

Forecastability and statistical characteristics of aggregate oil and gas investments on the Norwegian Continental Shelf¹

Sindre Lorentzen, University of Stavanger
4036 Stavanger, Norway
e-mail: sindre.lorentzen@uis.no

Petter Osmundsen, University of Stavanger,
4036 Stavanger, Norway
e-mail: petter.osmundsen@uis.no

Abstract

We investigate the potential for statistical forecasting of aggregate oil and gas investment on the Norwegian Continental Shelf (NCS). A unique and detailed dataset containing data from 109 different fields on the NCS between 1970 and 2015 is employed. A set of 1080 autoregressive distributed lag models are evaluated pseudo out-of-sample and tested for data mining by utilizing a Diebold-Mariano hypothesis test and the model confidence set procedure by Hansen and Lunde (2011). The main results are as follows. First, we find that it is indeed possible but challenging to outperform the parsimonious random walk benchmark in an out-of-sample environment. Second, lags of investment growth, crude oil price growth and realized volatility are found to be adequate predictors of the investment growth. Finally, there is a clear benefit from re-estimating the model's coefficient at every step.

Keywords: Investment, oil and gas sector, Norwegian Continental Shelf, pseudo out-of-sample forecasting.

¹ Thanks are due to Anders Toft at the Norwegian Petroleum Directorate for useful comments and suggestions.

1.0 Introduction

This paper attempts to forecast future aggregate oil and gas investments on the Norwegian Continental Shelf (NCS). The ability to predict future movements in aggregate investment might prove useful for the government in their decision-making. The petroleum sector exerts a prevalent influence over the Norwegian economy. Thus, cyclicalities in oil and gas investments could be a denominator of the overall domestic business cycle. From a governmental point of view, it is of great interest to employ measures to smoothen the business cycles; however, the timing of such measures is crucial. For instance, implementing measures aimed at halting a business boom might prove ill-advised if the economy on its own accord is about to revert into a downturn. As such, providing a model capable of forecasting future movements in investment could prove helpful in the timing of measures aimed at smoothing the cycles. In Norway, oil investment predictions are also necessary for budgetary purposes as the Norwegian state has a direct interest in many of the fields, via the fully state owned company Petoro.

Previous research has been dedicated to understand and forecast future investments, especially corporate investments. However, investments in the oil and gas sector exhibit several important characteristics (Bhattacharyya, 2011) that set them apart from other types of investments. Capital intensiveness: petroleum projects tend to be extensively capital intensive. Asset specificity: assets utilized in the projects tend to be idiosyncratic and thus have little alternative usage, which consequently increases the risk of the project. Long-life of assets: the projects have a long duration, and as the project will operate long into the future it becomes challenging to forecast the associated costs and benefits. Long gestation period: oil and gas projects tend to require a longer period to execute, thus making them more sensitive to the business environment. It can therefore be argued that the presence of these characteristics warrants an investigation of the forecastability of investments from the oil and gas sector specifically.

There are several broad approaches to forecasting. As pointed out by Hamilton (2009), predicting the future is frequently based on a statistical approach, economic theory or inspecting the fundamental denominators of supply and demand. The various approaches are often regarded as substitutes or competing methodologies serving the same purpose. However, it might be more prudent to consider these as complementary. Arguably, several forecasting techniques should be combined to achieve superior insight. Nevertheless, while valuable insight could potentially be gained from all methodologies, this study will primarily pursue the statistical approach in the sense that the future is deterministically derived from correlations and relationships gleaned from historical data by applying time series models. Time series models are typically considered to be atheoretical. That is, the estimated coefficients in these models are used to forecast the variables of interest, but unlike typical cross-sectional studies, we do not attempt to assign a causal interpretation. The selection of variables or predictors in the time series model will be based on expert knowledge or theoretical considerations, but the overall model itself is not necessarily indicative of any causal relationships among the predictors and the variable to be predicted. With this as a point of departure, there are several issues of paramount contention in the choice of methodology for statistical forecasting and forecast evaluation.

First, a potential issue is whether to forecast the aggregate investment or aggregate the forecasts of the components of investment. From a theoretical perspective, the latter might yield superior forecast as data on a disaggregate level could be more informative. On the other hand, as the level of disaggregation increases the amount of noise is likely to increase. Consequently, whether to use the former or latter approach to forecasting the aggregate investment is an empirical matter (Hendry and Hubrich, 2012; Grunfeld and Griliches 1960; Kohn, 1982; Giacomini and Granger, 2004).

Second, the measure of the forecast accuracy is a prevalent issue. Typically, a model's forecasting accuracy is based on either economic or statistical loss functions. The former utilizes the end user's actual utility function, while the latter measures forecast accuracy as the weighted sum of the forecast error (e_t), i.e. the difference between the observed (y_t) and predicted (\hat{y}_t) values of the variable of interest, $e_t = L(y_t - \hat{y}_t)$. An example of the former would be a trader on the stock exchange deciding to sell or buy a particular share based on a model's forecast – the profit gained by using this strategy and model would be an economic loss function. An economic loss function is widely considered to be the best measure of a model forecast accuracy as it is directly contingent on the end user's goal. However, the downside is that explicit knowledge about the user's utility function is required, which is seldom the case, and that a model's accuracy becomes dependent on the intended usage. Just as beauty is in the eye of the beholder, so is the model's forecast accuracy. In practice, it is easier to utilize a set of statistical loss functions, which attempt to emulate the unobservable utility function in a more simplified manner. The measure of the forecast accuracy becomes more generalized, but at the expense of potentially becoming arbitrary. An example voiced by Taleb (2007) is that using Value at Risk (VaR) that is correctly predicting the loss on a portfolio 95 percent of the time is similar to deciding to cross a river that on average is a meter deep. In other words, a blind reliance on statistical loss functions could be disastrous.

Third, the issue of how to restrict the information set used to estimate the coefficients in the time series model is less contested, but still a paramount issue. Depending on the information set, forecasts are either made in-sample or out-of-sample, where the latter can be further disaggregated into ex ante or ex post. In-sample indicates that the same sample of data is both used to estimate the model and to perform evaluation of the model's forecast accuracy. From a practical point of view, in-sample is relatively fast and easy to implement compared to out-of-sample. However, the in-sample approach is not realistic in the sense that we essentially need to know the future in order to predict the same future. The ever increasing ability of our statistical models to fit the observed data is paradoxically a problem in the forecast literature. Following the idea behind Wold's decomposition theorem, a stationary variable can be decomposed into two components, a deterministic part and a stochastic part. When utilizing the in-sample approach, models that capture the random noise and nuisance of the past are rewarded. This is problematic as this idiosyncratic noise is of little importance and is unlikely to repeat itself in future. More formally, by allowing the sample to be both used for estimation and evaluation the problem of overfitting and data-mining becomes more prevalent. In-short, the in-sample approach tends to understate the forecasting errors and to favour models that in practice are sub-optimal. For these reasons, the well-established consensus in the literature is that the forecast accuracy of any given models ought to be assessed through out-of-sample evaluation rather than in-sample testing.

Perhaps the most prevalent issue in statistical forecasting is the selection of predictors. As a point of departure, the underlying motivation for undertaking investments is to enable the realization of benefits that outweigh the costs after controlling for factors such as risk and the value of time. What constitutes a benefit is open to interpretation, but within the context of oil and gas projects, the most obvious goal of a project is profit. Thus, anything affecting the expected time and risk adjusted profit of the projects is a potential predictor of aggregate oil and gas investment. In this paper we will investigate the effect of three central exogenous predictors on the aggregate investments, specifically, the crude oil price (Brent), the realized volatility of the crude oil price and the USD/NOK exchange rate.

Summing up, the aim of this paper is twofold. First, we attempt to identify the statistical properties of oil and gas investments on the NCS. Second, we investigate the feasibility of the statistical approach

for forecasting future aggregate oil and gas investment in Norway with an ex post (pseudo) out-of-sample forecast accuracy evaluation.

There are several studies addressing cost and activity issues at the NCS, but they are focusing on causalities, not forecasting. Exploration activity is analysed by Mohn and Osmundsen (2008, 2011) and Mohn (2008). A report by the Norwegian Petroleum Directorate (2013) considers overruns of 5 megaprojects on the Norwegian continental shelf. The findings in the report were compared to NOU (1999), a similar report produced by the Investment Committee in 1998. Drilling efficiency at the NCS was analysed in Osmundsen et al. (2010, 2012). Osmundsen et al. (2015) analyse the effect of taxation on NCS investment. Investments on the UK shelf and the effect on production is addressed in Kemp and Kasim (20013).

The remainder of this paper is structured as follows. Section 2 displays the components of investments and shows how the aggregate oil and gas investment on the NCS is computed. In section 3, we investigate the statistical properties of the aggregate investment and in Section 4 we analyse the aggregate investment growth and the predictors. Section 5 discusses and presents the utilized methodology for forecasting and evaluating the forecastability of aggregate investment growth. Further, in section 6 the results from the ex ante out-of-sample forecast evaluation of the considered models is presented and discussed. Finally, section 7 draws conclusions.

2.0 Aggregation of investment

To analyse aggregate oil and gas investments on the NCS we utilize a unique and detailed dataset extracted from the Norwegian Petroleum Directorate (NPD). Aggregate investment is here computed as the inflation adjusted aggregate sum of twelve different time series (see Table 1), based on 109 development projects, pertinent to oil and gas production on the NCS between 1970 and 2015. The consumer price index from Statistics Norway is used to adjust investments to year 2015 NOK values. The aggregate investment of the development projects consists of investment related to wells, pipelines, offshore installations and onshore facilities. The data set is on a field basis, i.e., the investments in pipelines and onshore facilities are only those directly related to individual oil and gas fields, not separate facilities.

While the aggregate investment is the variable of primary interest, it is possible that it will exhibit characteristics that differ from its base component. Furthermore, the twelve different investment types used as a basis for the aggregate are likely to possess a great deal of heterogeneity, both in behaviour and data availability. Thus, a compelling argument can be made that it is necessary to understand the underlying drivers or determinants of the base component before understanding the aggregate. However, due to differing data consistency, it becomes challenging to evaluate this data adequately beyond a descriptive level.

Based on the twelve time series for investment on the NCS, we aggregate the variables into broader subcategories in accordance with the type of investment. The considered subcategories are wells, pipelines, offshore installations and onshore facilities. Figure 1 compares the aggregate investment to the four subcategories. As observed, aggregate investment was initiated with less than one billion in 1970 before experiencing an oscillating behaviour around an upward and stable trend throughout the sample period until 2004. With an aggregate investment of 57 billion NOK in 2004, the positive trend appears to intensify as the average yearly increase in investment shifted from 1.66 bn to 7.56 bn. Between 1970 and 1989 aggregate investment appears to exclusively consist of investment in offshore

facilities, only from 1990 do the other subcategories appear. While investment in offshore facilities diverges from the aggregate investment after 1990, both variables appear to remain predominantly collinear. Investment in wells appears to mostly trend upwards with little fluctuation, whereas pipelines and onshore facilities appear to be relatively stationary.

Table 1: List of aggregate investment components

Subcategories of investment, the twelve base components of the aggregate oil and gas investment and their respective data consistency. Number of panel data observations (N), number of fields (Fields), and sample period range (Time) is reported. All data was extracted from the Norwegian Petroleum Directorate.

Subcategory:	Component:	N/Fields/Time
Investment in wells	New non mobile units	656/43/1990-2015
	New mobile units	878/84/1990-2015
	Not classified	428/66/2006-2015
Investment in offshore installations	Modifications on existing installations	1110/85/1970-2015
	New bottom conditions and other	398/50/1992-2015
	New movable units	193/26/1995-2015
	New cargo and storage	13/4/1995-2000
	Subsea installations	598/71/1995-2015
	Other constructions	350/67/2007-2015
Investment in pipelines	Offshore pipelines	540/72/1990-2015
	Onshore pipelines	16/16/2015-2015
Investment in onshore facilities	Facilities	81/14/1998-2015

Figure 1: Investment on the NCS

Inflation adjusted investment in billion NOK on the Norwegian Continental Shelf for the oil and gas sector between 1970 and 2015. Investment is displayed both on an aggregate and disaggregate level. All data was provided by the Norwegian Petroleum Directorate.

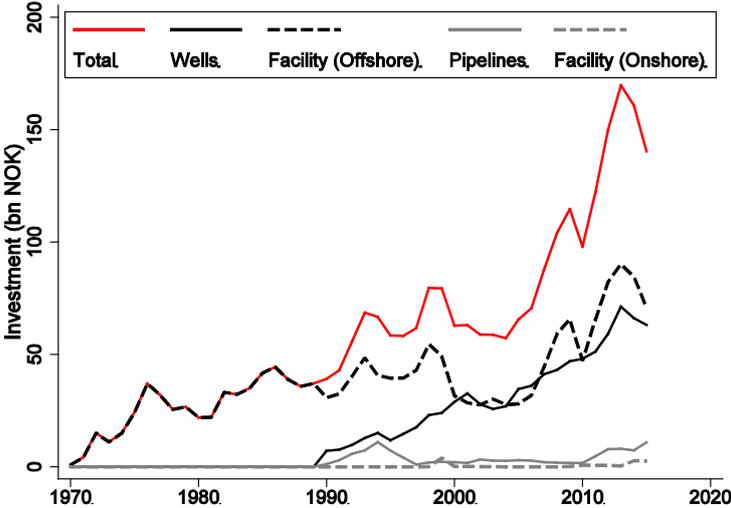


Figure 2 shows the four subcategories of the total investment and their respective base component. Subfigure (a) displays the development in investment related to offshore installations. As revealed, after a spike in the 70s, modifications of existing infrastructure has steadily increased until 2013, from a relatively low level in 1993, before seemingly plunging again. Investment in new offshore installations

exhibits a far more volatile and erratic behaviour with a non-obvious pattern. Subfigure (b) addresses investment related to wells. This type of investment appears to predominantly exhibit an upward sloping trend. Investment in pipelines, see subfigure (c), appears to spike in 1994 and 2012 while otherwise remaining on a lower and more stable level. Finally, subfigure (d) shows the development of investment in onshore facilities. With the exception of a few larger spikes, this type of investment appears to follow no particular discernible trend.

Figure 2: Disaggregated investment on the NCS

The inflation adjusted investment in billion NOK on the Norwegian Continental Shelf for the oil and gas sector between 1970 and 2015. Subfigure (a) shows investment in offshore installations, (b) investment in wells, (c) investment in pipelines both on- and offshore, and (d) onshore facilities. All data was provided by the Norwegian Petroleum Directorate.

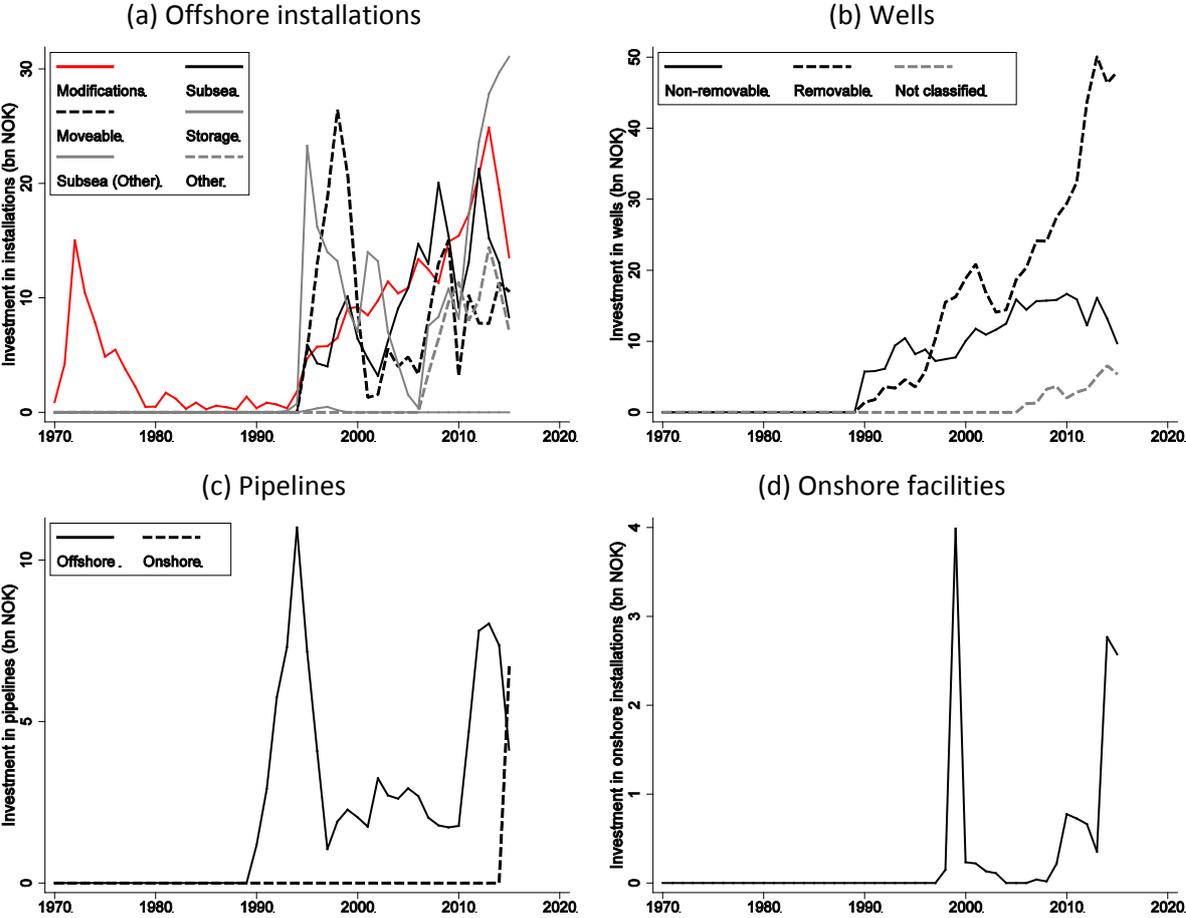
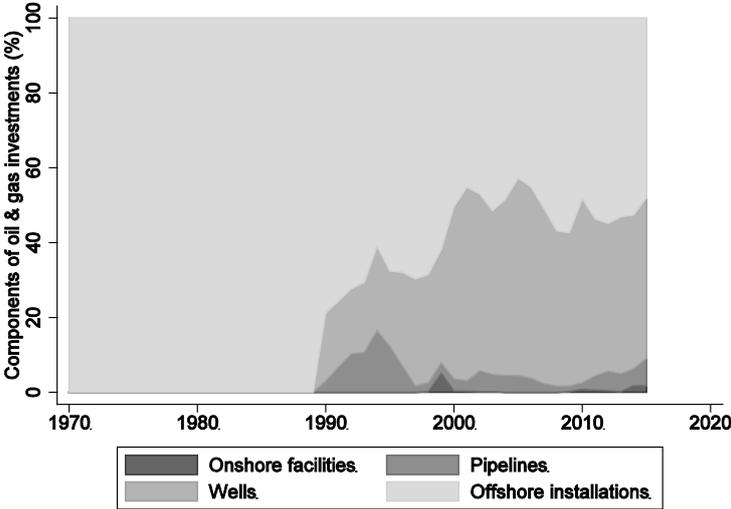


Figure 3 shows the percentage contribution from the four subcategories (wells, pipelines, offshore installations and onshore facilities) to the aggregate investment in the oil and gas sector on the NCS. Prior to 1989 investment was seemingly exclusively confined to offshore installations. This is likely due to a more crude level of details in the records. As such, it makes little sense to consider the investment disaggregated to the subcategories for the whole sample period in later econometric analysis. Nevertheless, considering the sample period from 1990 to 2015, the relative contribution of the subcategories appears to change over time, where offshore installations and wells are the most

important. In comparison, pipelines and onshore facilities have a minuscule effect on the aggregate investment.

Figure 3: **Components of oil and gas investments**

The percentage of each subcategories of the aggregate oil and gas investment on the Norwegian Continental Shelf between 1970 and 2015. All data was provided by the Norwegian Petroleum Directorate.



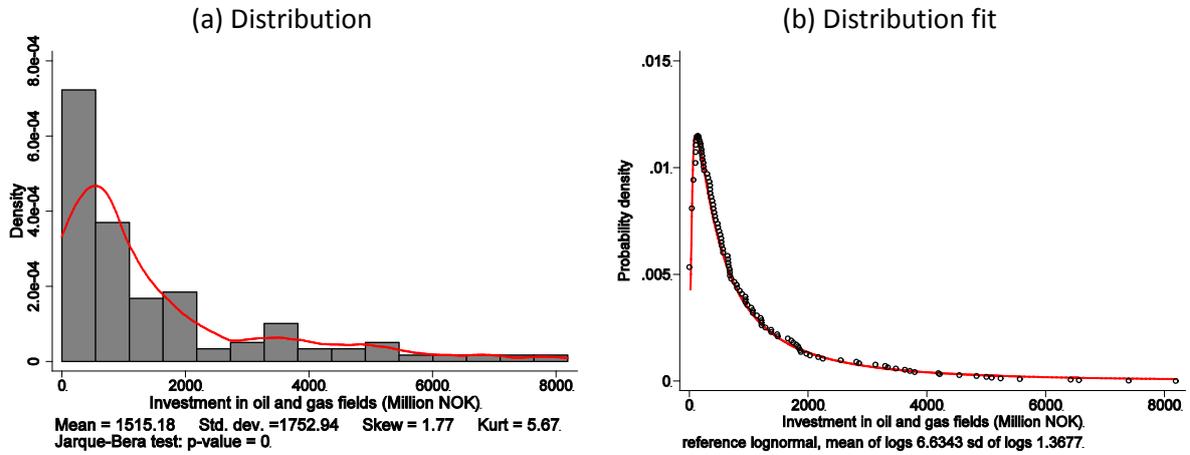
3.0 Statistical characteristics of investments

While the aggregate investment is the variable of primary concern, it is paramount to acknowledge that oil and gas investment on the NCS is derived from fields operated by oil companies forming license groups. Although the properties of a company’s investment behaviour could differ from the aggregate, it might be insightful to address the statistical properties of investment disaggregated down to fields.

Figure 4 (a) shows the distribution of the total investment across 109 oil and gas fields on the NCS between 1971 and 2015. As observed, the kernel density plot of the distribution appears to resemble a lognormal distribution. Figure 4 (b) compares the density-quantile function (the density function of investment computed through the quantile function) with the theoretical lognormal distribution with parameters derived from the sample. Based on visual inspection, the fit appears to be quite adequate. An implication of these observations is that we can expect the data to exhibit similarities to the Pareto principle, i.e. the 80/20 rule. Conferring with the data, 20 percent of the oil and gas fields accounts for 60.58 percent of the aggregate investment on the NCS. Consequently, not all fields are equally important to the aggregate investment. If a qualitative approach were to be undertaken in the pursuit of forecasting aggregate investment it might be advisable to not spend a uniform amount of effort on analysing the different fields.

Figure 4: **Total investment in oil and gas fields**

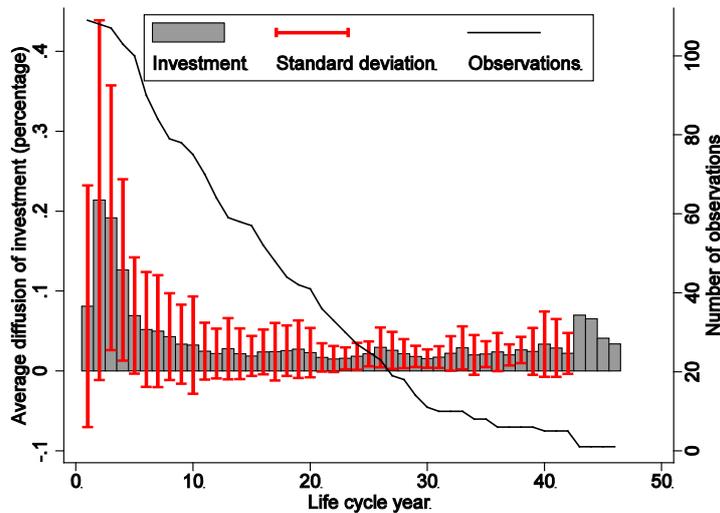
The distribution of the total inflation-adjusted investment for 109 oil and gas fields aggregate over time (1971-2015) on the NCS. The distribution is approximated through a histogram and an Epanechnikov kernel density plot.



Common knowledge would suggest that the dispersion of investments throughout the life cycle of the oil and gas field does not follow a uniform distribution. Rather, it seems more plausible to expect investment to be predominantly clustered around the early stages of the investment period. Investment behaviour can be observed in Figure 5. Here each field's yearly investment is relative to the fields total observed lifetime investment. Conforming to prior beliefs, the normalized investment diffusion appears to build up quickly before geometrically declining to a comparably lower and more stable level with the occasional spike. This characteristic of a field's investment dispersion might be pivotal for forecasting investment disaggregated down to fields, but not necessarily when considering the aggregate investment on the NCS.

Figure 5: **Investment in fields across life time**

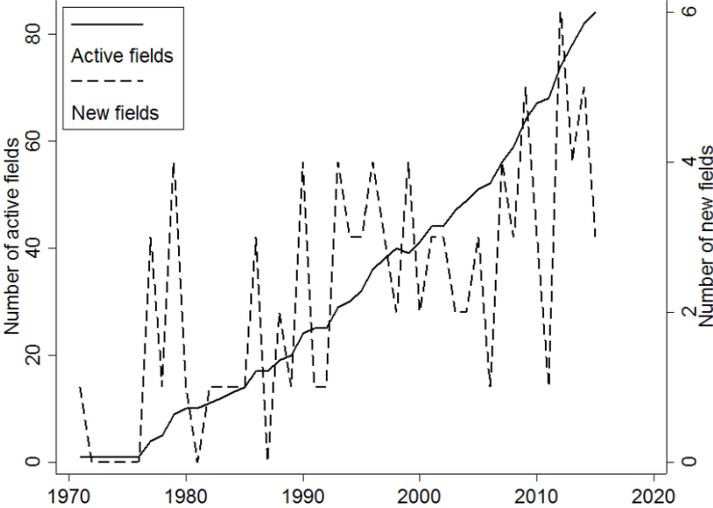
Field investment across the life cycle, normalized by the fields total investment size throughout the whole cycle.



If the projects are of similar size and if the initiation of new projects are uniformly distributed throughout the sample period, then the initial spike in the life cycle will be averaged out. However, these assumptions might not hold in this dataset. First, as revealed in Figure 4 (a), the first assumption appears to be violated as a relatively small subsample of the fields contributes a disproportionate amount of the total investment. Second, it is possible that the oil and gas companies' investment decisions are correlated given that their judgment could be affected by the business cycle. For instance, if the decision to either initiate or postpone a project is significantly driven by the oil price, then periods of low prices will cause an investment drought and conversely an investment boom when prices are high. Under this mechanism, the initial spike in the investment dispersion will amplify the effect on the aggregate investment from the postulated cascading behaviour of the oil and gas companies. As shown in Figure 6, the number of new fields on the NCS follows a fluctuating pattern, which does not appear to be uniform.

Figure 6: Oil and gas fields

The number of active fields on the NCS and the number of yearly additions of new fields.



4.0 Empirical analysis of key variables

Table A1 and A2 in appendix A present results from the unit root tests for the dataset both under panel data structure and under cross-sectional aggregation. As revealed by the Fisher panel data unit root test, the null hypothesis of non-stationarity is predominantly rejected by both the augmented Dickey-Fuller (ADF) and Phillips-Perron specification. However, when the data is subjected to cross-sectional aggregation, changing the data structure from panel to time series, the null is chiefly not rejected. Given this paper's primary interest in the aggregate investment, two approaches are immediately apparent – either apply first differences (if the data is integrated of order one) to make aggregate investment stationary or employ an error correction model (ECM) if investment is cointegrated with our explanatory variables (See table 4 for a list of evaluated predictors).

An augmented Dickey-Fuller and Phillips-Perron test reveal that aggregate investment, crude oil price and the USD/NOK exchange rate are integrated of the first order, while realized volatility is stationary in levels. As such, the realized volatility is excluded from further cointegration testing. The Johansen's cointegration test indicates that total investment and the oil price exhibit a long-term relation as the yielded trace statistics of 19.30 exceed the critical 5 percent value of 15.41 under the null hypothesis of zero cointegrated relations. The number of lags is specified by utilizing the Hanna-Quinn information criterion. However, with a yielded test statistics of 14.17 it appears that there is no long-term relation between aggregate investment and the USD/NOK exchange rate. Thus, it is possible to form an ECM with the aggregate investment and crude oil price, but neither volatility nor the exchange rate can be included. Consequently, we opt for applying log-difference to the aggregate investment, effectively transforming the variable into the growth of aggregate investment. Differencing aggregate investment changes the econometric model from addressing a long-term relationship to a short-term². However, this is inconsequential if only the one-step ahead forecast is of interest.

It is likely that a vast number of variables have the potential to forecast future growth in aggregate oil and gas investment on the NCS. Based on theory and insight gained from related studies of investment activity on the NCS, quoted in Section 1, we consider three potential predictors in addition to past investment growth. See Table 2 for a full list of all predictors and their respective definitions.

² Instead of forecasting investment (Inv_t) at time t , the continuously compounded return ($\ln(Inv_t) - \ln Inv_{t-1}$) between period $t - 1$ and t is forecasted. These are related in the following manner:

$$Inv_t = Inv_{t-1} * e^{\ln\left(\frac{Inv_t}{Inv_{t-1}}\right)} = Inv_{t-1} * \frac{Inv_t}{Inv_{t-1}}$$

Table 2: **List of investment growth predictors**

This table shows the definition of the variable to be forecasted and the list of evaluated predictors.

Variable	Description
Investment growth	Logarithmic change in aggregate oil and gas investment on the NCS with an annual frequency, $\log(investment_t) - \log(investment_{t-1})$.
Change in oil price	Logarithmic change in Brent crude oil price with an annual frequency, $\log(p_t) - \log(p_{t-1})$.
Volatility	Volatility of the crude oil price approximated through the realized volatility proxy, see Equation 5
Change in exchange rate	Logarithmic change in USD/NOK exchange rate with an annual frequency, $\log(FEX_t) - \log(FEX_{t-1})$.

Growth in aggregate investment at time t is the variable to be predicted, but given the considerable lead time between project execution and production start (see Figure 11), it is reasonable to use information about investment growth available in previous years as a predictor. Once a project has come to the construction phase it is seldom stopped, and since fabrication and installation takes several years, lagged variables of the investment should be included. Under the assumption of economic rationality, investment is fundamentally driven by expectations for future profitability. Given the validity of this claim, anything that exhibits the power to affect the profitability ranking of the companies' portfolio of investment opportunities is a potential predictor. On the income side, the oil price is perhaps the single most important driver of profitability. As such, the crude oil price growth is a good potential predictor of future aggregate investment growth.

Investments are measured in NOK, but material and equipment from the international market is a necessity. As such, the exchange rates matter for the monetary amount of NOK invested. Everything else being equal, an increase in the USD/NOK exchange rate will decrease the purchasing power of Norway and subsequently the amount invested will increase, all other things being equal.

Volatility is an important driver of investment. First, in terms of real options, the value of waiting increases as the volatility increases and the market become more erratic (Pindyck, 1994). Second, given the capital intensive nature of oil and gas investments, it is quite plausible that increased volatility makes decision-makers more hesitant to commit³. Consequently, prima facia, we expect volatility to exhibit a negative influence over the aggregate investments. Volatility is latent which implies that it is necessary to estimate it ex post. Volatility has a very precise definition in the literature (Andersen et al., 2006, 780): "[...] in financial economics, volatility is often defined as the (instantaneous) standard deviation (or "sigma") of the random Wiener-driven component in a continuous-time diffusion model." In practice, there is a wide variety of approaches to estimate the latent ex post volatility. In this paper, we choose to utilize the realized volatility proxy. Following the notation of Andersen and Bollerslev (1998) and Hansen and Lunde (2001), we define realized

³ Relevant issue here is the comparison objectives related to dividend policy and credit rating. (Osmundsen et al, 2005)

volatility in the following manner. First, let t exhibit the desired frequency of the time series (yearly in this case) and m be the number of intra-frequency observations. Then, the continuously compounded returns on the oil price is given as follows:

$$r_{(m),t+j/m} = \log(p_{t+j/m}) - \log(p_{t+(j-1)/m}) \quad \forall j = \{1, \dots, m\} \quad (1)$$

With the growth or returns in the oil price realized volatility is defined as:

$$\begin{aligned} \sigma_t^2 &\equiv \text{var}(r_t | \mathcal{F}_{t-1}) \\ &= E \left(\sum_{j=1}^m r_{(m),t+j/m} - E(r_{(m),t+j/m} | \mathcal{F}_{t-1}) \right)^2 \\ &= \sum_{j=1}^m \text{var}(r_{(m),t+j/m} | \mathcal{F}_{t-1}) + \sum_{i \neq j} \text{cov}(r_{(m),t+i/m}, r_{(m),t+j/m} | \mathcal{F}_{t-1}) \end{aligned} \quad (2)$$

Under the assumption of no autocorrelation in the continuously compounded intra-frequency return on the crude oil price, the identity of the volatility becomes:

$$\sigma_t^2 \equiv \text{var}(r_t | \mathcal{F}_{t-1}) = \sum_{j=1}^m \text{var}(r_{(m),t+j/m} | \mathcal{F}_{t-1}) \quad (3)$$

Further, given a sufficiently rapid measurement of the intra-frequency return we can further assume that the expected return becomes minuscule and subsequently neglectible.

$$E(r_{(m),t+j/m}^2 | \mathcal{F}_{t-1}) \approx \text{var}(r_{(m),t+j/m} | \mathcal{F}_{t-1}) \quad (4)$$

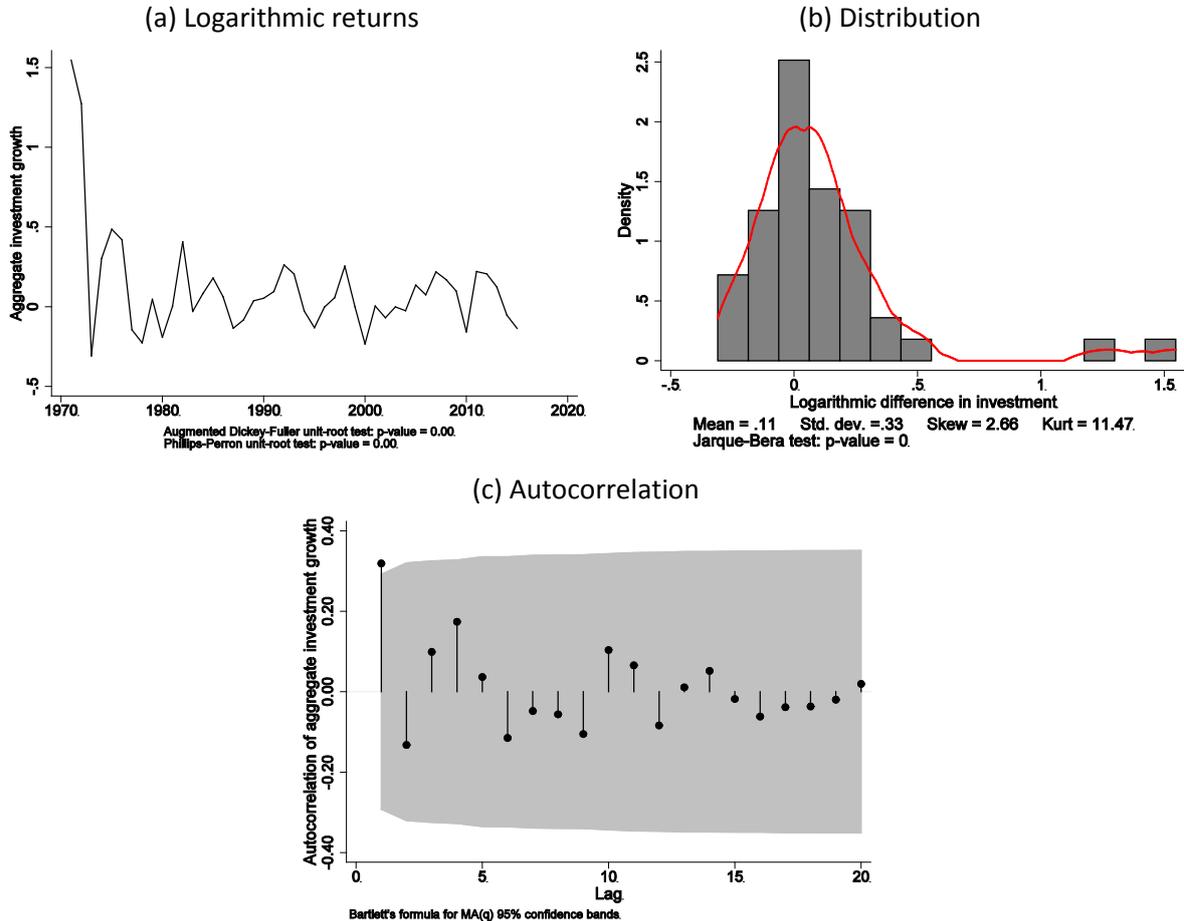
Finally, the realized volatility becomes the squared sum of the continuously compounded intra-frequency.

$$\hat{\sigma}_{(m),t}^2 \equiv \sum_{j=1}^m r_{(m),t+j/m}^2 \quad (5)$$

With the theoretical motivation and technical definitions of the predictors in mind, we further investigate the statistical properties of the proposed predictors of growth in aggregate oil and gas investment on the NCS.

Figure 7: **Growth in aggregate investment**

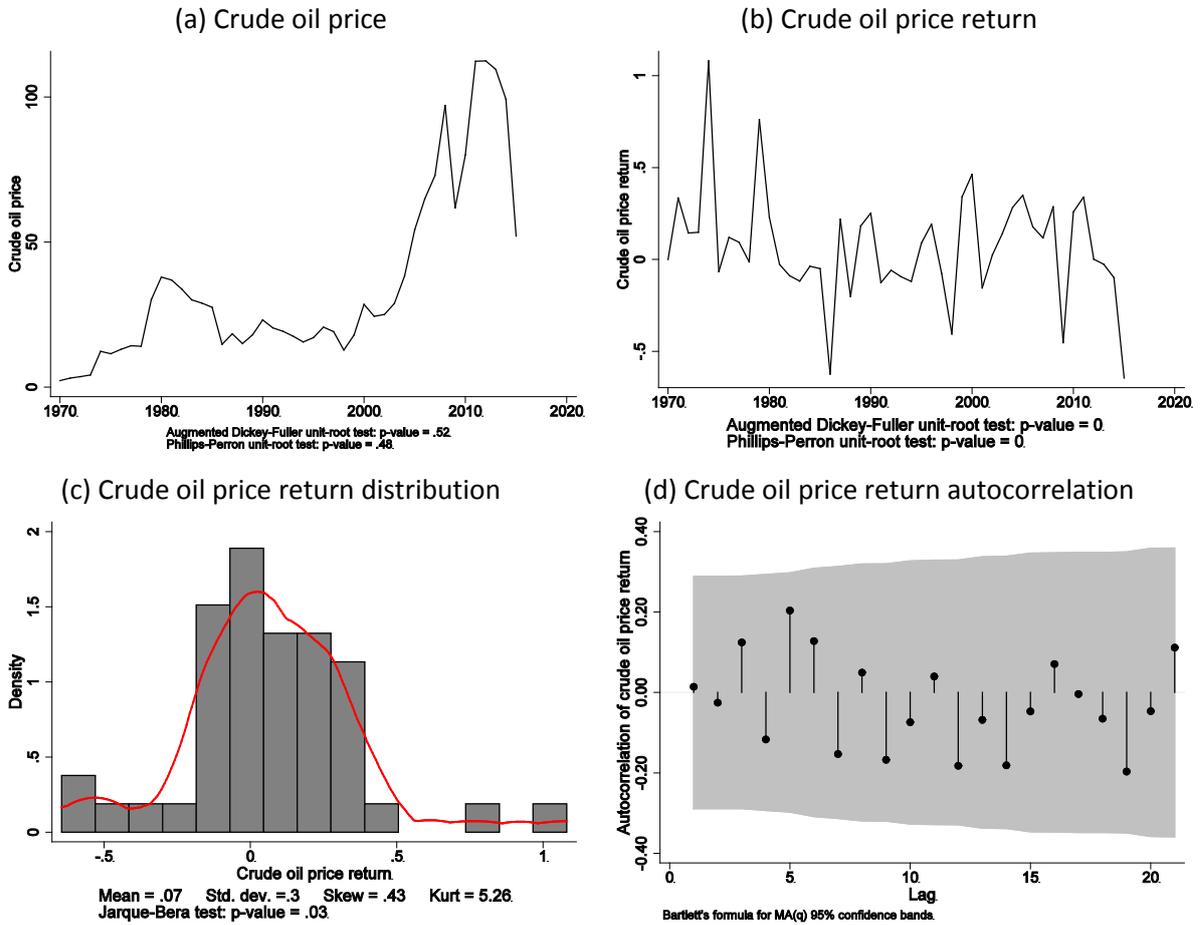
The growth in aggregate oil and gas investment on the Norwegian Continental Shelf between 1970 and 2015. Growth is here defined as the logarithmic change in investment. Subfigure (a) displays the development over time in investment growth, (b) the distribution approximated by a histogram and Epanechnikov kernel density plot, and (c) the autocorrelation function. All data was provided by the Norwegian Petroleum Directorate.



First, Figure 7 shows the statistical properties of the growth in aggregate investment. While aggregate investment is not stationary in levels, both the Augmented Dickey-Fuller and the Phillips-Perron test confirm that the logarithmic change, i.e. the growth, is stationary. By inspecting the distribution of the growth through a histogram and an Epanechnikov kernel density plot, we can see that it is reasonably well behaved. Based on the plot and summary statistics, there appears to be some outliers causing the distribution to exhibit positive skewness. A more formal Jarque-Bera test confirms that the distribution of the investment growth is indeed not normally distributed. The autocorrelation plot shows that the autocorrelation of the investment resembles a damped sinusoid, that is, it declines geometrically with the lags alternating between being positive and negative. However, only the first lag is significant on a five percent level. Consequently, it seems that little value can be gained from applying an autoregressive model on the investment growth with a too large lag structure.

Figure 8: **Crude oil price**

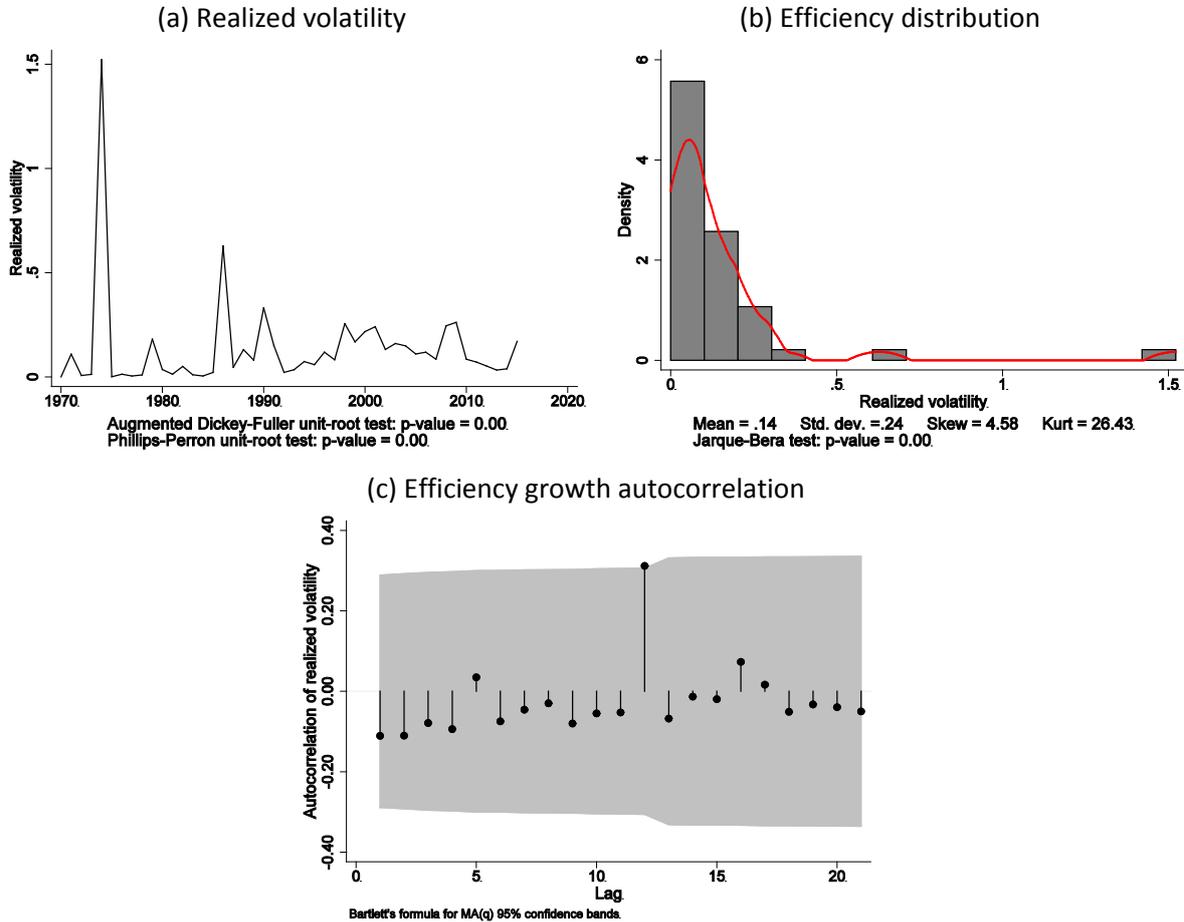
Descriptive statistics regarding the crude oil price and return. Subfigure (a) displays the price, (b) the return, (c) the distribution of the return approximated through a histogram and Epanechnikov kernel density plot, and (d) the autocorrelation function of the return.



Second, the characteristic of the oil price is well known in the literature. As observed in Figure 8, the variable is not stationary in level but stationary after applying the first difference. The distribution of the oil price growth exhibits considerable amount of leptokurtosis and consequently is not normally distributed. An inspection of the autocorrelation of the oil price growth strongly indicates there is little forecastability in the statistical sense.

Figure 9: **Realized volatility**

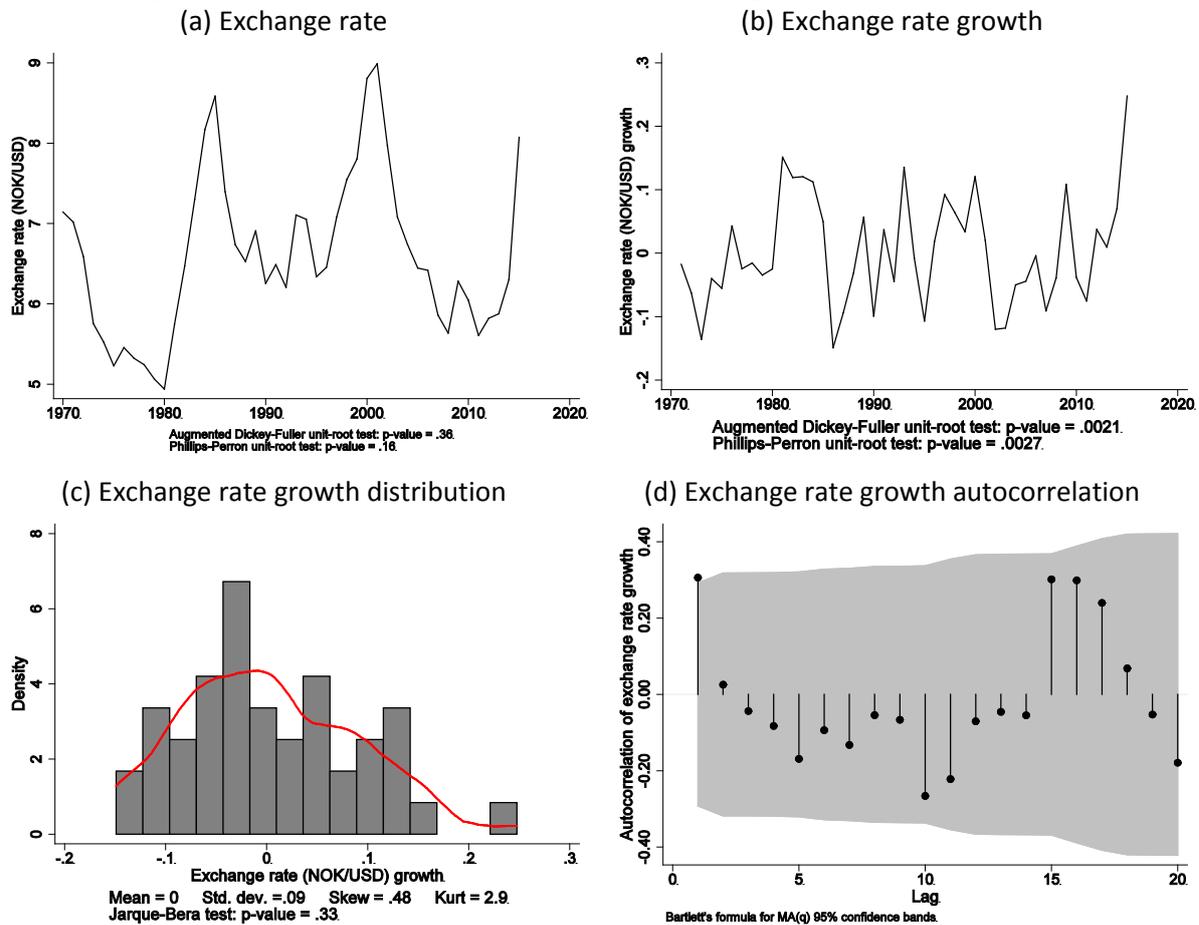
Descriptive statistics regarding the realized volatility of the crude oil price. Subfigure (a) realized volatility, (b) the distribution of the volatility approximated through a histogram and Epanechnikov kernel density plot, and (d) the autocorrelation function of the volatility.



Third, Figure 9 shows the characteristics of the realized volatility of the crude oil price. Confirming to stylized facts, the volatility is stationary and possesses a mean reverting behaviour, thus implying that shocks eventually die out. Thus, a negative investment impulse from increased oil price volatility, which we now observe, will be temporary. Being naturally bound in the interval of zero to infinity, the distribution of the realized volatility is obviously not normally distributed. Interestingly, there appears to be a miniscule amount of autocorrelation in the volatility.

Figure 10: USD/NOK exchange rate

Descriptive statistics regarding the USD/NOK exchange rate. Subfigure (a) displays the level, (b) the growth, (c) the distribution of the growth approximated through a histogram and Epanechnikov kernel density plot, and (d) the autocorrelation function of the growth.



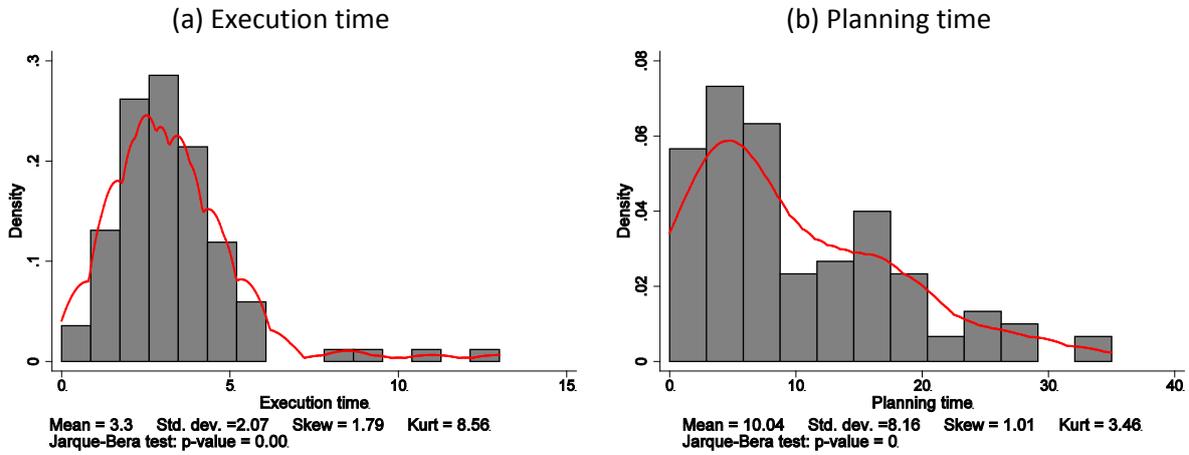
Finally, Figure 10 shows the statistical characteristics of the USD/NOK exchange rate. Confirming to prior beliefs, the exchange rate appears to contain a unit root in levels but is stationary in the first-difference. The exchange rate has changed extensively throughout the sample period, but cannot grow ad infinitum. The distribution of the exchange rate growth deviates from the normal by being characterized by a slight positive skewness and a considerable kurtosis. Besides a significant first lag, there appears to be a miniscule amount of autocorrelation. A fact that may weaken the exchange rate as a predictor, beyond the oil prices, is the high correlation with the oil prices.

Statistical forecasting in this context is utterly deterministic, which is to say that the forecast for the future is derived only from historical observations of the predictors. A prevalent point of contention in that regard is how much of the past is relevant. Admittedly, this is challenging to answer with any certainty. However, it is generally accepted that the distant past matters less than the recent. For the oil industry, however, a long time series is essential. One point of departure is to consider the timing aspects of investments.

Figure 11 (a) shows the distribution for the investment execution time, defined as the number of years between the acceptance of the Plan for Development and Operations (PDO) and production start. In other words, how long time it takes to develop a field.

Figure 11: **Project execution time**

The distribution of the length of project execution approximated through histogram and Epanechnikov kernel density plot. Execution time is defined here as the number of years in difference between the date for PDO approval and production start.



Given the irreversible nature of these types of investment, it is difficult to cancel a project once initiated. As discerned from the distribution of the execution time, it takes an average of 3.3 years from the time a project is initiated to its completion. This implies that even if a dramatic fall in oil prices were to effectively stop most new investments, it will still take years to see the full effect due to the long lead time in project execution. Analogously, Figure 11 (b) shows the distribution of planning time, which is here approximated as the number of years between the discovery of an oil and gas field and project sanctioning (date of PDO). Although this proxy is imperfect, it is evident that it can take a considerable amount of time before a project is initiated. Thus an oil price spike, even if it were perceived to be permanent, will have no immediate effect on investments. This seems to be indicating that predictors can have a long lasting effect.

The simplest statistical approach to gain insight regarding how far the temporal reach of the predictors are, is to consider the correlation between investment growth at time t and the predictors at time $t - i$ for i greater or equal to one. Table 3 displays the results. As it appears, the oil price growth has a noteworthy relation to the aggregate investment up until the fourth lag before becoming minuscule – a trait that appears to be shared by the volatility and exchange rate growth. The oil price growth has a positive correlation for the first to third lag and is negative at the fourth lag. The volatility has a positive correlation for the first two lags before becoming negative. Finally, the exchange rate growth begins with a correlation of zero before becoming negative and noteworthy. Besides being different from zero, the causal interpretation of the correlations is not of interest. It is possible that the predictors still possess an influence over the aggregate investment beyond the fifth lag, but due to practical considerations it is not feasible to pursue a longer lag length. First, as the lag length is increased, the sample period is decreased and, subsequently, statistical testing becomes more challenging. Second, incrementally increasing the number of lags yields an almost exponential increase in the computation time when evaluating all possible combinations of lags.

Table 3: **Correlation**

This table shows the correlation between the aggregate investment return for oil and gas projects at the Norwegian Continental Shelf and various lags of a set of proposed predictors. The included predictors are the growth in crude oil prices, the realized volatility of the crude oil price and the USD/NOK exchange rate growth.

Variable	Lag(1)	Lag(2)	Lag(3)	Lag(4)	Lag(5)
Oil price growth	0.21	0.37	0.14	-0.24	0.04
Realized volatility	0.23	0.26	-0.10	-0.25	0.05
Exchange rate growth	0.00	-0.26	-0.24	-0.12	-0.02

5.0 Methodology

The purpose of this research is to identify the model with highest forecast accuracy for oil and gas investment on the Norwegian Continental Shelf. We evaluate several models differing in specification based on their out-of-sample performance. The models' forecast accuracy are subsequently evaluated by a set of different statistical loss functions. The validity of the ranking of the models forecast accuracy is tested both with the Diebold and Mariano (1995) predictive accuracy test and with the Hansen and Lunde (2011) model confidence set procedure.

5.1 Model specification

The main model of interest in this research is the Autoregressive Distributed Lag model (ADL). The explanatory variables in the model are selected based on theoretical considerations. Several specifications of the ADL are considered, both in terms of number of lags and which regressors to include. Specifically, with all possible combinations of lags subjected to a chosen max lag⁴ and all possible inclusion schemes for the regressors, $m + \sum_{i=1}^q m^{i+1} \binom{q}{i}$ models are evaluated, where m is the maximum number of lags allowed and q is the number of regressors. The maxlag is determined both such that the sample contains sufficiently many observations for evaluation and such that it is computationally feasible to consider all specifications. The ADL model is given by:

$$y_t = \alpha + \sum_{i=1}^p \beta_i y_{t-i} + \sum_{j=0}^q \left(\sum_{k=1}^r \gamma_{jk} x_{jt-k} \right) + u_t \quad (6)$$

Where y is the dependent variable and x is the independent variable. Additionally to the ADL model, we include a driftless random walk as a more parsimonious and atheoretical benchmark. If the ADL model with its complexity is not able to outperform this benchmark, it is of limited value.

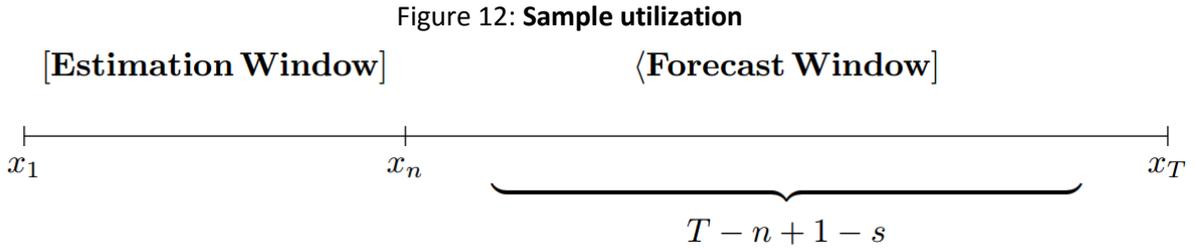
Random Walk:
$$y_t = y_{t-1} + u_t \quad (7)$$

5.2 Forecast scheme and evaluation

With the given set of models, the forecast is made both in-sample and ex post (pseudo) out-of-sample. While the in-sample procedure is fast, it is prone to data mining and overfitting. That is, random chance and capturing the idiosyncratic behaviour of the sample might lead to seemingly good model but with poor external validity. To address this problem, we apply out-of-sample forecasting procedures. However, it comes at the cost of being computationally more demanding. The out-of-sample

⁴ In general, the max lag is a trade-off between the models explanatory power and keeping it parsimonious. Additionally, as the max lag increases, the sample size decreases.

procedure is implemented as follows. With T observations ordered as $\{x_1, \dots, x_T\}$, we divide the data into two subperiods: $\{x_1, \dots, x_n\}$ and $\{x_{n+1}, \dots, x_T\}$. The former is utilized for estimating the model and the latter for evaluating the forecast made, see Figure 12. A rolling window scheme is utilized such that the forecast subset is incrementally decreased while the estimation subset remains of the same size. Thus, a total number of $T - n + 1 - s$ pseudo out-of-sample forecasts are made. Here s denotes how many periods into the future the forecast is made for. The split point n is of importance for the out-of-sample procedure. The demarcation between the estimation and forecast window must be specified such that both are sufficiently long to make valid inference.



The forecast accuracy of a particular model is evaluated by comparing the actual and the predicted investment with a weighing-scheme dictated by a loss function. Loss functions are typically either economic or statistical. To preserve the general applicability of this research only the latter is considered. While the statistical loss functions are arbitrary, they attempt to emulate the users' utility function. In practice, there is a wide variety of available loss functions. The main difference relates to (1) whether the loss function employs a symmetric or asymmetric penalty to over- and under-predictions, and (2) whether an increase in the absolute error is penalized linearly or nonlinearly. In this paper, we choose to utilize the root mean square error (RMSE) loss function, see Equation 8. First, RMSE is widely used in the forecasting literature. By using a frequently applied loss function, it becomes easier to compare the obtained results with related literature. Second, it can be argued that RMSE fits the intended usage as it has a symmetric and nonlinear penalty. That is, under and over predictions of equal size are regarded as equally undesirable, and large forecast errors are penalized over proportionally stricter compared to minor errors.

RMSE:
$$L(y, \hat{y}) = \sqrt{(\hat{y} - y)^2} \quad (8)$$

5.3 Model confidence set procedure

Based on each loss function we can order the models from highest to lowest forecast accuracy. To assure the validity of the forecast accuracy ranking, we perform a formal hypothesis test, specifically, the model confidence set procedure, proposed by Hansen et al (2011). Let y_t be the observed investment on the NCS at time t and $\hat{y}_{i,t}$ be the predicted value also at time t generated by a given model i . Given the observed and predicted values of investment, we can evaluate the forecast accuracy with any given loss function such that we let $L_{i,t}$ denote the loss at time t for model i .

$$L_{i,t} = L(y_t, \hat{y}_{i,t}) \quad (9)$$

With a vector of loss values for all M considered models, we want to define the superior set of models (SSM). That is, we attempt to find a subset of these models, $\widehat{M}_{1-\alpha}^* \subseteq M$, such that the subset contains models that possess equal accuracy, but are superior to the remaining set of models, given a confidence interval of $1 - \alpha$. To accomplish this, two concepts are of importance, the equal predictive ability (EPA) hypothesis and the elimination rule.

To specify the EPA hypothesis, both the loss differential and the relative loss differential must be defined. Let $d_{ij,t}$ denote the difference in forecast accuracy between model i and j at time t , i.e. the loss differential is given by:

$$d_{ij,t} = L_{i,t} - L_{j,t}, \quad i, j = 1, \dots, m, \quad t = 1, \dots, T \quad (10)$$

Further, let $d_{i,\cdot}$ be the accuracy of model i compared with the remaining set of models, i.e., the relative loss differential:

$$d_{i,\cdot} = (m - 1)^{-1} \sum_{j \in M} d_{ij,t} \quad i = 1, \dots, m \quad (11)$$

With $d_{ij,t}$ and $d_{i,\cdot}$, the EPA hypothesis is expressed as:

$$\begin{aligned} H_{0,M}: \quad \mathbb{E}(d_{ij}) &= 0 \quad \forall i, j = 1, \dots, m \\ H_{1,M}: \quad \mathbb{E}(d_{ij}) &\neq 0 \quad \text{for some } i, j = 1, \dots, m \end{aligned} \quad (12)$$

and

$$\begin{aligned} H_{0,M}: \quad \mathbb{E}(d_{i,\cdot}) &= 0 \quad \forall i, j = 1, \dots, m \\ H_{1,M}: \quad \mathbb{E}(d_{i,\cdot}) &\neq 0 \quad \text{for some } i, j = 1, \dots, m, \end{aligned} \quad (13)$$

where the former is valid for loss differential and the latter for the relative. Under the null hypothesis, all models possess equal predictive ability such that no model is superior to any other considered model. To implement these hypothesis tests, the statistics t_{ij} and $t_{i,\cdot}$ are required as an intermediate step to obtain the test statistics $T_{R,M}$ and $T_{max,M}$.

$$t_{ij} = \frac{\bar{d}_{ij}}{\sqrt{\widehat{\text{var}}(\bar{d}_{ij})}} \quad \text{for } i, j \in M \quad (14)$$

$$t_{i,\cdot} = \frac{\bar{d}_{i,\cdot}}{\sqrt{\widehat{\text{var}}(\bar{d}_{i,\cdot})}} \quad \text{for } i \in M \quad (15)$$

The numerators in the test statistics are given respectively by $\bar{d}_{ij} = m^{-1} \sum_{t=1}^m d_{ij,t}$ and $\bar{d}_{i,\cdot} = (m - 1)^{-1} \sum_{j \in M} \bar{d}_{ij}$ while the denominators are simply the bootstrapped estimate of the variance of \bar{d}_{ij} and $\bar{d}_{i,\cdot}$. Finally, t_{ij} and $t_{i,\cdot}$ are used to define $T_{R,M}$ and $T_{max,M}$.

$$T_{R,M} = \max_{i,j \in M} |t_{ij}| \quad \text{and} \quad T_{max,M} = \max_{i \in M} t_{i,\cdot} \quad (16)$$

Given the test statistics, if the null hypothesis is rejected such that there exists at least one inferior model, an elimination rule is applied to remove models with lower performance. This process is repeated until the null hypothesis can no longer be rejected. In the best case scenario, the set of superior models will contain only one model. The elimination rule is specified as follows:

$$e_{max,M} = \arg \max_{i \in M} \frac{\bar{d}_{i.}}{\sqrt{\widehat{var}(\bar{d}_{i.})}} \quad \text{and} \quad e_{R,M} = \arg \max_i \left\{ \sup_{j \in M} \frac{\bar{d}_{ij}}{\sqrt{\widehat{var}(\bar{d}_{ij})}} \right\} \quad (17)$$

6.0 Empirical results

Based on the ex post out-of-sample forecast evaluation of the 1080 ADL model specifications, there are clear findings indicating that it is possible to outperform the driftless random walk benchmark in forecasting future aggregate investment on the NCS. By utilizing the RMSE loss function for measuring the forecast errors, we find that 61 different specifications of the model are capable of producing forecasts with a greater accuracy than the benchmark in this particular sample. However, in line with fundamental statistical theory, any set of data is regarded as a random realization of an underlying and unobservable data generating process. Consequently, it is possible that the model forecast accuracy ranking obtained by comparing the models' associated loss function values is distorted by noise. Different samples are not guaranteed to produce the same ranking. To address this sample uncertainty, we employ the Diebold and Mariano test of equal predictive accuracy by comparing each subsequent model against the benchmark. As revealed by Table B1 in Appendix B, the null hypothesis of equal predictive accuracy is rejected on a ten per cent level for 30 different specifications of the ADL model. For the 31 models that are seemingly superior to random walk - but not able to reject the null hypothesis - it would seem that the variability within the models' forecast errors are considerable. When the variability is extensive it is challenging to determine whether the model is genuinely accurate or just happens to make an accurate prediction in this particular sample of data due to luck. Thus, the conservative course of action is to dismiss such models.

Comparing the models' forecast accuracy against random walk with the Diebold and Mariano test is useful for eliminating the inferior models, but it does not determine what the overall best model is. To answer this question, we utilize the Hansen and Lunde model confidence set procedure to simultaneously compare several models against each other. The procedure works by using an iterative procedure where the weakest model, in terms of forecast accuracy, is eliminated until it is no longer possible to differentiate between the models. To address the sample uncertainty, the procedure estimates the variability in the models' forecast errors by bootstrapping. As revealed by Table 6, the set of superior models appears to depend on whether the $T_{R,M}$ or $T_{max,M}$ test statistic is applied. While both approaches are conceptually similar, the underlying mathematical formulation differs. Hence, there is no obvious argument for favouring one over the other. Following the $T_{R,M}$ test results, it would seem that the set of superior models consists of ten different specifications of the ADL model. Among these superior models, the $ADL(3,1,2,0)$ model has the lowest RMSE realization. On the contrary, when applying the $T_{max,M}$ statistic, none of the 30 models superior to random walk can be differentiated based on forecast accuracy. Thus, in this set of superior models, the $ADL(1,1,1,1)$ possesses the lowest RMSE, so in summary, we have demonstrated that it is indeed possible to outperform the parsimonious random walk forecast on a statistically significant level by using a simple ADL approach in an out-of-sample evaluation.

Whether to base the model selection the $T_{R,M}$ or $T_{max,M}$ test statistic is dependent on subjective preferences. As revealed in Table B1, $T_{R,M}$ eliminates more models compared to $T_{max,M}$. Thus, it would

seem that the latter is more conservative in the sense that it requires more evidence to make an elimination. On the other hand, being overly conservative potentially means that models that demonstrate a high accuracy in this sample might just be the result of luck. In the pursuit of a good model it can be argued that it is preferable to eliminate too many rather than too few as it reduces the probability of selecting a model that could potentially make considerable forecast errors. Consequently, we opt to base the model selection on the $T_{R,M}$ statistic. Under this scenario, the selected model becomes $ADL(3,1,2,0)$, which implies that three lags are included for aggregate investment growth $\Delta \ln(Inv)$, one lag for crude oil price growth $\Delta \ln(OilPrice)$ and two lags for realized volatility of the crude oil price $\Delta \ln(Vol)$. The USD/NOK exchange rate growth $\Delta \ln(FEX)$ is not included. The equation for the proposed regression model is given in Equation 18.

$$\Delta \ln(Inv_t) = \alpha + \sum_{i=1}^3 \beta_i \Delta \ln(Inv_{t-i}) + \gamma \Delta \ln(OilPrice_{t-1}) + \sum_{j=1}^2 \beta_j Vol_{t-j} + u_t \quad (18)$$

Given the stated objective of forecasting future movements in aggregate oil and gas investments on the NCS and the atheoretical nature of time series models in general, a causal interpretation of the obtained coefficient estimates and their respective significance is not of primary interest or even advisable. The model is designed for investment forecasting. Nevertheless, Figure 13 shows the development of the obtained coefficient estimates and associated p-values from the regression model presented in Equation (21). As observed in subfigure (a), the coefficient for the first lag of investment growth is positive, thus indicating that aggregate investments possess momentum. However, Subfigure (b) and (c) show that the second and third lags are estimated to be negative. Thus, it appears that investment goes through a cycle of momentum followed by a correction. Subfigure (d) shows the coefficient for the crude oil price growth. The coefficient appears to start out negative, but quickly becomes positive. Thus, later in the sample, investment tends to increase when the oil price increases. Subfigure (e) shows the coefficient for the first lag of the realized volatility of the crude oil price. The coefficient is negative throughout all subsamples. This implies that investment tends to decrease when volatility increases. Similarly, Subfigure (f) shows the coefficient of the second lag of the volatility. In this case, the coefficient lies close to zero, thus alternating between being positive or negative.

Figure 13: **Coefficients**

The development of the coefficients of the ADL(3,1,2,0) with the corresponding 95 % confidence interval and p-values, as the estimation rolls over the sample period. The initial estimation window covers the period of 1975 to 1995. The regression equation is given as follows:

$$\Delta \ln(Inv_t) = \alpha + \sum_{i=1}^3 \beta_i \Delta \ln(Inv_{t-i}) + \gamma \Delta \ln(OilPrice_{t-1}) + \sum_{j=1}^2 \beta_j Vol_{t-j} + u_t$$

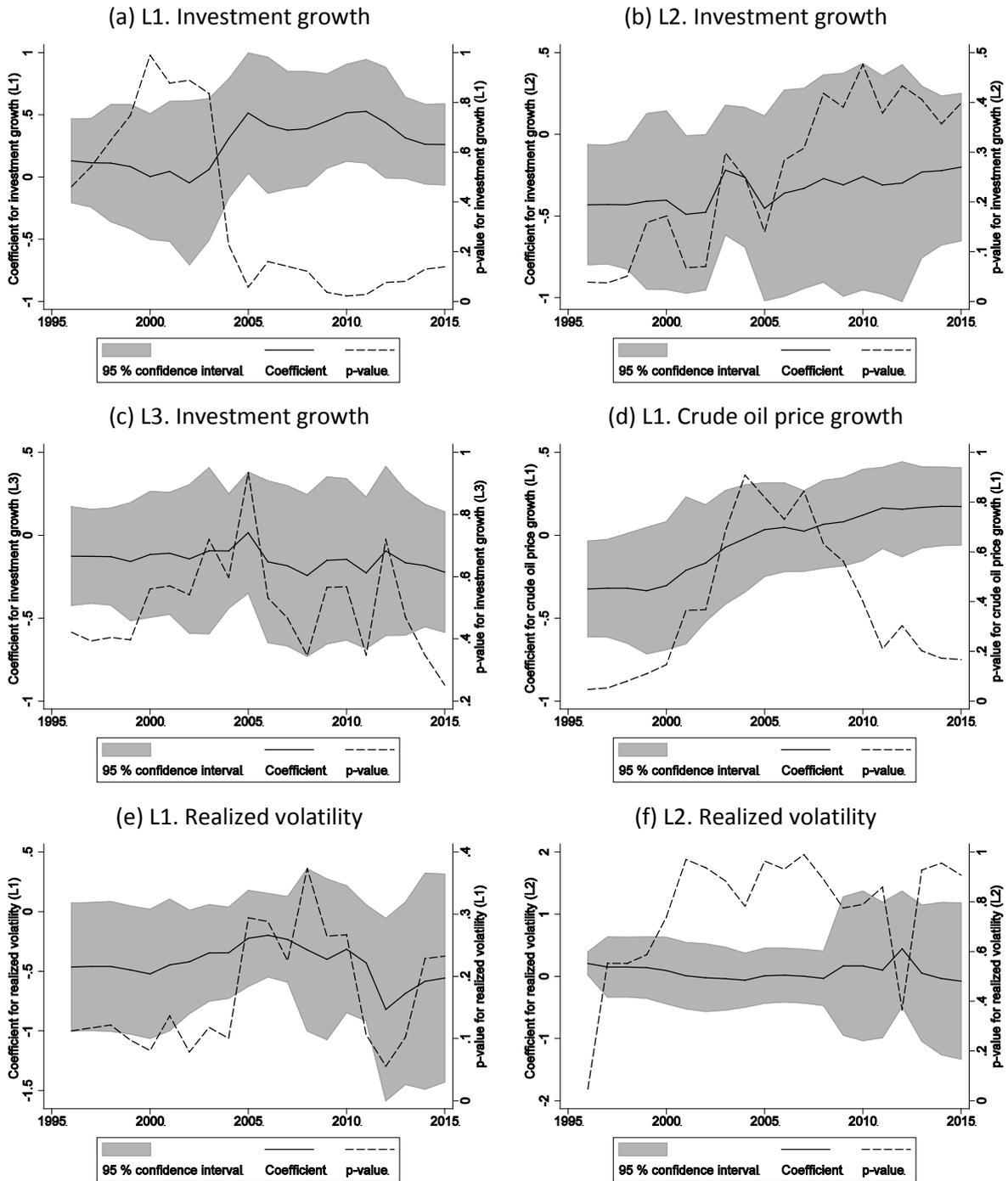


Figure 14: **Model forecast accuracy**

Illustration of the forecast accuracy of the $ADL(3,1,2,0)$ model by comparing the predicted and realized investment growth. Subfigure (a) shows the investment growth during the full sample period, the predicted values during in forecast window and the forecast accuracy represented as the area between these time series. Subfigure (b) shows the forecast accuracy as the area between the predicted and realized investment growth. Subfigure (c) compares the predicted and realized investment growth during the forecast window. Subfigure (d) compares the prediction made from the model both with and without updating the coefficient estimates with the realized investment growth.

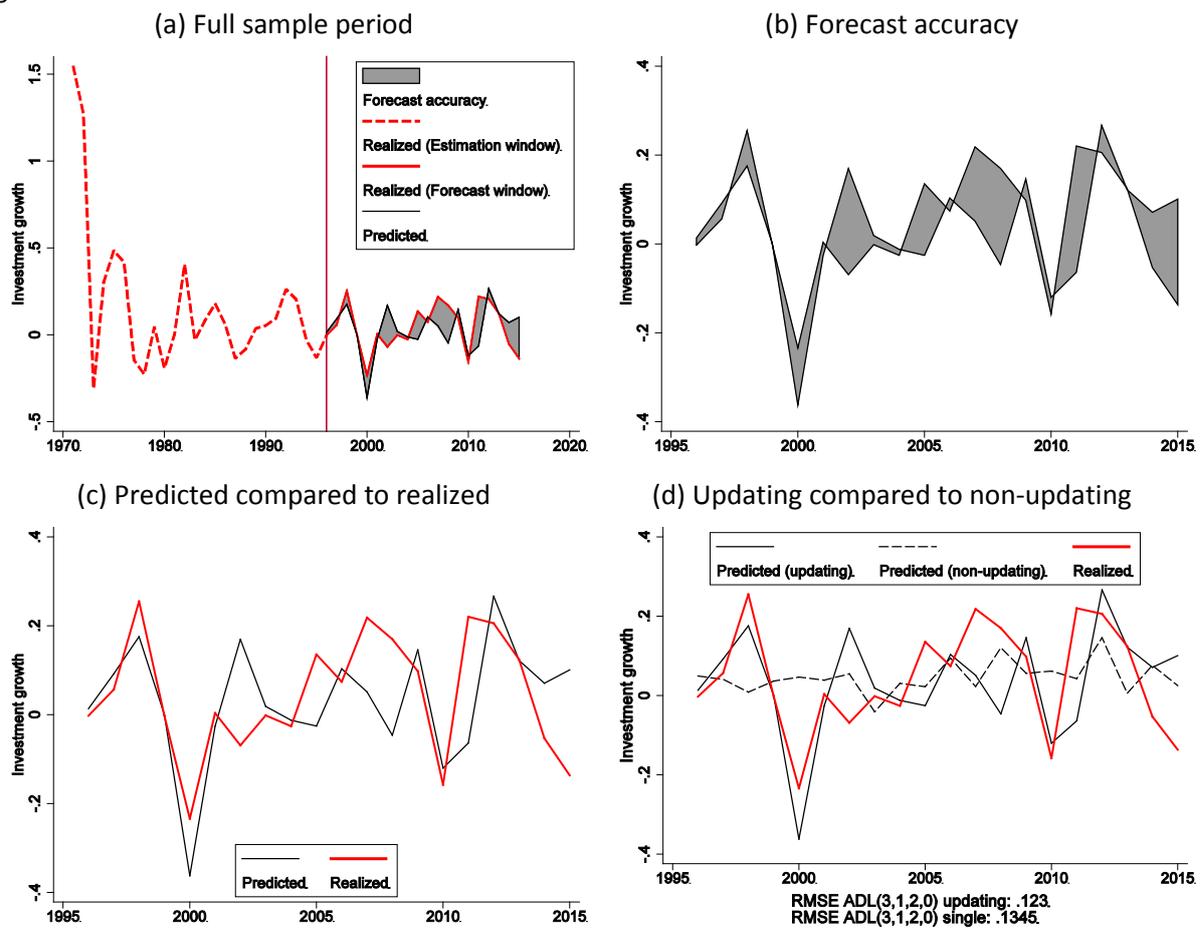


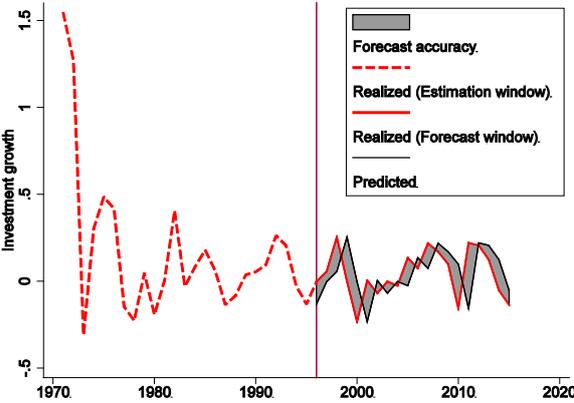
Figure 14 shows the predicted compared to realized aggregate investment. Subfigure (a) illustrates how the whole sample is initially divided into an estimation and an evaluation window. Other than ensuring that the former is sufficiently large to estimate the model and the latter is sufficiently large to carry out the hypothesis testing, the split point between the two is essentially arbitrary. It is quite possible that an alternate split point would yield different results. Subfigure (b) illustrates the forecast error of the $ADL(3,1,2,0)$ represented as the area between the realized and predicted aggregate investment. An important distinction is that this is an illustration of the forecast error and not the forecast accuracy. Subfigure (c) compares the realized and predicted aggregate investment throughout the evaluation window. Finally, subfigure (d) additionally adds the $ADL(3,1,2,0)$ forecast without re-estimation of the coefficients at every step. As it appears, re-estimating the models coefficients yields

better forecast accuracy. Analogous to Figure 14, Figure 15 shows the predictive power of the random walk benchmark model. Conforming to the conclusions made by the Diebold and Mariano test, the $ADL(3,1,2,0)$ model outperforms the random walk.

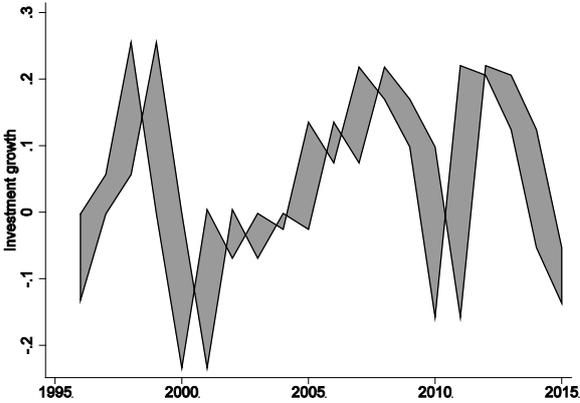
Figure 15: **Benchmark forecast accuracy**

Illustration of the forecast accuracy of the random walk benchmark model by comparing the predicted and realized investment growth. Subfigure (a) shows the investment growth during the full sample period, the predicted values during in forecast window and the forecast accuracy represented as the area between these time series. Subfigure (b) shows the forecast accuracy as the area between the predicted and realized investment growth. Subfigure (c) compares the predicted and realized investment growth during the forecast window. Subfigure (d) compares both the random walk and $ADL(3,1,2,0)$ forecast with the realized investment growth.

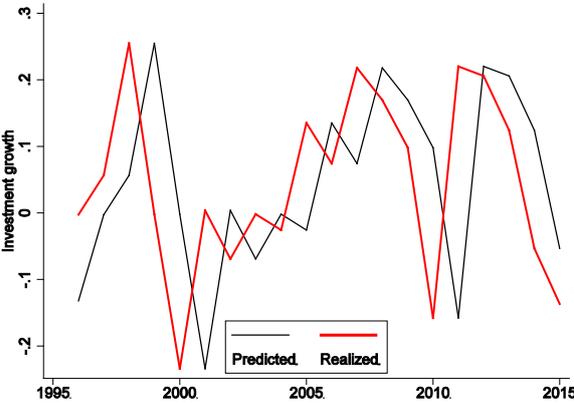
(a) Full sample period



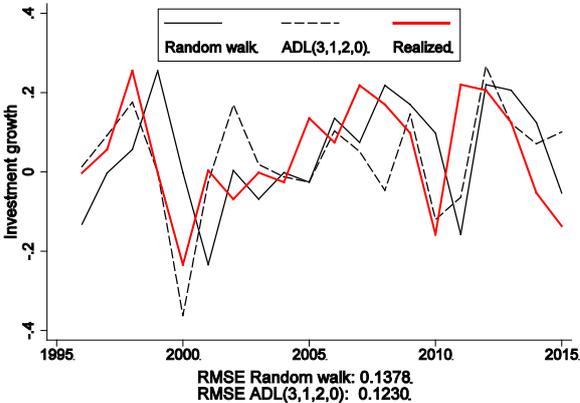
(b) Forecast accuracy



(c) Predicted compared to realized



(d) Benchmark compared to ADL



7.0 Conclusion

It is difficult to make predictions, especially about the future. In this paper, we have investigated the forecastability of aggregate oil and gas investments on the NCS. By empirically testing whether the set of statistical ADL models can significantly outperform a driftless random walk forecast, we reach the following conclusion: it is challenging but possible. Out of the 1080 different evaluated models, only 30 are capable of producing a RMSE significantly lower than the RMSE associated with random walk. Based on the $T_{R,M}$ test statistic of the model confidence set procedure, the model with the highest realization of forecast accuracy is the $ADL(3,1,2,0)$ model. This is a forecasting model with three lags of investment growth, one lag of crude oil price growth and two lags of realized oil price volatility. The model sheds light on the negative investment climate over the last few years as we have been in a situation with simultaneous low oil price and high oil price volatility. The oil price now seems to pick up. If better OPEC coordination in addition may provide reduced oil price volatility, one may speak of a recovery of the oil sector.

The choice of loss function is crucial for the choice of forecasting model. Among the models that generate predictions that outperform random walk, we have chosen the model with the traditional criteria of lowest RMSE. By applying a method that squares the errors, we implicitly assume risk aversion on part of the estimator. Whether the forecasts are to be used by government, oil companies or oil service companies, we believe this to be a reasonable presumption. Small deviations in aggregate investments are manageable. It is the large forecasting errors that are problematic, as they may lead to suboptimal decisions. For the government, e.g., a failure to predict a large reduction in aggregate activity may mean that accommodating measures like tax concessions and an increase in exploration acreage come too late. Failure to predict a large increase in aggregate investments may, e.g. for an oil company cause an underestimation of cost and lead to overinvestment.

Further insight about the forecastability of the aggregate oil and gas investment is gained. First, as to be expected, it appears to be beneficial for the models' forecasting accuracy to re-estimate the model coefficients as new information becomes available. Based on the results obtained from the $ADL(3,1,2,0)$ model using the various subsamples of data, the coefficients tend to change over time but predominantly the signs of the coefficients remain the same.

Reference

Andersen, T. G., T. Bollerslev, P. F. Christoffersen, and F. X. Diebold (2006). Volatility and correlation forecasting. *Handbook of economic forecasting* 1, 777-878.

Andersen, T. G. and T. Bollerslev (1998). Answering the skeptics: Yes, standard volatility models do provide accurate forecasts. *International economic review*, 885-905.

Bhattacharyya, Subhes C. "Energy Economics: Concepts, Issues, Markets and Governance." Springer 2011

Diebold, Francis X., and Robert S. Mariano. "Comparing predictive accuracy." *Journal of Business & economic statistics* (2012).

Dixit, Avinash K., and Robert S. Pindyck. *Investment under uncertainty*. Princeton university press, 1994.

Giacomini, Raffaella, and Clive WJ Granger. "Aggregation of space-time processes." *Journal of econometrics* 118, no. 1 (2004): 7-26.

Grunfeld, Yehuda, and Zvi Griliches. "Is aggregation necessarily bad?." *The Review of Economics and Statistics* (1960): 1-13.

Hansen, P. and A. Lunde (2003). Does anything beat a GARCH (1, 1)? A comparison based on test for superior predictive ability. In *Computational Intelligence for Financial Engineering, 2003. Proceedings. 2003 IEEE International Conference on*, pp. 301-307. IEEE.

Hendry, David F., and Kirstin Hubrich. "Combining disaggregate forecasts or combining disaggregate information to forecast an aggregate." *Journal of Business & Economic Statistics* (2012).

Kemp, A. and S. Kasim (2013), "An Econometric Model of Oil and Gas Exploration Development and Production in the UK Continental Shelf: A Systems Approach", *The Energy Journal*, Vol. 24, No. 2, pp. 113-141.

Kohn, Robert. "When is an aggregate of a time series efficiently forecast by its past?." *Journal of Econometrics* 18, no. 3 (1982): 337-349.

Maddala, Gangadharrao S., and Shaowen Wu. "A comparative study of unit root tests with panel data and a new simple test." *Oxford Bulletin of Economics and statistics* 61, no. S1 (1999): 631-652.

Mohn, K (2008). "Efforts and Efficiency in Oil Exploration: A Vector Error-Correction Approach", *Energy Journal* 29, 4, 53-78.

Mohn, K. and P. Osmundsen (2011). "Asymmetry and uncertainty in capital formation: an application to oil investment", *Applied Economics*, Volume 43, Issue 28, November 2011, 4387-4401.

Mohn, K. and P. Osmundsen (2008). "Exploration economics in a regulated petroleum province: The case of the Norwegian continental shelf", *Energy Economics* 30, 303-320.

Norwegian Petroleum Directorate (2013). "Evaluation of projects implemented on the Norwegian shelf", Norwegian Petroleum Directorate.

NOU (1999). "NOU1999:11 Investeringsutvalget. Analyse av investeringsutviklingen på kontinentalsokkelen. (The investment commission. "Analysis of the investment development on the continental shelf")

Osmundsen, Petter, Frank Asche, Klaus Mohn, and Bard Misund. "Valuation of international oil companies-the RoACE era." (2005).

Osmundsen, P., Roll, K., and R. Tveterås (2010), Exploration Drilling Productivity at the Norwegian Shelf, *Journal of Petroleum Science and Engineering*, 73, 122-128.

Osmundsen, P., K.H. Roll and R. Tveterås (2012), "Drilling speed - the relevance of experience", *Energy Economics* 34, 786-794.

Appendix A

Table A1: **Unit root test result cross-sectionally aggregate data**

Unit root test results for the cost overrun and surprise variables. The Fisher unit root test by Maddala and Wu (1999) with both the augmented Dickey-Fuller and Phillips-Perron specification. The tests were considered both with and without a deterministic trend. The null hypothesis under both the ADF and PP test states that a unit root is present in the data.

Variable	Fisher (ADF)		Fisher (PP)		n
	Without trend	With trend	Without trend	With trend	
Totaleinvesteringer	0.00	0.00	0.00	0.00	1643
Aggr_Brønn	0.00	0.00	0.00	0.00	1117
Aggr_Innr	0.00	0.00	0.00	0.00	1186
Aggr_Rør	0.00	0.00	0.00	0.00	284
Aggr_Landanl	0.00	0.00	0.73	0.00	45
Inv_Brønn_Nye_Fast	0.00	0.00	0.00	0.00	500
Inv_Brønn_Nye_Flyttblnrr	0.00	0.00	0.00	0.00	605
Inv_Brønn_Ufordelt		0.77	0.00	0.00	273
Inv_Innr_Eksist_Modifik	0.00	0.00	0.00	0.00	928
Inv_Innr_Nye_BunnfOgAndr	0.00	0.00	0.00	0.00	197
Inv_Innr_Nye_Flyttb	0.00	0.16	0.00	0.00	96
Inv_Innr_Nye_LastLager	1.00	1.00	0.48	0.42	8
Inv_Innr_Nye_UndervAnnl	0.00	0.00	0.00	0.00	361
Inv_Innr_Utbygg_Andre	0.00	0.00	0.00	0.00	220
Inv_Landanl	0.00	0.00	0.73	0.00	45
Inv_Rør	0.00	0.52	0.00	0.00	283
Inv_Rør_Landanl	1.00	1.00	1.00	1.00	16

Table A2: Unit root test result cross-sectional aggregated data

Unitroot test results for the economic activity variables. Augmented Dickey-Fuller and Phillips-Perron tests were utilized both with and without a deterministic trend. The null hypothesis under both the ADF and PP test states that a unit root is present in the data.

Variable	Augmented Dickey-Fuller		Phillips-Perron		n
	Without trend	With trend	Without trend	With trend	
Totaleinvesteringer	0.93	0.72	0.93	0.64	46
Aggr_Brønn	1.00	0.83	1.00	0.86	26
Aggr_Innr	0.87	0.56	0.85	0.51	46
Aggr_Rør	0.89	0.83	0.74	0.57	26
Aggr_Landanl	0.00	0.00	0.00	0.00	15
Inv_Brønn_Nye_Fast	0.71	0.69	0.72	0.65	26
Inv_Brønn_Nye_FlyttbInnr	1.00	0.98	1.00	0.98	26
Inv_Brønn_Ufordelt	0.99	0.97	0.99	0.98	10
Inv_Innr_Eksist_Modifik	0.60	0.66	0.57	0.63	46
Inv_Innr_Nye_BunnfOgAndr	0.88	0.61	0.89	0.56	24
Inv_Innr_Nye_Flyttb	0.28	0.27	0.18	0.16	21
Inv_Innr_Nye_LastLager	0.11	0.34	0.07	0.23	4
Inv_Innr_Nye_UndervAnnl	0.48	0.14	0.58	0.18	21
Inv_Innr_Utbygg_Andre	0.78	0.72	0.74	0.68	9
Inv_Landanl	0.00	0.00	0.00	0.00	15
Inv_Rør	0.38	0.45	0.22	0.24	26
Inv_Rør_Landanl	1.00	1.00	1.00	1.00	1

Appendix B

Table B1: **Model rank**

This table shows the forecast accuracy values for all models that statistically outperform the random walk benchmark. The p-value is the outcome of the Diebold-Mariano test of superior predictability comparing the ADL against the random walk benchmark, where random walk is superior under the null. Finally, $T_{R,M}$ and $T_{max,M}$ refer to the yielded result from the model confidence procedure. Outcome “A” implies that the model was eliminated from the set of superior models at a significance level of one percent, “B” at five percent, “C” at ten percent and “-” implies that the model is never eliminated.

Model	RMSE	p-value	$T_{R,M}$	$T_{max,M}$
ADL(1,1,1,1)	0.1159	0.0000	B	-
ADL(3,1,1,1)	0.1211	0.0084	B	-
ADL(1,1,2,1)	0.1228	0.0000	B	-
ADL(3,1,2,0)	0.1230	0.0813	-	-
ADL(2,1,1,1)	0.1240	0.0120	B	-
ADL(1,0,3,0)	0.1258	0.0285	-	-
ADL(3,1,2,1)	0.1266	0.0151	B	-
ADL(1,1,1,2)	0.1268	0.0354	B	-
ADL(1,0,4,0)	0.1272	0.0721	-	-
ADL(1,0,1,0)	0.1272	0.0453	-	-
ADL(3,1,1,0)	0.1273	0.0847	-	-
ADL(3,1,1,2)	0.1278	0.0141	B	-
ADL(2,1,2,1)	0.1283	0.0195	B	-
ADL(3,1,4,1)	0.1285	0.0104	-	-
ADL(1,0,2,0)	0.1289	0.0751	-	-
ADL(1,1,3,1)	0.1295	0.0002	B	-
ADL(2,1,1,2)	0.1304	0.0499	A	-
ADL(1,1,2,2)	0.1306	0.0205	B	-
ADL(3,1,4,2)	0.1314	0.0186	B	-
ADL(1,1,4,1)	0.1317	0.0034	B	-
ADL(3,1,0,0)	0.1318	0.0919	-	-
ADL(3,1,3,1)	0.1318	0.0257	B	-
ADL(2,1,3,1)	0.1335	0.0460	B	-
ADL(2,1,4,1)	0.1340	0.0803	B	-
ADL(1,1,0,1)	0.1341	0.0943	-	-
ADL(3,1,2,2)	0.1343	0.0254	B	-
ADL(2,1,2,2)	0.1344	0.0454	A	-
ADL(1,1,5,1)	0.1346	0.0270	B	-
ADL(1,0,0,0)	0.1349	0.0068	-	-
ADL(2,1,4,2)	0.1365	0.0978	B	-