.48 ; 0.00]	1.044 [0.53 ; 0.00]	1.153 [0.08]	1.518
	4	1.514 [0.94 ; 0.00]	1.408 [0.61 ; 0.00]	1.414 [0.60 ; 0.00]	1.366 [0.54 ; 0.00]	1.405 [0.62 ; 0.00]	1.533 [0.06]	2.098
	5	1.932 [0.91 ; 0.00]	1.782 [0.72 ; 0.00]	1.787 [0.70 ; 0.00]	1.692 [0.58 ; 0.00]	1.754 [0.68 ; 0.00]	1.896 [0.04]	2.675
	6	2.398 [0.82 ; 0.00]	2.224 [0.83 ; 0.00]	2.220 [0.79 ; 0.00]	2.062 [0.60 ; 0.00]	2.153 [0.72 ; 0.00]	2.309 [0.04]	3.304
	7	2.899 [0.78 ; 0.00]	2.702 [0.89 ; 0.00]	2.693 [0.85 ; 0.00]	2.492 [0.63 ; 0.00]	2.618 [0.77 ; 0.00]	2.768 [0.03]	3.978
	8	3.391 [0.77 ; 0.00]	3.189 [0.94 ; 0.00]	3.162 [0.87 ; 0.00]	2.930 [0.65 ; 0.00]	3.099 [0.82 ; 0.00]	3.231 [0.03]	4.645
	9	3.835 [0.76 ; 0.00]	3.621 [0.97 ; 0.00]	3.601 [0.92 ; 0.00]	3.330 [0.66 ; 0.00]	3.558 [0.89 ; 0.00]	3.648 [0.03]	5.267
	10	4.237 [0.76 ; 0.00]	4.040 [0.99 ; 0.00]	4.008 [0.96 ; 0.00]	3.701 [0.67 ; 0.00]	3.995 [0.95 ; 0.00]	4.034 [0.04]	5.844
	11	4.592 [0.77 ; 0.00]	4.421 [0.96 ; 0.01]	4.374 [0.99 ; 0.00]	4.025 [0.67 ; 0.00]	4.389 [0.99 ; 0.00]	4.378 [0.04]	6.371

Evaluation	h	CY(M)	CY(Da)	CY(Db)	CY(Ca)	CY(Cb)	RWA	FMH
starting in Jul 00	1	0.371 [0.00 ; 0.00]	0.371 [0.00 ; 0.00]	0.371 [0.00 ; 0.00]	0.371 [0.00 ; 0.00]	0.371 [0.00 ; 0.00]	0.505 [0.13]	0.587
	2	0.746 [0.04 ; 0.00]	0.758 [0.03 ; 0.00]	0.785 [0.04 ; 0.00]	0.758 [0.04 ; 0.00]	0.782 [0.07 ; 0.00]	0.974 [0.14]	1.192
	3	1.219 [0.15 ; 0.00]	1.155 [0.06 ; 0.00]	1.219 [0.11 ; 0.00]	1.186 [0.12 ; 0.00]	1.260 [0.23 ; 0.00]	1.467 [0.08]	1.909
	4	1.670 [0.32 ; 0.00]	1.522 [0.13 ; 0.00]	1.633 [0.23 ; 0.00]	1.589 [0.26 ; 0.00]	1.719 [0.49 ; 0.00]	1.893 [0.03]	2.614
	5	2.111 [0.49 ; 0.00]	1.904 [0.22 ; 0.00]	2.029 [0.30 ; 0.00]	2.014 [0.42 ; 0.00]	2.162 [0.66 ; 0.00]	2.295 [0.01]	3.324
	6	2.717 [0.60 ; 0.00]	2.484 [0.32 ; 0.00]	2.611 [0.38 ; 0.00]	2.611 [0.54 ; 0.00]	2.773 [0.77 ; 0.00]	2.864 [0.00]	4.207
	7	3.394 [0.67 ; 0.00]	3.138 [0.42 ; 0.00]	3.302 [0.46 ; 0.00]	3.309 [0.67 ; 0.00]	3.487 [0.91 ; 0.00]	3.522 [0.00]	5.186
	8	4.038 [0.68 ; 0.00]	3.764 [0.46 ; 0.00]	3.952 [0.48 ; 0.00]	3.957 [0.72 ; 0.00]	4.141 [0.92 ; 0.00]	4.173 [0.00]	6.141
	9	4.641 [0.67 ; 0.00]	4.357 [0.50 ; 0.00]	4.585 [0.48 ; 0.00]	4.574 [0.76 ; 0.00]	4.771 [0.96 ; 0.00]	4.784 [0.00]	7.063
	10	5.190 [0.66 ; 0.00]	4.915 [0.59 ; 0.00]	5.156 [0.54 ; 0.00]	5.151 [0.82 ; 0.00]	5.321 [0.92 ; 0.00]	5.357 [0.00]	7.932
	11	5.616 [0.65 ; 0.00]	5.329 [0.62 ; 0.00]	5.614 [0.55 ; 0.00]	5.581 [0.82 ; 0.00]	5.750 [0.87 ; 0.00]	5.813 [0.00]	8.671

The tables report the MEs in U.S. dollars. The figures in brackets are p-values of t-tests for equal mean errors. In the CY($\cdot$) columns, the first figure in the brackets refers to the test against the RWA and the second to that against the FMH. In the RWA column, the brackets contain the p-value of the test against the FMH.

Owing to this definition, the positive values documented everywhere in Table 2 imply that forecasts are, on average, biased downward. To some extent, this result is an artifact of the evaluation mainly characterized by a sharply rising oil price. Focusing on their relative performance, the forecasting devices can clearly be sorted according to the bias. In the long evaluation period (upper panel), the FMH has the largest MEs at all forecast horizons, whereas the CY(Pa) comes out with the lowest. The ME differences between the CY models and the RWA are marginal overall. As a consequence, the hypothesis of equal MEs cannot be rejected in these cases. Instead, the t-tests indicate that the biases produced by the CY models are significantly smaller than those of the FMH.

The AIC models tend to produce lower biases than their BIC/HQ counterparts. Hence, specifications with longer lag lengths turn out to have advantages in comparison with more parsimonious model choices. Furthermore, the plug-in technique seem to outperform the direct estimation method in terms of the bias. Both pieces of evidence contrast to the conclusions drawn from the RMSE evaluation. The lower panel of Table 2 shows that the combined variants are inferior to the best CY performers at all horizons. This observation is generally in line with the RMSE results.

Finally, Table 3 presents the relative frequencies with which the forecasting devices are able to predict the direction of change. Formally, the statistic

$$\mathrm{DC}_i(h) = \mathrm{Prob}\Big\{ \big[p(\tau+h) - p(\tau)\big]\big[p_i^f(\tau+h) - p(\tau)\big] > 0 \Big\} \tag{14}$$

is considered.[29] A figure larger than one half means that the forecasting device performs better than flipping a coin.

In the upper panel, the FMH does not come up with a relative frequency exceeding 0.5 at any horizon. At short horizons, all CY models succeed more frequently than they fail to predict the direction of change. While the CY(M) as well as the plug-in AR forecasts seem to lose this property at longer horizons, the relative frequencies documented for the direct AR forecasts tend to increase over the forecast horizon, even reaching values slightly below 0.7 for 10- and 11-months ahead predictions. Furthermore, at horizons longer than half a year, both the Pesaran-Timmermann tests and Fisher's exact tests reject the hypothesis that the direction-of-change predictions of these CY variants are equal to a naive coin flip at conventional significance levels.

The results documented in the lower panel confirm the view that the CY models are able to predict the direction of change, while the FMH is not. At every horizon under review, the relative frequencies of all CY variants are clearly larger than 0.5, mostly between 0.6 and 0.7. Although less obvious than in the upper panel, the direct AR forecasts tend to show the best performance overall. With very few exceptions at individual horizons, the CY models are significantly better than a naive coin flip.

Some general conclusions can be drawn from the recursive out-of-sample forecasting exercises. Most importantly, it is never optimal to consider the futures price as a direct predictor of the future spot price for crude oil. The use of futures market information in the

[29] As the RWA assumes no change, the direction-of-change statistic is meaningless.

Table 3: Direction of change

Evaluation	h	CY(M)	CY(Da)	CY(Db)	CY(Pa)	CY(Pb)	FMH
starting in Jan 97	1	0.518 [0.47 ; 0.16]	0.518 [0.47 ; 0.16]	0.518 [0.47 ; 0.16]	0.518 [0.47 ; 0.16]	0.518 [0.47 ; 0.16]	0.500
	2	0.523 [0.52 ; 0.16]	0.541 [0.36 ; 0.15]	0.514 [0.58 ; 0.16]	0.514 [0.58 ; 0.16]	0.532 [0.46 ; 0.16]	0.450
	3	0.528 [0.41 ; 0.15]	0.537 [0.36 ; 0.15]	0.500	0.500	0.509 [0.57 ; 0.16]	0.481
	4	0.458	0.523 [0.48 ; 0.16]	0.505 [0.53 ; 0.16]	0.495	0.477	0.393
	5	0.472	0.509 [0.63 ; 0.15]	0.500	0.491	0.491	0.387
	6	0.476	0.533 [0.37 ; 0.15]	0.505 [0.56 ; 0.16]	0.486	0.486	0.381
	7	0.529 [0.27 ; 0.13]	0.567 [0.14 ; 0.09]	0.577 [0.08 ; 0.06]	0.442	0.490	0.423
	8	0.495	0.583 [0.09 ; 0.07]	0.583 [0.08 ; 0.06]	0.485	0.466	0.417
	9	0.500	0.608 [0.04 ; 0.03]	0.588 [0.06 ; 0.05]	0.480	0.471	0.431
	10	0.495	0.653 [0.00 ; 0.00]	0.693 [0.00 ; 0.00]	0.465	0.505 [0.55 ; 0.16]	0.416
	11	0.470	0.680 [0.00 ; 0.00]	0.680 [0.00 ; 0.00]	0.460	0.490	0.420

Evaluation	h	CY(M)	CY(Da)	CY(Db)	CY(Ca)	CY(Cb)	FMH
starting in Jul 00	1	0.618 [0.08 ; 0.08]	0.618 [0.08 ; 0.08]	0.618 [0.08 ; 0.08]	0.618 [0.08 ; 0.08]	0.618 [0.08 ; 0.08]	0.515 [0.21 ; 0.15]
	2	0.642 [0.05 ; 0.05]	0.627 [0.05 ; 0.06]	0.612 [0.10 ; 0.09]	0.627 [0.07 ; 0.07]	0.642 [0.03 ; 0.04]	0.433
	3	0.667 [0.01 ; 0.02]	0.591 [0.15 ; 0.12]	0.606 [0.18 ; 0.14]	0.636 [0.04 ; 0.05]	0.636 [0.02 ; 0.03]	0.515 [0.04 ; 0.06]
	4	0.646 [0.03 ; 0.03]	0.631 [0.06 ; 0.06]	0.569 [0.45 ; 0.22]	0.646 [0.04 ; 0.05]	0.631 [0.03 ; 0.03]	0.415
	5	0.656 [0.02 ; 0.03]	0.641 [0.05 ; 0.06]	0.688 [0.01 ; 0.02]	0.609 [0.10 ; 0.10]	0.641 [0.02 ; 0.03]	0.406
	6	0.635 [0.03 ; 0.04]	0.635 [0.05 ; 0.06]	0.651 [0.06 ; 0.07]	0.635 [0.03 ; 0.04]	0.587 [0.08 ; 0.08]	0.413
	7	0.661 [0.01 ; 0.02]	0.661 [0.02 ; 0.03]	0.661 [0.03 ; 0.04]	0.645 [0.02 ; 0.03]	0.613 [0.04 ; 0.04]	0.452
	8	0.672 [0.01 ; 0.02]	0.656 [0.03 ; 0.05]	0.705 [0.01 ; 0.01]	0.607 [0.07 ; 0.08]	0.656 [0.01 ; 0.01]	0.426
	9	0.617 [0.09 ; 0.09]	0.650 [0.04 ; 0.05]	0.667 [0.04 ; 0.05]	0.617 [0.06 ; 0.07]	0.600 [0.09 ; 0.09]	0.417
	10	0.593 [0.21 ; 0.17]	0.678 [0.01 ; 0.02]	0.678 [0.02 ; 0.04]	0.610 [0.07 ; 0.08]	0.627 [0.03 ; 0.04]	0.339
	11	0.638 [0.07 ; 0.08]	0.655 [0.03 ; 0.04]	0.621 [0.10 ; 0.10]	0.603 [0.10 ; 0.10]	0.638 [0.05 ; 0.06]	0.345

The tables report the relative frequencies of a correct prediction. When 0.5 is exceeded, the Pesaran-Timmermann test and Fisher's exact test are employed in order to test whether the direction-of-change prediction is "valuable". The p-value of the Pesaran-Timmermann test is the first figure and the p-value of Fisher's exact test is the second figure in brackets.

proposed setup, however, leads to a substantial improvement in forecast accuracy. The best CY performers are also unbeaten by the RWA even though differences are not statistically significant. Within the CY variants, smaller models tend to outperform larger models and the direct estimation method tends to beat the plug-in technique. In terms of the RMSE, the CY(M) performs best at short horizons, while the CY(Db) is significantly better than the FMH at longer horizons. The direct estimation method wins the forecast competition in the direction-of-change category, whereas the plug-in technique has advantages when the bias is of particular interest. Finally, the combined models do not perform best at any horizon. However, against the background that the relative advantages of the other CY models are horizon-dependent, they can be seen as good compromise devices.

6 Conclusion

The paper has argued that, in the context of crude oil, the futures price for delivery in h months should not be used as a predictor of the spot price h months ahead. In a present value model of rational commodity pricing, Pindyck (1993) established that, under the assumption of risk-averse market participants, futures prices are systematically biased. His setup forms the theoretical basis for a forecasting device which processes the futures market information more appropriately. Apart from the existence of a risk premium, the second main ingredient of the approach is the concept of convenience yields. Implicit market expectations of this entity can be drawn from futures prices and money market rates by means of the cost-of-carry relationship. As convenience yields are known to be governed by mean-reverting behavior, forecasts can be performed on the basis of an autoregressive model. Another variant combines both ideas, with the weight being estimated by a regression model. In general, the proposed approach is indirect, as it transfers the forecasting problem to the marginal convenience yield—a variable which have been proven to be more predictable than the oil price percentage change. The evaluation of recursive out-of-sample forecasts between one and eleven months ahead has shown that the proposed forecasting model outperforms the approach of using futures prices as direct predictors of future spot prices in terms of root mean squared errors and mean errors. With respect to these criteria, the random-walk model is not clearly beaten. In contrast to the latter benchmark, however, the proposed forecasting tool may provide valuable statements on the direction of change.

The relatively good performance of the autoregressive variants at the longer horizons under review might support the view that oil price forecasts benefit from imposing an attractor. An advantage of the present setup is that the attractor can be interpreted as an "equilibrium oil price level" because it depends on structural parameters such as the risk-free interest rate, the oil-specific risk premium and the mean marginal convenience yield. Consequently, the model avoids deriving the attractor from ad hoc considerations and/or pure time series techniques. Vis-à-vis the random walk hypothesis (suggesting that the equilibrium level always coincides with the currently observed value), the autoregressive convenience yield model is more flexible in the sense that oil price changes may be of a permanent *and* temporary nature.

The main objective of the analysis was to show that, although the common practice leads to biased predictors, futures prices provide helpful information. Central components of the present value model of rational commodity pricing, such as the oil-specific risk premium and the (marginal) convenience yields, can be derived from them. A cointegration approach has been used to estimate the risk premium. This methodology is fully consistent with the theoretical model, but perhaps too restrictive in practice because the risk premium is supposed to be constant over the estimation period. Forecast improvement might be envisaged by letting it be time-variant. Furthermore, although a simple univariate autoregressive model works well for forecasting the marginal convenience yield, more comprehensive specifications might further improve forecast quality. In Pindyck (1994, 2001), for instance, the marginal convenience yield (understood as the price of storage) is linked to the demand for inventories and impact factors such as the oil market volatility. In French (2005), oil inventories are modelled using a stock-adjustment mechanism known from the macroeconomic literature on inventory behavior. In addition, a non-linear response of convenience yields to inventory changes is incorporated.

Evaluating the forecast performance of structural approaches to convenience yield modelling, although insightful, would have been beyond the focus of this paper. With the same limiting clarification, the set of benchmarks has been restricted to the technical devices commonly used in macroeconomic projection exercises. It might be argued that these are actually beatable in the comparably short horizons under review. However, a more comprehensive assessment should naturally include a larger set of oil price models.

References

Akaike, Hirotsugu (1973), *Information Theory and an Extension of the Maximum Likelihood Principle*; in: Petrov, B. N. and F. Csáki (eds.), *2nd International Symposium on Information Theory*, Budapest: Académai Kiadó: 267-281.

Banerjee, Anindya, Juan J. Dolado and Ricardo Mestre (1998), *Error-Correction Mechanism Tests for Cointegration in a Single-Equation Framework*, Journal of Time Series Analysis 19, 3: 267-283.

Bhansali, Rajendra J. (1999), *Parameter Estimation and Model Selection for Multistep Prediction of a Time Series*, in: Gosh, Subir (ed.), *Aymptotics, Nonparametrics and Time Series*, New York: Marcel Dekker: 201-225.

Brennan, Michael J. (1991), *The Price of Convenience and the Valuation of Commodity Contingent Claims*, in: Lund, Diderik and Øksendal, Bernt (eds.), *Stochastic Models and Option Values*, Amsterdam et al.: North-Holland: 33-71.

Brüggemann, Ralf and Helmut Lütkepohl (2001), *Lag Selection in Subset VAR Models with an Application to a U.S. Monetary System*, in: Friedmann, Ralph, Lothar Knüppel and Helmut Lütkepohl (eds.), *Econometric Studies—A Festschrift in Honour of Joachim Frohn*, Münster: LIT: 107-128.

Campbell, John Y. and Robert F. Shiller (1987), *Cointegration and Tests of the Present Value Models*, Journal of Political Economy 95, 5: 1062-1088.

Chernenko, Sergey V., Krista B. Schwarz and Jonathan Wright (2004), *The Information Content of Forward and Futures Prices: Market Expectation and the Price of Risk*, Board of Governors of the Federal Reserve System: International Finance Discussion Paper 808.

Chinn, Menzie D., Michael LeBlanc and Olivier Coibion (2005), *The Predictive Content of Energy Futures: An Update on Petroleum, Natural Gas, Heating Oil and Gasoline*, NBER Working Paper 11033.

Clark, Todd E. and Michael W. McCracken (2001), *Test of Equal Forecast Accuracy and Encompassing for Nested Models*, Journal of Econometrics 105: 85-110.

Clark, Todd E. and Kenneth D. West (2006), *Using Out-of-Sample Mean Squared Prediction Errors to Test the Martingale Difference Hypothesis*, Journal of Econometrics, forthcoming.

Clements, Michael P. and David F. Hendry (1998), *Forecasting Economic Time Series*, Cambridge et al.: Cambridge University Press.

Coimbra, Carlos and Paulo Soares Esteves (2004), *Oil Price Assumptions in Macroeconomic Forecasts: Should We Follow Futures Market Expectations?*, OPEC Review 28, 2: 87-106.

Considine, Timothy J. and Donald F. Larson (2001), *Risk Premiums on Inventory Assets: The Case of Crude Oil and Natural Gas*, Journal of Futures Markets 21, 2: 109-126.

Diebold, Francis X. (1989), *Forecast Combination and Encompassing: Reconciling Two Divergent Literatures*, International Journal of Forecasting 5: 589-592.

Diebold, Francis X. and Lutz Kilian (2001), *Measuring Predictability: Theory and Macroeconomic Applications*, Journal of Applied Econometrics 16: 657-669.

Diebold, Francis X. and Roberto S. Mariano (1995), *Comparing Predictive Accuracy*, Journal of Business and Economic Statistics 13, 3: 253-263.

Dixit, Avinash K. and Robert S. Pindyck (1994), *Investment under Uncertainty*, Princeton and New Jersey: Princeton University Press.

Engel, Charles and Kenneth D. West (2005), *Exchange Rates and Fundamentals*, Journal of Political Economy 113, 2: 485-517.

Findley, David F. (1983), *On the Use of Multiple Models for Multi-Period Forecasting*, Proceedings of Business and Economic Statistics, American Statistical Association: 528-531.

French, Mark W. (2005), *Why and When do Spot Prices of Crude Oil Revert to Futures Price Levels?*, Board of Governors of the Federal Reserve System: Finance and Economics Discussion Paper 2005-30.

Gibson, Rajna and Eduardo S. Schwartz (1990), *Stochastic Convenience Yield and the Pricing of Oil Contingent Claims*, Journal of Finance 45, 3: 959-976.

Hannan, E.J. and B.G. Quinn (1979), *The Determination of the Order of an Autoregression*, Journal of the Royal Statistical Society B41: 190-195.

Harvey, David, Stephen Leybourne and Paul Newbold (1997), *Testing the Equality of Prediction Mean Squared Errors*, International Journal of Forecasting 13: 281-291.

Henriksson, Roy D. and Robert C. Merton (1981), *On Market Timing and Investment Performance. II. Statistical Procedures for Evaluating Forecasting Skills*, Journal of Business 54, 4: 513-533.

Johansen, Søren (1991), *Estimation and Testing of Cointegration Vectors in Gaussian Vector Autoregressive Models*, Econometrica 59, 6: 1551-1580.

Johansen, Søren (1992), *Cointegration in Partial Systems and the Efficiency of Single-Equation Analysis*, Journal of Econometrics 52: 389-402.

Johnston, H. N. (1974), *A Note on the Estimation and Prediction Inefficiency of "Dynamic" Estimators*, International Economic Review 15, 1: 251-255.

Litzenberger, Robert H. and Nir Rabinowitz (1995), *Backwardation in Oil Futures Markets: Theory and Empirical Evidence*, Journal of Finance 50, 5: 1517-1545.

Marcellino, Massimiliano, James H. Stock and Mark W. Watson (2006), *A Comparison of Direct and Iterated Multistep AR Methods for Forecasting Macroeconomic Time Series*, Journal of Econometrics, forthcoming.

Moosa, Imad A. and Nabeel E. Al-Loughani (1994), *Unbiasedness and Time Varying Risk Premia in Crude Oil Futures Markets*, Energy Journal 16, 2: 99-105.

Nelson, Charles R. and Myung J. Kim (1993), *Predictable Stock Returns: The Role of Small Sample Bias*, Journal of Finance 48, 2: 642-661.

Newey, Whitney K. and Kenneth D. West (1987), *A Simple, Positive Semi-Definite, Heteroskedasticity and Autocorrelation Consistent Covariance Matrix*, Econometrica 55, 3: 703-708.

Newey, Whitney K. and Kenneth D. West (1994), *Automatic Lag Selection in Covariance Matrix Estimation*, Review of Economic Studies 61: 631-653.

Pesaran, M. Hashem and Allan Timmermann (1992), *A Simple Nonparametric Test of Predictive Performance*, Journal of Business and Economic Statistics 10, 4: 461-465.

Pindyck, Robert S. (1993), *The Present Value Model of Rational Commodity Pricing*, Economic Journal 103: 501-530.

Pindyck, Robert S. (1994), *Inventories and the Short-Run Dynamics of Commodity Prices*, RAND Journal of Economics 25, 1: 141-159.

Pindyck, Robert S. (2001), *The Dynamics of Commodity Spot and Futures Markets: A Primer*, Energy Journal 22, 3: 1-29.

Schwarz, Gideon (1978), *Estimating the Dimension of a Model*, Annals of Statistics 6: 461-464.

Stekler, H.O. (1994), *Are Economic Forecasts Valuable?*, Journal of Forecasting 13: 495-505.

Weiss, Andrew A. (1991), *Multi-Step Estimation and Forecasting in Dynamic Models*, Journal of Econometrics 48: 135-149.

The following Discussion Papers have been published since 2005:

Series 1: Economic Studies

1	2005	Financial constraints and capacity adjustment in the United Kingdom – Evidence from a large panel of survey data	Ulf von Kalckreuth Emma Murphy
2	2005	Common stationary and non-stationary factors in the euro area analyzed in a large-scale factor model	Sandra Eickmeier
3	2005	Financial intermediaries, markets, and growth	F. Fecht, K. Huang, A. Martin
4	2005	The New Keynesian Phillips Curve in Europe: does it fit or does it fail?	Peter Tillmann
5	2005	Taxes and the financial structure of German inward FDI	Fred Ramb A. J. Weichenrieder
6	2005	International diversification at home and abroad	Fang Cai Francis E. Warnock
7	2005	Multinational enterprises, international trade, and productivity growth: Firm-level evidence from the United States	Wolfgang Keller Steven R. Yeaple
8	2005	Location choice and employment decisions: a comparison of German and Swedish multinationals	S. O. Becker, K. Ekholm, R. Jäckle, M.-A. Muendler
9	2005	Business cycles and FDI: evidence from German sectoral data	Claudia M. Buch Alexander Lipponer
10	2005	Multinational firms, exclusivity, and the degree of backward linkages	Ping Lin Kamal Saggi

11	2005	Firm-level evidence on international stock market comovement	Robin Brooks Marco Del Negro
12	2005	The determinants of intra-firm trade: in search for export-import magnification effects	Peter Egger Michael Pfaffermayr
13	2005	Foreign direct investment, spillovers and absorptive capacity: evidence from quantile regressions	Sourafel Girma Holger Görg
14	2005	Learning on the quick and cheap: gains from trade through imported expertise	James R. Markusen Thomas F. Rutherford
15	2005	Discriminatory auctions with seller discretion: evidence from German treasury auctions	Jörg Rocholl
16	2005	Consumption, wealth and business cycles: why is Germany different?	B. Hamburg, M. Hoffmann, J. Keller
17	2005	Tax incentives and the location of FDI: evidence from a panel of German multinationals	Thiess Buettner Martin Ruf
18	2005	Monetary Disequilibria and the Euro/Dollar Exchange Rate	Dieter Nautz Karsten Ruth
19	2005	Berechnung trendbereinigter Indikatoren für Deutschland mit Hilfe von Filterverfahren	Stefan Stamfort
20	2005	How synchronized are central and east European economies with the euro area? Evidence from a structural factor model	Sandra Eickmeier Jörg Breitung
21	2005	Asymptotic distribution of linear unbiased estimators in the presence of heavy-tailed stochastic regressors and residuals	J.-R. Kurz-Kim S.T. Rachev G. Samorodnitsky

22	2005	The Role of Contracting Schemes for the Welfare Costs of Nominal Rigidities over the Business Cycle	Matthias Paustian
23	2005	The cross-sectional dynamics of German business cycles: a bird's eye view	J. Döpke, M. Funke S. Holly, S. Weber
24	2005	Forecasting German GDP using alternative factor models based on large datasets	Christian Schumacher
25	2005	Time-dependent or state-dependent price setting? – micro-evidence from German metal-working industries –	Harald Stahl
26	2005	Money demand and macroeconomic uncertainty	Claus Greiber Wolfgang Lemke
27	2005	In search of distress risk	J. Y. Campbell, J. Hilscher, J. Szilagyi
28	2005	Recursive robust estimation and control without commitment	Lars Peter Hansen Thomas J. Sargent
29	2005	Asset pricing implications of Pareto optimality with private information	N. R. Kocherlakota Luigi Pistaferri
30	2005	Ultra high frequency volatility estimation with dependent microstructure noise	Y. Aït-Sahalia, P. A. Mykland, L. Zhang
31	2005	Umstellung der deutschen VGR auf Vorjahres-preisbasis – Konzept und Konsequenzen für die aktuelle Wirtschaftsanalyse sowie die ökono-metrische Modellierung	Karl-Heinz Tödter

32	2005	Determinants of current account developments in the central and east European EU member states – consequences for the enlargement of the euro erea	Sabine Herrmann Axel Jochem
33	2005	An estimated DSGE model for the German economy within the euro area	Ernest Pytlarczyk
34	2005	Rational inattention: a research agenda	Christopher A. Sims
35	2005	Monetary policy with model uncertainty: distribution forecast targeting	Lars E.O. Svensson Noah Williams
36	2005	Comparing the value revelance of R&D reporting in Germany: standard and selection effects	Fred Ramb Markus Reitzig
37	2005	European inflation expectations dynamics	J. Döpke, J. Dovern U. Fritsche, J. Slacalek
38	2005	Dynamic factor models	Sandra Eickmeier Jörg Breitung
39	2005	Short-run and long-run comovement of GDP and some expenditure aggregates in Germany, France and Italy	Thomas A. Knetsch
40	2005	A“wreckers theory” of financial distress	Ulf von Kalckreuth
41	2005	Trade balances of the central and east European EU member states and the role of foreign direct investment	Sabine Herrmann Axel Jochem
42	2005	Unit roots and cointegration in panels	Jörg Breitung M. Hashem Pesaran
43	2005	Price setting in German manufacturing: new evidence from new survey data	Harald Stahl

1	2006	The dynamic relationship between the Euro overnight rate, the ECB's policy rate and the term spread	Dieter Nautz Christian J. Offermanns
2	2006	Sticky prices in the euro area: a summary of new micro evidence	Álvarez, Dhyne, Hoeberichts Kwapil, Le Bihan, Lünnemann Martins, Sabbatini, Stahl Vermeulen, Vilmunen
3	2006	Going multinational: What are the effects on home market performance?	Robert Jäckle
4	2006	Exports versus FDI in German manufacturing: firm performance and participation in international markets	Jens Matthias Arnold Katrin Hussinger
5	2006	A disaggregated framework for the analysis of structural developments in public finances	Kremer, Braz, Brosens Langenus, Momigliano Spolander
6	2006	Bond pricing when the short term interest rate follows a threshold process	Wolfgang Lemke Theofanis Archontakis
7	2006	Has the impact of key determinants of German exports changed? Results from estimations of Germany's intra euro-area and extra euro-area exports	Kerstin Stahn
8	2006	The coordination channel of foreign exchange intervention: a nonlinear microstructural analysis	Stefan Reitz Mark P. Taylor
9	2006	Capital, labour and productivity: What role do they play in the potential GDP weakness of France, Germany and Italy?	Antonio Bassanetti Jörg Döpke, Roberto Torrini Roberta Zizza
10	2006	Real-time macroeconomic data and ex ante predictability of stock returns	J. Döpke, D. Hartmann C. Pierdzioch

11	2006	The role of real wage rigidity and labor market frictions for unemployment and inflation dynamics	Kai Christoffel Tobias Linzert
12	2006	Forecasting the price of crude oil via convenience yield predictions	Thomas A. Knetsch

Series 2: Banking and Financial Studies

1	2005	Measurement matters – Input price proxies and bank efficiency in Germany	Michael Koetter
2	2005	The supervisor's portfolio: the market price risk of German banks from 2001 to 2003 – Analysis and models for risk aggregation	Christoph Memmel Carsten Wehn
3	2005	Do banks diversify loan portfolios? A tentative answer based on individual bank loan portfolios	Andreas Kamp Andreas Pfingsten Daniel Porath
4	2005	Banks, markets, and efficiency	F. Fecht, A. Martin
5	2005	The forecast ability of risk-neutral densities of foreign exchange	Ben Craig Joachim Keller
6	2005	Cyclical implications of minimum capital requirements	Frank Heid
7	2005	Banks' regulatory capital buffer and the business cycle: evidence for German savings and cooperative banks	Stéphanie Stolz Michael Wedow
8	2005	German bank lending to industrial and non-industrial countries: driven by fundamentals or different treatment?	Thorsten Nestmann
9	2005	Accounting for distress in bank mergers	M. Koetter, J. Bos, F. Heid C. Kool, J. Kolari, D. Porath
10	2005	The eurosystem money market auctions: a banking perspective	Nikolaus Bartzsch Ben Craig, Falko Fecht
11	2005	Financial integration and systemic risk	Falko Fecht Hans Peter Grüner

12	2005	Evaluating the German bank merger wave	Michael Koetter
13	2005	Incorporating prediction and estimation risk in point-in-time credit portfolio models	A. Hamerle, M. Knapp, T. Liebig, N. Wildenauer
14	2005	Time series properties of a rating system based on financial ratios	U. Krüger, M. Stötzel, S. Trück
15	2005	Inefficient or just different? Effects of heterogeneity on bank efficiency scores	J. Bos, F. Heid, M. Koetter, J. Kolatri, C. Kool
01	2006	Forecasting stock market volatility with macroeconomic variables in real time	J. Döpke, D. Hartmann C. Pierdzioch
02	2006	Finance and growth in a bank-based economy: is it quantity or quality that matters?	Michael Koetter Michael Wedow
03	2006	Measuring business sector concentration by an infection model	Klaus Düllmann

Visiting researcher at the Deutsche Bundesbank

The Deutsche Bundesbank in Frankfurt is looking for a visiting researcher. Among others under certain conditions visiting researchers have access to a wide range of data in the Bundesbank. They include micro data on firms and banks not available in the public. Visitors should prepare a research project during their stay at the Bundesbank. Candidates must hold a Ph D and be engaged in the field of either macroeconomics and monetary economics, financial markets or international economics. Proposed research projects should be from these fields. The visiting term will be from 3 to 6 months. Salary is commensurate with experience.

Applicants are requested to send a CV, copies of recent papers, letters of reference and a proposal for a research project to:

Deutsche Bundesbank
Personalabteilung
Wilhelm-Epstein-Str. 14

D - 60431 Frankfurt
GERMANY