Staff Working Paper ERSD-2006-05 **May 2006**

World Trade Organization
Economic Research and Statistics Division

FORECASTING TRADE

Alexander Keck	WTO
Alexander Raubold	University of Munich
Manuscript date	*April 2006*

FORECASTING TRADE

Alexander Keck[§] and Alexander Raubold[*]

April 2006

Abstract

This paper develops a set of time series models to provide short-term forecasts (6 to 18 months ahead) of international trade both at the global level and for selected regions. Our results compare favourably to other forecasts, notably by the International Monetary Fund, as measured by standard evaluation measures, such as the root mean square forecast error. In comparison to other models, our approach offers several methodological advantages, *inter alia*, a focus on import growth as the core variable, the avoidance of certain difficulties affecting the performance of structural models, the selection of variables and lags on the basis of theoretical considerations and empirical testing as well as a full documentation of the modelling process.

Keywords: Forecasting, time series, international trade, WTO

***JEL* classification**: F17, C53, C32, C22

[§] Economic Research and Statistics Division, World Trade Organization (WTO), Centre William Rappard, Rue de Lausanne 154, 1211 Geneva 21, Switzerland, alexander.keck@wto.org

[*] Department of Economics, University of Munich, Ludwigstr. 28, 80539 Munich, Germany, alexander.raubold@lrz.uni-muenchen.de

Disclaimer and acknowledgements: The opinions expressed in this paper should be attributed to the authors. They are not meant to represent the positions or opinions of the WTO and its Members and are without prejudice to Members' rights and obligations under the WTO.

The authors would like to thank Patrick Low, George Rapsomanikis and Robert Teh for helpful comments on earlier drafts of this paper. All remaining errors and omissions are the fault of the authors.

I. INTRODUCTION

Over the last years, the WTO Secretariat regularly has produced short-term forecasts of global trade growth to accompany its updates of international trade statistics and inform Members of possible future trends that can be discerned from the data. More specifically, forecasts are normally made for the current and following year when the International Trade Statistics (ITS) are published in October, which also contain an analysis of trends in global trade over the recent past. Then, in April, in the context of a Press Release,[1] which reports on first estimates of the trade performance in the previous year, these forecasts are updated in light of new data and trends. In both cases, the timing roughly coincides with forecasts made by other institutions, such as the International Monetary Fund (IMF) and Organisation for Economic Co-operation and Development (OECD) as the leading macroeconomic forecasters among international organizations.[2] These organizations predict a much larger set of variables, in particular GDP and inflation, and the forecasting process involves structural models at disaggregate levels.

The OECD acknowledges that the results of its structural model only serve "as a starting point to help animate the early stages of the OECD's forecasting round" (Rae and Turner, 2001: 4). The forecasting proper for the OECD Economic Outlook heavily relies on an exchange between country experts and topic specialists. Individual country projections emanating from this process are then fed into the OECD's larger INTERLINK world economic model in an iterative fashion to ensure consistency through international financial and trade linkages (OECD, 2004).[3] According to the OECD's own assessment, the resulting future trends of key macroeconomic variables are better characterized as "conditional projections" rather than forecasts, since they depend on the continuity of a range of influential factors, such as macroeconomic policies, nominal exchange rates and commodity price developments, in particular oil. Similarly, the IMF states that its MULTIMOD model "has not been designed to be a forecasting tool" (IMF, 2003). For its World Economic Outlook, it first produces projections that reflect the detailed knowledge and judgements of the IMF's country economists on the basis of information gathered through consultations with member countries (IMF, 2005). These are then used in MULTIMOD to generate a baseline scenario and simulate possible alternative settings.

It is part of the OECD's and IMF's mandate to identify structural economic problems and possible policy responses. Their projections are directed at policy-makers and are expected to provide a range of scenarios rather than mere forecasts (Lenain, 2002; IMF, 2005). In fact, the projections and analyses of the IMF contained in the biannual World Economic Outlook are "an integral element of the IMF's ongoing surveillance of economic development and policies in its member countries" (IMF, 2005). Their purpose is different from commercial activities, such as "Consensus Economics",[4] which present averages of private sector forecasts, mainly of investment banks, from a large range of countries. According to some, the latter have outperformed the OECD and IMF, especially on

---

[1] See, for instance, WTO (2006).

[2] The OECD and IMF each publish economic forecasts twice a year. The OECD *Economic Outlook* is usually published in June and December, the IMF's *World Economic Outlook* appears in May and September. More importantly, however, the underlying data are limited to what is available in April and September (possibly October in the case of the OECD). For an extensive analysis of forecasting by these two organizations see Batchelor (2000).

[3] For its Economic Outlooks, the OECD has also developed an "International Trade Model" for 24 countries, which has recently been simplified by abandoning separate relationships for manufacturing, non-manufacturing and services. It now focuses its attention on aggregate relationships for total trade in goods plus services. Pain et al. (2005) describe the updating of the trade matrices, model restrictions and the re-estimation of model parameters, notably a set of long-run elasticities.

[4] Each month, Consensus Economics taps economic and financial forecasters for their predictions of a range of variables, including growth, inflation, interest rates, exchange rates and others. Surveys cover more than 1,000 variables from over 70 countries in North America, Europe, Asia Pacific and Latin America. See http://www.consensuseconomics.com, visited on 10 November 2005.

individual variables, such as real GDP growth (Blix et al., 2001; Batchelor, 2000).[5] One reason might be that international institutions usually give projections based on policies announced by their member governments, whereas private forecasters, especially if they are based locally, may reflect insider information that allows them to anticipate likely policy changes and other relevant events. On the other hand, official sources may have an information advantage as far as the reliability, consistency and completeness of the data are concerned.

Both certain institutional and private forecasts have in common that they do not fully disclose their methodology. The expectations of country specialists are usually a key input, which are somehow amalgamated into the final forecast. About "Consensus Economics" this is all that is known. It simply states that experts in over 70 countries provide individual predictions on a monthly basis and that survey responses are then checked for accuracy, completeness and integrity and "processed using proprietary software" to arrive at simple arithmetic averages (Consensus Economics, 2005). While both the OECD and IMF, somewhere along the way, use well documented structural models in order to ensure consistency of individual predictions at the international level, it is not entirely clear to what extent and in what manner expert knowledge is utilized, be it to generate baselines or to adjust model outcomes.

WTO forecasts do not compete with any of these approaches. Quite to the contrary, forecasts are compared, checked against various data sources (including WTO trade data) and evaluated in the light of assumptions about other relevant variables, such as oil prices, and about the interplay of regional and sectoral developments. The construction of an alternative structural model of the world economy would neither be feasible due to resource constraints nor useful given the predictive power of existing structural models. However, the intimate knowledge of trade flows and policies enables the WTO Secretariat to examine trade growth predictions in more depth and focus on the dependency of trade on other economic factors rather than on simultaneous predictions of all endogenous variables.

The WTO forecasting process combines quantitative and qualitative elements. Econometric methods are "myopic" dealing only with "hard" data from the past. Expert judgement acts as a complement in order to take account of expectations for the future triggered by recent events or pending changes the effects of which have not yet been realized. Experts are also able to make conjectures about events that have occurred in the past but are not expected to recur in the future, such as natural disasters, or, vice versa, events that have not occurred in the past but are deemed likely to occur in the future, for instance a looming political crisis.[6]

This paper describes some of the quantitative aspects shoring up the WTO's appraisal of trade developments, notably a set of time series models used to produce a basic short-term forecast of global trade growth. In the next section, time series modelling is briefly introduced in terms of its advantages and disadvantages compared to structural models. Section III presents the data. Section IV describes how appropriate time series models can be identified. The fifth and main section consists of two parts: First, the actual models used are categorized in various types and the estimation

[5] Lenain (2002) is a reply to Blix et al. (2001) in which he also emphasizes other aspects that render a comparison difficult. In particular, the cut-off date for available data should coincide to make different forecasts comparable, which was not the case in Blix et al. (2001). However, even after the appropriate adjustments are made, 42 per cent of the (arguably simpler) private sector forecasts continue to perform better than the OECD (Lenain, 2002).

[6] Ideally, experts provide information that is not captured by the econometric forecast. While such judgement is indispensable to improve forecast performance, there are also risks that experts may see more in the data or in recent/expected events than is warranted, for instance due to "double counting" or simply "optimism". For more on the integration of econometric methods and expert judgement for time series forecasting see Armstrong and Collopy (1998). The authors propose several "integration" approaches, survey an extensive body of studies to examine the gains of integration, develop a set of screening criteria to identify the conditions under which integration is feasible and useful and, finally, develop some principles for integration. For several illustrations on how to formalize the interaction between decision-makers and econometricians in the forecast modelling process see, for instance, Manganelli (2006).

as well as forecasting procedures are explained. Second, the results obtained in terms of both ex post and actual forecasts are discussed for both the OECD-25 as well as selected countries from that group. Section VI concludes.

II. TIME SERIES MODELLING

Time series forecasting models use the past movements of variables in order to predict their future behaviour.[7] Unlike structural models that relate the variable of interest to a set of other variables in a causal framework, time series regressions need not be based on economic theory. What counts is their explanatory power, the precision of coefficients and, in order to make predictions, the reliability of the estimated equation once applied out-of-sample (Stock and Watson, 2003). While unable to explain causation, a time series model can still produce quite accurate forecasts, especially when causal relationships are manifold and multidirectional like in the global economy. In fact, owing to the complexity of international economic relations, large structural models are likely to suffer from omitted variable bias, misspecifications, simultaneous causality and other problems leading to substantial forecast errors.[8]

Figure 1 below shows the development of quarterly growth rates of imports and domestic demand for the OECD-25.[9] It can be seen that imports are more volatile than GDP.[10] Wider swings in the observed values of a variable imply a higher degree of uncertainty about influencing factors for our analysis. Hence, while a structural model may yield good results in forecasting more stable variables, such as GDP growth, a time series model may provide a much simpler and more effective means to make forecasts when causal relationship are less clear, like in the case of import growth. A structural model would seek to establish the links between imports and a range of other endogenous and exogenous variables on the basis of economic theory. Conversely, a time series analysis of imports relates that variable to its past values and to any additional variables (and their respective lags) that can account for part of the random nature of its past movements. Even if the resulting coefficients have no causal interpretations, a time series model can produce reliable forecasts if the regression explains much of the variation and is stable over time.

---

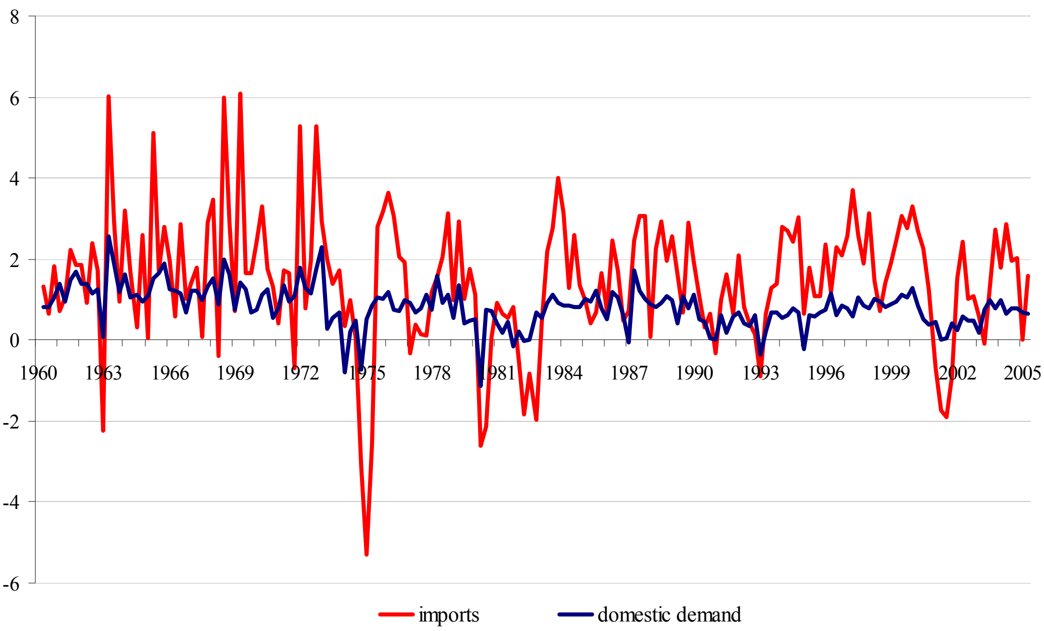
[7] For an introduction into time series analysis see Pindyck and Rubinfeld (1998).

[8] See Diebold (1997) for a critical account of the history of "systems-of-equations" econometrics and an overview of recent research marrying non-structural forecasting traditions with new approaches to macroeconomic modelling. Hendry and Clements (2003) develop a forecast error taxonomy, focusing in particular on shifts in the coefficients of deterministic terms, and propose ten principles to evaluate forecast performance and ten areas for further research if stationarity and well-specified models cannot be presumed.

[9] OECD-25 refers to Australia, Austria, Belgium, Canada, Denmark, Finland, France, Germany, Greece, Iceland, Ireland, Italy, Japan, Luxembourg, Mexico, Netherlands, New Zealand, Norway, Portugal, Spain, Sweden, Switzerland, Turkey, United Kingdom, United States, for which longer time series are available. In the meantime, the Czech Republic, Hungary, Korea, Poland and Slovak Republic have joined the OECD.

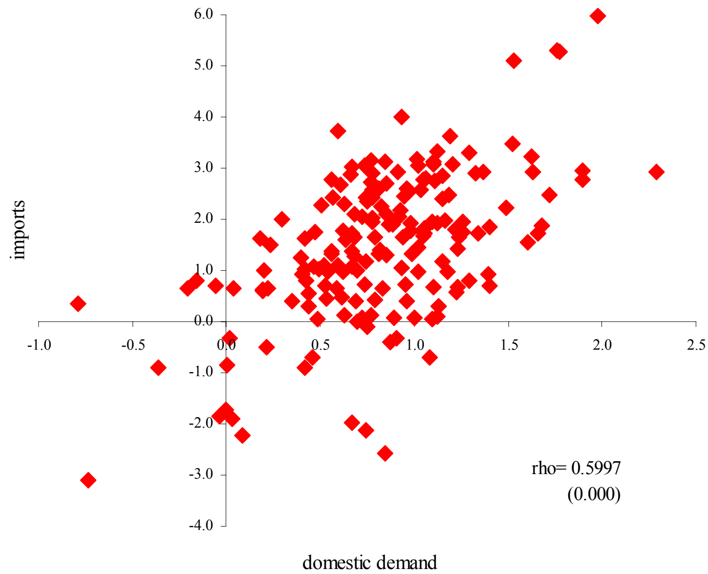
[10] This phenomenon hitherto has not been explained by economic theory and would certainly merit further study.

Figure 1: Quarterly growth rates of OECD-25 imports and domestic demand, 1960-2005 (per cent)



The selection of additional variables need not be void of theoretical considerations. For instance, from a theoretical point of view, domestic demand should impact the amount of imported goods and services positively. A simple plot of the data, like in Figure 2, featuring a positive correlation (at a 1 per cent significance level) between the quarterly growth rates in domestic demand and imports, supports that view. However, Figure 1 has also shown that growth rates are more volatile for imports than domestic demand, which poses additional difficulties for forecasting despite the positive correlation. Furthermore, the increase in imports, especially over the last 20 years,[11] has been more pronounced than the increase in domestic demand. Both a higher volatility of quarterly growth rates and a stronger absolute increase of imports over time suggest that other factors are at work that could usefully be considered in our forecasting exercise.

Figure 2: Correlation between quarterly growth rates of imports and domestic demand, 1960-2005, (per cent)



[11] See Figure 4 below.

When exogenous variables are added, projections of their future values may be required in order to make forecasts. This applies to both multivariate time series and structural models. Future values can be taken from other outside sources, such as central banks for monetary targets, or must themselves be estimated, for instance through other macroeconomic models or separate time-series analyses. Forecast errors in the exogenous variables may filter through the model and lead to unacceptably vague forecasts of the variable of interest. Simple univariate time series estimations or vector autoregressions (VAR), whereby two or more time series variables are determined simultaneously, can be used to forecast the variable of interest directly.[12]

To summarize, some of the distinguishing features of our approach are the following:

- Import growth is forecasted directly and not as a "by-product" of models constructed principally to forecast other variables or conduct policy simulations;
- The use of time series analysis avoids some of the difficulties affecting the performance of more complex structural models;
- In turn, we can include a more sophisticated lag structure of imports and other variables that are selected on the basis of theoretical considerations and tested for their explanatory power of variations in trade over time;[13]
- The modelling process is fully documented and can be replicated by other researchers.

III. DATA

Our analysis mainly relies on OECD Quarterly National Accounts (QNA) data.[14] A first dataset is constructed for the group of former OECD-25 countries. The raw data are sourced from the OECD Olisnet database in aggregated form. The data are seasonally adjusted and expressed in constant dollars. From this, we calculate volume indices (year 2000 = 100). The dependent variable are quarterly imports of goods and services by the former OECD-25 countries.[15] The maximum length of the time series we use (published in October 2005) stretches from 1960 to the second quarter of 2005.

A second variable is domestic demand, consisting of the sum of private consumption, gross fixed capital formation and government spending. These data are available from the OECD for the same time frame and in the same format as imports. Economic theory suggests that domestic demand is more closely related to import developments than GDP, although the difference is not large. GDP includes export production that could be responsible for large parts of GDP growth during times of stagnant domestic demand, as has been the case in Germany for quite some time. Future values for

[12] Diebold (2004) holds that simple, parsimonious models tend to be best for out-of-sample forecasting in many areas of business, finance and economics.

[13] Although being theoretically based, our time series models are specified differently for different regions/countries in view of data availability and econometric test results. This approach contrasts with, for instance, a theory-based gravity equation that could be estimated separately for each country in order to gain insights in country-specific elasticities (different estimated coefficients). However, the basic set-up of the model remains the same for each country. Structural models, such as the "Fair" macroeconomic model which is freely available on the internet, feature a unified framework linking countries together, but use a different variable range (up to 15) at the country level, owing also to restricted data availability in some countries. See Fair (2004). We also run this model in order to acquire further benchmarks for our forecasts. See http://fairmodel.econ.yale.edu/fp/fp.htm.

[14] The OECD regularly revises their data series. Newly downloaded time series from Olisnet may not exactly reproduce the results presented in this paper.

[15] We focus on imports rather than exports to measure trade developments, since import data collected by customs authorities are more reliable and available earlier. Also, there is a theoretical link between domestic demand and imports. For exports, a world demand variable would need to be constructed, for which data are not readily available. In any case, at the OECD level, the difference between export and import growth over recent years has only been minor.

the forecast time period are constructed from IMF predictions of GDP.[16] Instead of contemporaneous domestic demand, some estimation equations include a leading indicator, the ifo World Economic Climate index,[17] which reflects expectations two quarters ahead. Both imports and domestic demand are also estimated simultaneously using VAR, which avoids the need to obtain contemporaneous values for the exogenous variable.

Finally, specific commodity prices, especially the oil price, may be relevant in explaining import growth.[18] For the oil price, we use the petroleum average crude price, which is composed of the Dubai, UK Brent und West Texas Intermediate petroleum prices in dollars per barrel. The United States CPI is used to determine real oil prices in constant 2000 dollars. These data are sourced from the IMF's International Financial Statistics (IFS). For future time periods, oil prices are assumed to remain constant at the level of the last quarter for which data are available.[19]

For selected countries/regions, namely the United States, the European Monetary Union (EMU) (i.e. the 12 countries of the Euro area), Japan and Germany, a second set of data is constructed. Again, import and domestic demand data are in real terms (seasonally adjusted, constant dollars) expressed as volume indices (year 2000 = 100). For Germany, we use both the country-specific ifo index and, alternatively, the ZEW indicator,[20] which (rather than ifo's business-oriented outlook) reflects the expectations of financial institutions for the next six months. Again, we also include the oil price. Import and demand data are from the OECD, oil price data are sourced from the IMF's IFS.

Since the data are observed at quarterly intervals, they might exhibit seasonality. Macroeconomic forecasting is geared towards projecting non-seasonal fluctuations,[21] and seasonality should therefore be removed to the extent possible.[22] For the data we use, some seasonal adjustment has already been

---

[16] Of course, we could generate a set of alternative forecasts conditional on differing assumptions about the exogenous variables, such as GDP.

[17] The ifo World Economic Climate index is based on information gathered during the quarterly World Economic Surveys (WES). The WES are conducted in co-operation with the International Chamber of Commerce in Paris (ICC) and receive financial support from the European Commission. The surveys consist of qualitative information drawn from appraisals and expectations by economic experts of multinational firms and institutions in the countries surveyed. For more information on the ifo Institute for Economic Research and this indicator see http://www.cesifo-group.de/portal/page?_pageid=36,34788&_dad=portal&_schema=PORTAL, visited on 17 October 2005.

[18] See, for instance, Fair (2004) who uses similar variables (lagged imports, consumption, investment and government spending as well as the price of imported goods relative to the price of domestic goods) in order to predict imports in the context of a large structural model. Rather than deriving future values of exogenous variables from expert assessments or simple autoregressive schemes, like in the Fair model, our approach could supply more objective forecasts based on other influential factors and extensive empirical testing for use in structural models with trade as an exogenous variable.

[19] Like in the case of domestic demand, we could also use forecasts from other institutions to project the oil price into the future. Forward prices could be taken from the monthly oil market reports by the International Energy Agency (IEA), available at http://omrpublic.iea.org/. Of course, prices are more difficult to forecast over a two to six quarter horizon due to their higher volatility compared to real variables, such as domestic demand. Alternatively, we could fit any exogenous variable with an autoregressive (AR) model determining lag length in the usual manner. For the current application of our model, we have opted to stick with the assumption of constant real oil prices at a historically rather elevated (although not record) level. This assumption has proven quite realistic, as most forecasters in early 2006 continue to foresee slightly higher oil prices than in 2005 paired with small increases in inflation. See, for instance, United Nations (2006).

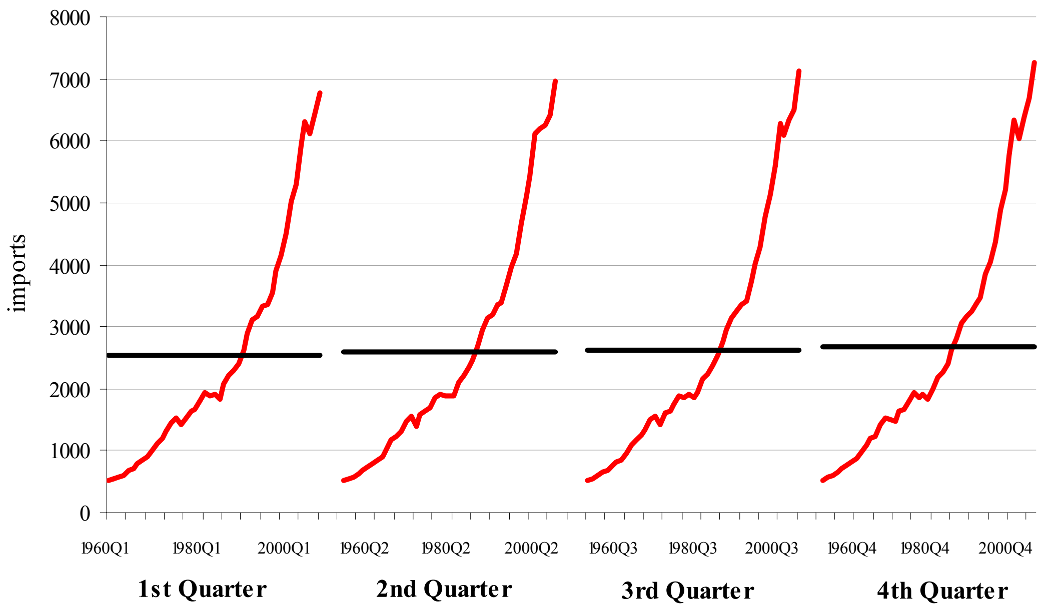
[20] The Zentrum für Europäische Wirtschaftsforschung (ZEW) constructs this leading indicator of the economic situation in Germany on the basis of monthly surveys of the expectations of 350 financial experts. For more see www.zew.de.

[21] Conversely, business forecasters work with seasonally unadjusted data, as they need to forecast all variation in a time series and not just the non-seasonal part.

[22] Abeysinghe (1994) observes that the use of seasonal dummies, generally in economic forecasting, tends to produce poor forecasts. It is therefore preferable to use seasonally adjusted series instead of mis-specifying the seasonal. For instance, Box-Jenkins analysis requires seasonally adjusted time series.

made by the OECD.[23] Remaining seasonality may be detected graphically by a seasonal subseries plot. The quarterly means, shown by the horizontal lines in Figure 3, are almost identical, indicating the absence of seasonality in our import data. A comparison of the first and the fourth quarter exhibits the largest difference in quarterly means, with the latter exceeding the former by 4.5 percent. By contrast, seasonally unadjusted import data for the OECD-25 countries show a difference in the mean between the first and the fourth quarter of 10.5 percent.[24] This means that although seasonal factors are mostly removed we still expect some degree of correlation between current imports and imports four quarters ago. However, there are no significant seasonal patterns that would have to be modelled explicitly.

Figure 3: OECD-25 imports, seasonally adjusted, 1960-2004, by quarter (million dollars, constant 1995 prices)



IV. IDENTIFICATION OF THE APPROPRIATE TIME SERIES MODEL

In order to identify a suitable time series model, simple plots of the data are usually a good starting point. For the OECD-25, Figure 4 shows the quarterly development of imports, domestic demand and the petroleum price (all in real terms) for the years 1960 to 2005.[25] During that period of time, imports and domestic demand increase more or less continuously. The two series each follow their own trend, which suggests that both are nonstationary[26] and have to be tested for unit roots. However,

[23] We make our own moving average adjustment on the unadjusted OECD data without obtaining better results. A weighted moving average is calculated over 5 quarterly observations centred at each observation in turn. Each observation is then divided by its moving average, and a simple average of all of the resulting terms is calculated for each quarter Q1 to Q4. These seasonal factors are normalized to average to 1, and the seasonally adjusted time series is computed by dividing the old series through the seasonal factors.

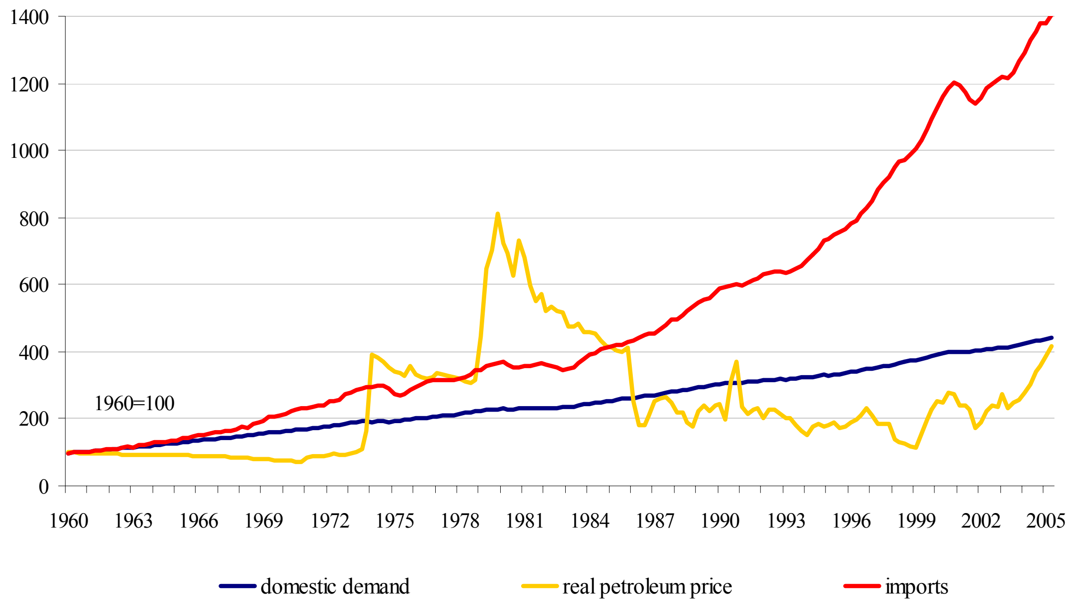
[24] Compared to other time series, even the unadjusted import data shows no strong seasonal component, and the OECD adjustment removes most of the minor seasonality left.

[25] The series are converted to a common base of 100 for the year 1960 for ease of comparison of trends.

[26] If the underlying stochastic structure of a time series is changing over time, i.e. if it is nonstationary, it is impossible to predict the future accurately on the basis of the past. At a minimum, its mean and covariance structure (i.e. the covariances between current and past values) have to be stable over time ("covariance stationarity"). In other words, the autocovariance between two observations y_t and $y_{t-\tau}$ of a time series should only depend on the displacement τ and not on time t. This condition is also called "weak stationarity". Therefore, transforming a time series variable such that its distribution does not change over time is crucial in

while domestic demand has quadrupled within 45 years, trade has increased by a factor of 14. The widening gap between the volume of imports and domestic demand, which might be explained with the increased international fragmentation of production,[27] does not appear to support the hypothesis of co-integrated time series following a common trend.[28] Given the strong fluctuations of the oil price, no major insights about its relationship with imports can be derived from the graph. Further statistical testing is required in order to determine the stationarity of a time series and the relationship with other variables. Similarly, plots of the autocorrelation structure or, alternatively, formal criteria can be used to select the optimal number of lags for each variable.[29]

Figure 4: Volumes of OECD-25 imports and domestic demand and real petroleum price, 1960-2005, (1960 = 100)



A. STATIONARITY

For univariate analyses, a Box-Jenkins[30] graphical plot is a straightforward tool to examine the existence of a unit root. While it is impossible to model the whole underlying stochastic process, the autocorrelation (AC) function provides at least a partial description of this process. The plot of the

---

making forecasts. In general, a time series is stationary if the series looks flat, has no trend and shows no seasonal fluctuations. For an in-depth discussion see Diebold (2004).

[27] Owing to declines in transport and communication costs as well as tariffs over the last decades different stages of production of a product have increasingly been geographically separated (or fragmented) to different countries. Yi (2003) finds that more than half of United States trade growth since the 1962 can be explained by taking account of such vertical specialization. Hummels et al. (2001) note that the import content of a country's exports increased by 30 per cent between 1970 and 1990, using data for 13 OECD countries. Other studies on intra-firm trade (or on trade in intermediate inputs between parent companies and foreign affiliates more specifically) as well as on international outsourcing reach similar conclusions.

[28] An absence of cointegration would not be surprising in view of the fact that imports refer to gross amounts, whereas domestic demand is determined on a value-added basis.

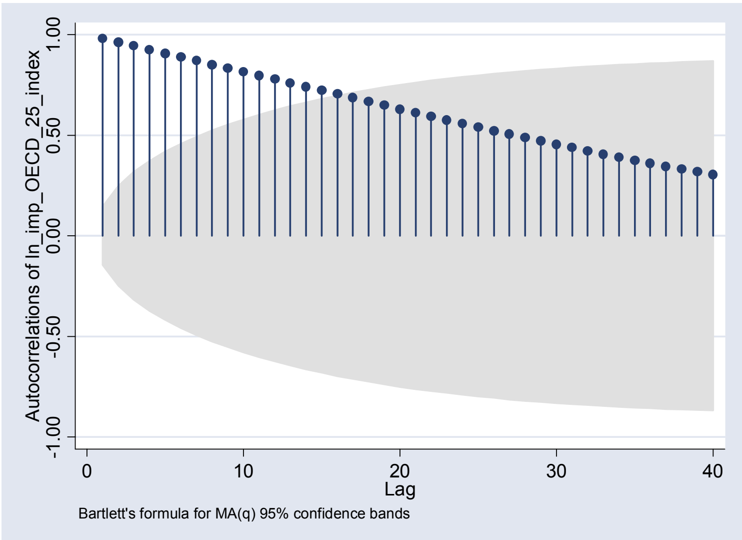
[29] These tests and analyses, which are only discussed in an exemplary fashion in the paper, have been carried out for each variable and model specification.

[30] Box and Jenkins provide a systematic methodology for identifying and estimating models that could incorporate autoregressive as well as moving average approaches. The Box-Jenkins forecasting method only uses the own past behaviour of a variable to forecast its future values and is therefore univariate.

AC function (the so-called correlogram) illustrates the degree to which a given value is correlated with its past values.[31]

Figure 5 shows that the time series of the logs of OECD-25 imports is non-random, since its autocorrelations are non-zero. The series has a rather high degree of autocorrelation between adjacent and near-adjacent observations. The slow decay of logarithmized imports for the structure under consideration (40 lags) clearly indicates that the time series exhibits a unit root. Bartlett's formula is used in order to verify whether a particular value of the sample AC function is equal to zero. The 95 per cent confidence interval in Figure 5 reveals that autocorrelations are not significantly different from zero only after the fifteenth lag.

Figure 5: Correlogram of OECD-25 imports (in logs)



The autocorrelation plot of the first differences of logarithmized OECD-25 imports in Figure 6 displays a mixture of an exponentially decreasing and dampened sinusoidal process, which eventually decays to zero. This pattern provides a strong indication for the fact that the nonstationary component of the time series has been removed by first differencing.[32] The graph also shows that the differenced time series fulfils the condition of weak dependence. As the time distance between two observations gets large, the observations are almost independent. In fact, already the fourth lag is not significantly different from zero. With this transformation, the two conditions of the time series equivalent of the i.i.d. assumption for cross-sectional data hold.[33]

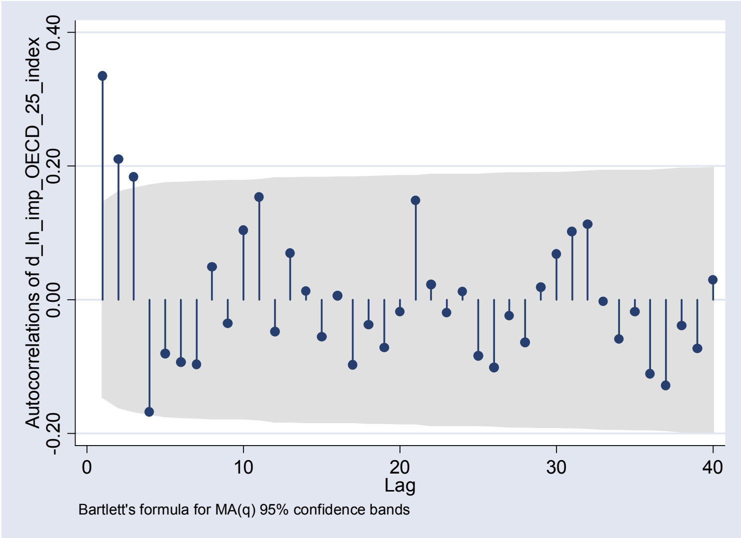
---

[31] The autocorrelation function gives the total correlation between an observation and a given lag, i.e. it includes all the "indirect effects" on the observation of the intervening lags. The partial autocorrelation function (see further below) measures the association of two observations of the series after controlling for the correlation with observations in between these two values.

[32] I.e. in this case, a trend has been removed. If a time series has a stochastic trend, i.e. if it has a unit root, the first difference of the series does not have a trend. If a nonstationary series possesses this desirable property, it is termed homogeneous.

[33] I.e. the distribution of the variable does not change over time ("identically distributed") and observations are "independently distributed" when they are separated by long time periods ("weak" dependence"). The latter condition ensures that there is sufficient randomness in large samples for the law of large numbers and the central limit theorem to hold.

Figure 6: Correlogram of first differences of OECD-25 imports (in logs)



As a complement to Box-Jenkins, we apply the Augmented Dickey-Fuller test to check for stationarity. We cannot reject the null hypothesis of the existence of a unit root (p-value equal to 0.523) for the logs of the OECD-25 import data. This corroborates the conclusions drawn from the Box-Jenkins method. Similarly, the augmented Dickey-Fuller test for the transformed series confirms at the one percent significance level that the time series in first differences is stationary.[34] It comes in handy that the resulting import time series in logarithmic form, which is integrated of the order one, amounts to growth rates, i.e. the variable format we are interested in. The same transformations have been made for domestic demand. For the oil price, the transformed variable can be interpreted as a proxy for energy inflation.

B. LAG STRUCTURE

After having identified the necessary degree of differencing *d* (to get a stationary time series), we turn to the "lag structure" of the model, i.e. the number of the autoregressive and moving average terms that should be included. Too few lags result in the omission of potentially valuable information contained in more distant values. By the same token, too many lags entail a loss of usable observations and an estimation of more coefficients than necessary, thus introducing additional estimation errors into the forecast. Following Box-Jenkins, we examine both the autocorrelation and partial autocorrelation functions in order to determine an integrated autoregressive-moving average (ARIMA(p,d,q)) model, i.e. find the optimal combination of p autoregressive terms and q moving average terms of a time series integrated of order d. We begin with the autocorrelation plot, which reveals the presence of autoregressive (AR) and/or moving average (MA) processes.[35] The autocorrelation function in Figure 6 above features an exponential decrease and then a sine wave-like pattern, i.e. a typical AR(p) process of an order higher than one and no moving average process.[36] A

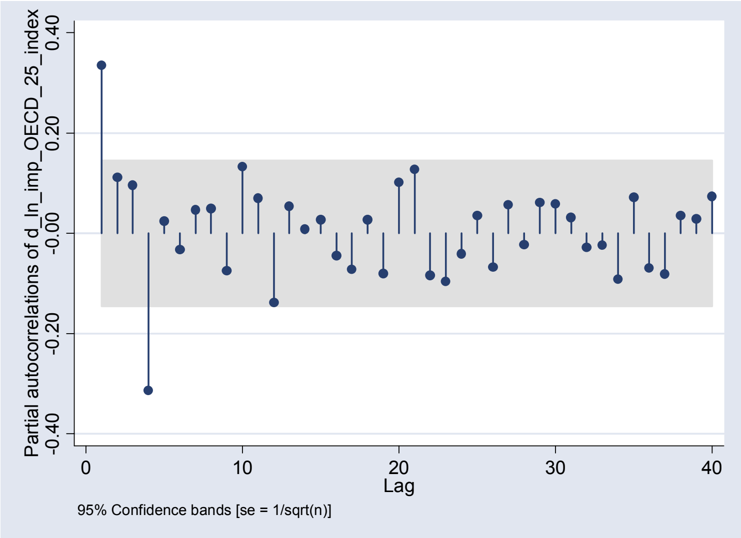
[34] This result is also confirmed by the Phillips-Perron test, which tests for a unit root against the alternative hypothesis of a stationary series with a structural break, i.e. a one-time change in the mean.

[35] Both the Ljung–Box portmanteau (Q) test and Bartlett's periodogram-based test for white noise reject at a highly significant level the hypothesis that our time series is generated by a white noise process.

[36] An AR(1) process is characterized by an autocorrelation coefficient θ that either monotonically declines (if θ>0) or exhibits a "sawtooth"pattern (if θ<0). If it monotonically declines in a dampened sine wave pattern, a higher-order AR process is indicated, as in the present case. A pure MA(q) process would sharply drop off to zero at one lag past the order of the process q. A mixed ARIMA(p,d,q) process would therefore show a rather irregular pattern over the initial q terms (for instance in the form of distinctive spikes) that are complicated functions of both AR and MA parameters, and then follow a simple AR(p) pattern.

partial autocorrelation plot is useful to determine the order of the AR and/or MA process, i.e. the number of lags. The plot of the partial autocorrelation (PAC) function of logarithmized OECD-25 imports in first differences in Figure 7 gives a good indication of the appropriate number of lags: The significant spike at lag one and the partially uncorrelated lags thereafter, which are not statistically significant different from zero, point to an autoregressive process of order one. However, the plot shows a statistically significant spike at lag four, which, as expected, hints at the remaining seasonality in the seasonally adjusted data.[37]

Figure 7: Partial correlogram of first differences of OECD-25 imports (in logs)



On the basis of this graphical analysis the model should be specified as an ARIMA (1,1,0) or (2,1,0) model, i.e. an ARIMA (p,d,q) model that is integrated of order 1 (d=1) with an optimal number of one or two autoregressive lags (p=1 or 2). In spite of only one distinctive spike at lag one in both the AC and PAC functions, we opt for two autoregressive lags. In the case of one lag as optimal lag structure, the AC function would have exhibited an exponential decay rather than the sine wave-like pattern we find (or set of exponential decays). The Box-Jenkins analysis has not detected an underlying moving average process (q=0).[38]

Besides the Box-Jenkins method for selecting the most appropriate lag structure, which quite heavily relies on the experience of the modeller in discerning patterns in the AC and PAC functions, information criteria are commonly used in time series analysis. Such formal criteria can help automate the model identification process. We have focused on the Akaike information criterion, which penalizes the addition of right-hand side variables, and hence reduction of degrees of freedom, more heavily than other criteria, such as the corrected R^2. Further lags are added until the Akaike information criterion reaches its minimum. In the case of OECD-25 imports, this is the case for five lags. This number of lags is also confirmed by the Schwartz's Bayesian information criterion. In a similar fashion, we obtain six lags for OECD-25 domestic demand and seven lags for the oil price.

[37] In fact, we could include the fourth lag in our time series regression to account for seasonality. This has been done in any case, since the formal information criteria we consider below suggest an even higher lag structure. Also, TSP takes the quarterly cycle into account by default if the time structure is specified as quarterly FREQ Q=4.

[38] For the assumptions of an ARIMA model to hold, the residuals should resemble a white noise process. This is confirmed by a plot of the residuals and their correlogram which do not show any structural change and serial correlation. In addition, the Ljung–Box portmanteau (Q) test does not reveal any significant autocorrelations among the residuals. In other words, the model successfully captures all systematic movements in the data with the remaining residuals being essentially random.

C. GRANGER CAUSALITY AND COINTEGRATION

In order to know whether the inclusion of domestic demand and the oil price contributes to the explanatory power of the regression, we apply a "Granger-causality" test. Variable X is said to "Granger-cause" Y if the addition of past values of X to the regression results in an improvement in the prediction.[39] The number of lags to be included in these regressions is arbitrary and we have chosen to run the tests for the lags that have proven optimal for each individual variable. The tests indicate that domestic demand and the oil price both jointly and individually "Granger-cause" import growth, although the results for the domestic demand variable alone may be sensitive to the choice of lags. The test values, including for domestic demand taken separately at its optimal lag structure, are highly significant and we therefore expect the inclusion of both variables and their appropriate lags in our models to improve our forecasts.

Since import and domestic demand volumes as well as the real oil price clearly exhibit different trends (see Figure 4), they cannot be cointegrated. As discussed above, each series has been integrated of order one in order to remove their individual trends. Once the time series are made stationary, they have to be re-examined for possible cointegration. If this was the case, our forecast could be improved by defining a vector error correction model (VECM) which makes use of the common stochastic trend.[40] We apply the Augmented Dickey-Fuller and Johansen cointegration tests to the growth of imports, domestic demand as well as the real oil price. On the basis of these tests, the null hypothesis of no cointegration of the growth rates of imports and domestic demand can be rejected. The same is not the case for the growth rates of imports and the real oil price. Even though the growth rates of imports and domestic demand are cointegrated, Stock and Watson (2003) warn that cointegration tests can be misleading, since they frequently improperly reject or fail to reject the null hypothesis of no cointegration. If variables that are not cointegrated are modelled using a VECM, the error term is assumed to have an unit root, thus introducing a trend that can lead to a poor forecasting performance of the model. We are not in a position to postulate cointegration on the basis of theoretical arguments and therefore estimate VECMs only as a robustness check to our VARs.[41] In any case, for the short-run forecasts we are interested in it is not advisable to overemphasize low frequency (i.e. long-run) variation of the data.

V. ESTIMATIONS AND FORECASTS

In this section, we first describe the estimation of model parameters and categorize the various specifications we have estimated into several types. The second part discusses the application of the models. This includes both ex post forecasts (in-sample forecasts) and the extrapolation of the time series beyond the sample period using the estimated parameters (out-of-sample forecasts). Results are compared among themselves and in relation to other forecasts on the basis of standard evaluation measures, most notably the root mean square forecast error (RMSFE).

[39] "Granger causality" is concerned with how information is sequenced in time and how useful it is in making predictions. Hence, the concept differs from the common understanding of "causation" in an economic sense.

[40] As has been seen in Figure 1, the growth rates of the stationary time series of both imports and domestic demand show considerable variation over time, but usually move in the same direction, with the swings in imports being more extreme than those in domestic demand. As a rule of thumb, an elasticity of import to GDP growth of about 2 on average has often been presumed, up from lower values of about 1.8 some 20 years ago.

[41] See Abeysinghe (1998) who follows a similar approach.

A. THE MODELS

The two main types of models we employ are univariate and multivariate time series equations.[42] The univariate models in this paper are autoregressive (AR(p)) models of various orders p, for instance of the second (Box-Jenkins analysis) and fifth order (Akaike and Schwartz's Bayesian information criteria) in the case of OECD-25 imports. Hence, they are of the following form (for instance AR(5) scheme for imports M at time t, with u_t as the error term):

$$M_t = \beta_0 + \beta_1 M_{t-1} + \beta_2 M_{t-2} + \ldots + \beta_5 M_{t-5} + u_t$$

We assume $E(u_t|M_{t-1}, \ldots) = 0$, i.e. the errors u_t are serially uncorrelated, which allows us to pursue a simple ordinary least-square (OLS) estimation procedure.

The multivariate specifications are either autoregressive distributed lag (ADL) or vector autoregression (VAR) models. The ADL(p,q) forecast model for imports with domestic demand includes five own lags (p = 5 lags) as well as the contemporaneous values and six lags of domestic demand (q = 6 lags). It is of the form:

$$M_t = \beta_0 + \beta_1 M_{t-1} + \beta_2 M_{t-2} + \ldots + \beta_5 M_{t-5} + \delta_0 DD_t + \delta_1 DD_{t-1} + \ldots + \delta_6 DD_{t-6} + u_t$$

While our ADL models have been estimated using OLS under the standard assumption that the errors have a conditional mean of zero given all past values of M and DD and constant variance, we notice in the plot of United States import growth rates (see Figure 8) that the absolute percentage changes, on average, are larger in the more distant past, for example in the early 1970s, than throughout the 1990s and thereafter. The chart also shows volatility clusters, albeit less pronounced, for the OECD-25.

A plot of the residuals confirms our suspicion of volatility clustering. This means that the variance of the error term is not constant and clusters over time, i.e. a small variance of the regression error in one period tends to imply a small variance in the next (time-varying heteroskedasticity). Formally, the Lagrange multiplier test for autoregressive conditional heteroskedasticity (ARCH) effects allows us to reject the null hypothesis of no ARCH effects.[43] In order to take account of periods with higher volatility and those with relative tranquillity – and hence exploit the fact that forecasting is "easier" at some times than others - we re-estimate our regressions using a generalized autoregressive conditional heteroskedasticity (GARCH) model which generates more efficient parameter estimates. In the GARCH model - in addition to the ADL equation - the error u_t is modelled explicitly as being normally distributed with mean zero and variance σ_t^2. The variance depends both on its own lags and on the lags of the squared error. In our models, we only estimate GARCH (1,1) equations, i.e. we only consider the first lags of u_t and σ_t:[44]

---

[42] This section large relies on the modelling procedures described in Stock and Watson (2003). Univariate equations relate a time series variable, imports in this case, to its past values. This is called an autoregressive model with p lags (AR(p)). Multivariate equations include the lags of additional predictors besides the lagged values of the dependent variable. In the simplest case there are p lags of the dependent variable and q lags of one additional predictor. This is called an autoregressive distributed model (ADL(p,q)). In several time series regression, we include multiple predictors, i.e. besides the lagged values of the dependent variable (p), the equation contains the lags of k additional predictors ($q_1, \ldots, q_k$).

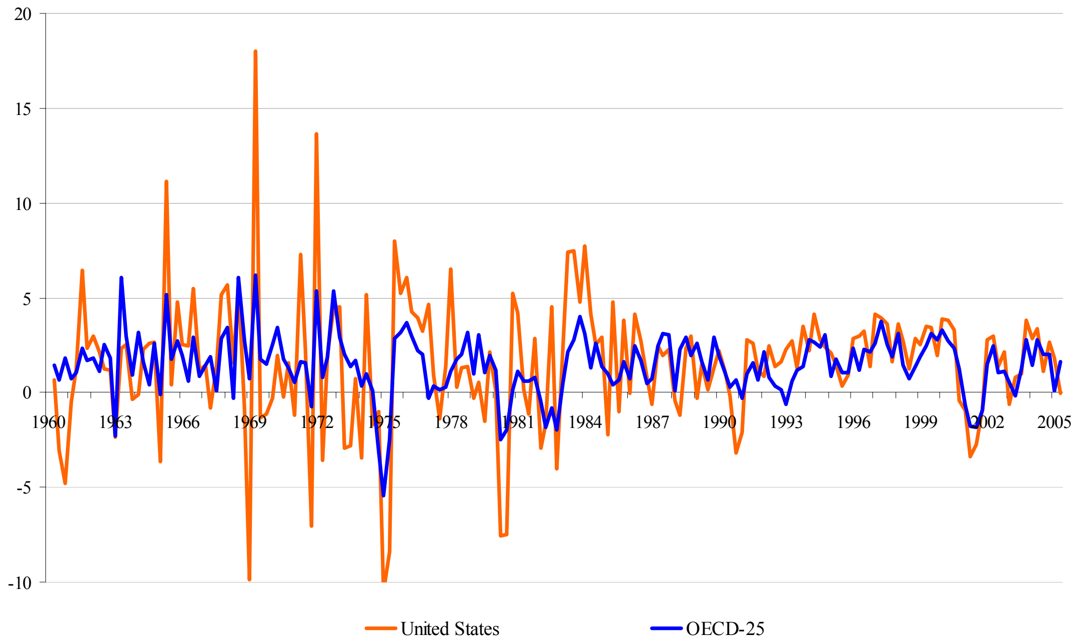
[43] For instance, for forecast 2, the Lagrange multiplier test indicates the possibility of conditional variance at the 10 per cent significance level. In addition, we conduct a Breusch-Godfrey test for serial correlation which confirms a moving average process in the residuals at the 10 per cent level of significance.

[44] GARCH (1,1) specifies the current volatility as a linear combination of lagged volatility and the lagged squared errors and, as such, models current volatility as an exponentially weighted moving average of past squared errors. Alternatively, we could have obtained the residuals from OLS and use standard Box-Jenkins techniques on the squared residuals to identify the order of the GARCH process. However, for our purposes a simple GARCH (1,1) model suffices, which is by far the most important case in practical applications. See

$$\sigma_t^2 = \alpha_0 + \alpha_1 u_{t-1}^2 + \phi_1 \sigma_{t-1}^2$$

The GARCH models are estimated using maximum likelihood.

Figure 8: Quarterly import growth rates, United States and OECD-25, 1960-2005



One of the problems with the ADL-based forecasts is the need to obtain future values of the additional predictors from elsewhere, for example from other studies or expert assessments. VAR models have the advantage that the other key variables can be forecasted as well. This is done in a multi-equation model, which makes the forecasts mutually consistent. However, it may be considered a drawback that, in a VAR(p) model, the number of lags is the same for all variables and that the number of coefficients to be estimated (number of variables times the lags, plus the intercept) needs to be kept sufficiently small in order not to lose too many degrees of freedom. More importantly, estimating too many coefficients increases the amount of estimation error entering the forecast, which can diminish its accuracy. Lag lengths can be determined using F-tests or the usual information criteria (for which the formulae are somewhat modified in the case of VARs), or else through the model itself using trial and error and examining the significance of the coefficients. With five lags, our VAR model for imports and domestic demand consists of the following two equations estimated by OLS:

$$M_t = \beta_{10} + \beta_{11} M_{t-1} + \ldots + \beta_{15} M_{t-5} + \gamma_{11} DD_{t-1} + \ldots + \gamma_{15} DD_{t-5} + u_{1t}$$
$$DD_t = \beta_{20} + \beta_{21} M_{t-1} + \ldots + \beta_{25} M_{t-5} + \gamma_{21} DD_{t-1} + \ldots + \gamma_{25} DD_{t-5} + u_{2t}$$

Since we are operating with quarterly data, but are called to make forecasts between half a year and one and a half years into the future, we need to obtain values for our endogenous variables beyond a one-step forecast. This applies to all our models. We pursue two methods to make such multistep forecasts. The multiperiod regression method involves using more distant lags, i.e. to make an h-period ahead forecast of a variable with p lags, the variable is regressed on its p lags with the most

---

Diebold (2004), in particular page 392 and footnote 7. The GARCH process makes h-step ahead forecasts of the conditional variance, which is time-varying, with the distant horizon forecast being just the unconditional variance, which is fixed, as is required for the stationarity of the time series to hold. See Engle (2001).

recent date of the regressor being t-h. In other words, for a simple AR(2) model of imports, a three quarters ahead forecast is computed by estimating the following multiperiod regression:

$$M_t = \delta_0 + \delta_1 M_{t-3} + \delta_2 M_{t-4} + u_t$$

Conversely, the iterated forecast strategy computes the one-period ahead forecast, which is then used in a second step to compute the two-period ahead forecast. For more distant horizons, this process is iterated until the target period is reached. In other words, the two- and three-quarters ahead iterated forecasts for an AR(5) model of imports would be:

$$\hat{M}_{t|t-2} = \hat{\beta}_0 + \hat{\beta}_1 \hat{M}_{t-1|t-2} + \hat{\beta}_2 M_{t-2} + \ldots + \hat{\beta}_5 M_{t-5}$$

$$\hat{M}_{t|t-3} = \hat{\beta}_0 + \hat{\beta}_1 \hat{M}_{t-1|t-3} + \hat{\beta}_2 \hat{M}_{t-2|t-3} + \hat{\beta}_3 M_{t-3} + \ldots + \hat{\beta}_5 M_{t-5}$$

($\hat{\beta}$s denote OLS estimates of the coefficients)

Each method has advantages and disadvantages depending on the specification of the model. If it is rather well specified, the iterated forecast method is preferable, since it uses coefficient estimators in a one-period ahead forecast that are more efficient than the estimators from the multiperiod regression.[45] Contrary to the multiperiod approach, the iterative forecast method is a *conditional* forecast since the values for more than one-step ahead forecasts are not known and have to be generated iteratively. Consequently, more than one-step ahead forecasts are conditional on forecasted values of used lags. Both methods can be used for multivariate forecasts as well (ADL and VAR).

In order to assess the relative quality of our models, we compare their respective root mean square forecast errors (RMSFEs), a standard evaluation tool. We estimate the RSMFE through pseudo out-of-sample forecasts. Since such "ex post forecasts" are computed using only data prior to the forecast date, the pseudo out-of-sample forecast errors reflect both the uncertainty associated with future values of the error term and the uncertainty inherent in the estimation of the regression coefficients. However, the RMSFE is not without problems. Two aspects are particularly noteworthy: First, since the seriousness of an error increases with the square of its size (e.g. an error of 2 per cent is treated as four times more important than a one per cent error), models with a few large errors may appear inferior to those with a larger number of small errors. Hence, a model that overall does well, but performs poorly in predicting the recession in 2001 (which, in particular in light of 9/11 was almost impossible to predict in its magnitude), may be considered worse than a model that is consistently somewhat off the point, including in 2001. Second, all ex post forecasts are given the same weight in the calculation of the RMSFE, although shorter time series are used the further back the ex post forecast is made. One therefore may wish to assign a higher weight to more recent forecasts, which make a fuller use of the available data set.

Another caveat regarding the RMSFE estimation from pseudo out-of-sample forecasts relates to the fact that for the additional predictors in the models, such as domestic demand, their actual values are used in the calculation. Historical time series exist for the actual observations of a variable, but are not normally kept for its forecast values that are the genuine input in the forecasting model. In order to illustrate this difference in an exemplary fashion, we assemble a time series of IMF GDP forecasts from 1991-2004 and, for one of the models, run individual regressions for each year with the data that would have been available at that point in time. From these ex post forecasts we compute the RMSFE

[45] Essentially, the multiperiod method throws away information that is already known by restricting the coefficients on lags 1 and 2 in our case to zero. This leads to a loss in efficiency. For a proof of the optimality of the iterated one-period ahead forecast see Harvey (1988). However, if the model is incorrectly specified and does not provide a good approximation to the correlations in the data, extrapolating these forecasts by iterating can lead to biased forecasts, and the multiperiod regression forecasts can be more accurate. See Stock and Watson (2003).

and compare it to the one calculated for that model in the usual manner. All of our results are discussed in detail below.

B. Forecast results

As is common practice in macroeconomic forecasting, we make point forecasts, i.e. provide a single number for the forecast period. In the annual fall forecasts (around September) growth rates of import volumes are forecasted for both the remainder of the year and the following year. Hence, the number of periods to be forecasted varies between two ("current year") and six quarters ("next year"). Figure 9 portrays the forecast time horizon for the best-performing model (forecast 2 in Table 1). The two quarterly forecasts in 2005 are combined with actual import growth rates in the first and second quarters to get the current year forecast. For 2006, all quarters must be forecasted to obtain the annual growth rate. We do not focus on the precise quarterly results, which, in the case of 2006, feature quite some variation, which may be difficult to explain. Rather, we are interested in annual developments, which we calculate from the quarterly data and which from 2005 to 2006 clearly show an upward tendency.

Figure 9: Quarterly growth rates of OECD-25 imports, 1991-2006, observed and forecasted (per cent)

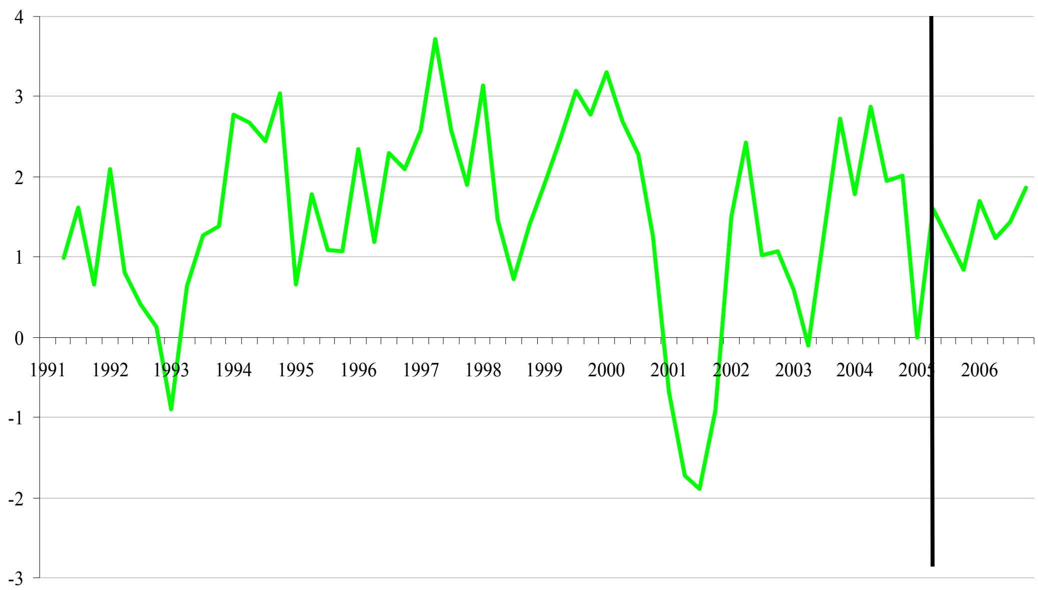


Table 1 shows ex post and current forecasts of annual import growth rates of goods and services for the group of OECD-25 countries as well as actual rates for the years in which they are known. We begin with a simple AR(2)[46] model following Box-Jenkins with a RMSFE for the current year of 0.84 and 3.60 for the next (forecast 1 in Table 1). This model has the advantage that no future values of exogenous variables are needed as inputs. The ADL model with domestic demand as an additional predictor using the iterative method (forecast 2 in Table 1) already performs considerably better; in fact, it is our best forecasting model for both the current and following year with RMSFEs of 0.49 and 2.72, respectively.[47] However, in order to calculate RMSFEs of the ADL models actual values of domestic demand growth are employed. This is equivalent to assuming that domestic demand predictions as an input in the model have proved to be correct. In reality, this is not the case and

[46] RMSFEs of the AR(5) model are virtually the same.

[47] The ADL coefficients estimated by OLS and by maximum likelihood with the GARCH model differ. We only exhibit the GARCH results in the paper, since the RMSFEs of the latter consistently outperform the OLS-based forecasts.

RMSFEs may therefore be higher. To illustrate this divergence we re-run the ADL model with domestic demand as an additional predictor using the actual IMF forecasts made in the years 1991 to 2004 (forecast 3 in Table 1). The results for the current year are still considerably better than the outcomes obtained from the AR(2) model.

Table 1: Annual import growth rates, OECD-25, 1995-2006, observed and forecasted (per cent)

		1995	1996	1997	1998	1999	2000	2001	2002	2003	2004	2005	2006	RMSFE
observed import growth		6.94	6.72	10.17	8.55	7.80	10.99	-0.43	2.04	3.78	8.11	-	-	-
forecast 1: IMP: 2 lags, Box-Jenkins	current year	6.85	5.75	9.22	8.60	6.33	10.38	1.30	2.05	3.05	7.42	5.10	-	0.84
	next year	5.20	4.77	4.60	5.79	4.92	5.29	5.60	3.27	5.23	3.92	5.43	4.74	3.60
forecast 2: IMP: 5 lags, DD: contemp + 6 lags, GARCH	current year	7.13	6.17	10.05	9.30	7.31	10.43	0.28	2.81	3.63	8.06	5.50	-	0.49
	next year	2.70	5.41	5.68	7.68	6.78	7.92	3.32	1.26	6.35	5.98	5.94	6.67	2.72
forecast 3: IMP: 5 lags, DD (IMF forecasts): contemp + 6 lags, GARCH	current year	7.68	5.63	9.65	8.40	6.98	11.10	0.61	2.82	3.19	8.21	4.97	-	0.64
	next year	5.13	5.37	3.92	5.84	3.48	6.39	6.97	2.27	6.84	4.48	6.42	4.54	3.78

Notes: Imports refer to import volumes of goods and services. For forecasting, quarterly data are used starting from 1960. Cells in the row "current year" contain the forecast for the year in the column header, i.e. forecasts of annual import growth when the first and second quarters of the year are already known. Cells in the row "next year" refer to forecasts for the year in the column header made in the *previous* year, i.e. the "next year" forecast in the 1996 column is the six quarter forecast of annual import growth in 1996 made in 1995 on the basis of data up to the second quarter of 1995. RMSFEs are calculated on the basis of the years 1990-2004.

Forecast 4 in Table 2 includes both domestic demand and the real oil price.[48] It does almost as well as forecast 2 (Table 1). Both models provide quite accurate forecasts for several years, especially 1998 and 1999. The two-digit import growth rates (10 and 11 per cent) in 1997 and 2000 are foreseen in the current year, but somewhat underestimated by both models in the preceding year. Forecast 2 still captures a strongly optimistic outlook at around 7 to 8 per cent. The years 2001-2003 have posed problems in forecasting owing to the dent in 2001 as well as the relatively swift recovery of global trade thereafter. For 2001, one of the few years (besides 1975 and 1982) when actual import growth rates were slightly negative in the last 45 years, forecast 4 predicts a rate of nearly zero even six quarters ahead. Forecast 2 largely misses the actual value in 2001, but quite accurately foresees import growth of about 2 per cent in 2002. Here, forecast 4 remains a bit lower, while still indicating a turnaround. The unusual developments in these years carry through to the 2003 forecasts, which tend to be overoptimistic.

[48] We have also run an ADL model with the real oil price as the only additional predictor. However, its performance, especially in making current year predictions, is clearly inferior to alternative specifications and its results are not presented here. We have also substituted relative import prices, which are available for the United States from the IMF's IFS, for the oil price. The United States import price index is deflated with the consumer price index (CPI) in order to obtain real index numbers (year 2000 = 100). In the country/regional equations, we have introduced real exchange rates in order to take account of changes in relative prices. We have employed the real effective exchange rate, which measures the volume of imports that can be afforded for a given volume of exports already weighed by principal trading partners. None of these changes have resulted in any improvements in forecast performance, and the respective models are therefore not further discussed in the text.

In an attempt to reduce the weight of extreme historical observations of imports, we include various time dummies in the ADL model with both growth in the volume of domestic demand and the real oil price as additional predictors (forecast 5 in Table 2). The dummies take a value of one for each quarter in which the import growth rate deviates positively or negatively by at least two standard deviations from the mean. Inclusion of dummies reduces the variation in the forecasts with no noteworthy improvement of the RMSFE. Some years that are rather accurately forecasted by the model without dummies are underestimated by the dummy version. The RMSFE remains about the same, as errors are reduced in the years that are difficult to forecast.

A closer look at the results of the VAR (forecast 6 in Table 2) reveals RMSFEs similar to forecast 3 (Table 1).[49] While the VAR avoids possible imprecision in the forecasted values of the exogenous variable, six quarter ahead import growth forecasts do not show much variation, oscillating between 5.5 and 6.5 per cent. This problem is similar to the AR(2) model (forecast 1 in Table 1) with projected values of between 4 and 5.5 per cent. These forecasts essentially revolve around the long-term trend resulting in rather conservative predictions and may therefore consistently be somewhat off the point but rarely completely wrong. For instance, the VAR model considerably underestimates import growth in the years 1996 and 1998, whereas the ADL model (forecast 3 in Table 1) produces better results (over- or underestimating actual values) in those years, while suffering from higher deviations in others.

Table 2: Annual import growth rates, OECD-25, 1995-2006, observed and forecasted (per cent)

		1995	1996	1997	1998	1999	2000	2001	2002	2003	2004	2005	2006	RMSFE
observed import growth		6.94	6.72	10.17	8.55	7.80	10.99	-0.43	2.04	3.78	8.11	-	-	-
forecast 4: IMP: 5 lags, DD: contemp + 6 lags, OIL: contemp + 7 lags, GARCH	current year	6.51	6.26	9.78	9.45	7.58	9.67	0.03	2.83	3.28	8.19	4.82	-	0.69
	next year	2.87	4.46	5.13	7.18	7.41	7.72	0.58	0.47	6.48	5.81	5.76	4.01	2.85
forecast 5: IMP: 5 lags, DD: contemp + 6 lags, OIL: contemp + 7 lags, Time Dummies, GARCH	current year	6.52	6.27	9.80	9.44	7.57	9.71	-0.03	2.81	3.31	8.20	4.85	-	0.69
	next year	2.87	4.50	5.19	7.21	7.35	7.79	0.72	0.42	6.43	5.93	5.79	4.23	2.84
forecast 6: IMP: 5 lags, DD: 5 lags, VAR	current year	6.30	6.12	9.25	8.93	6.94	10.77	0.66	2.72	3.00	7.80	5.22	-	0.75
	next year	2.13	1.41	5.69	3.58	5.18	6.08	5.78	3.43	4.70	4.39	4.94	5.58	4.19

Notes: Imports refer to import volumes of goods and services. For forecasting, quarterly data are used starting from 1960. Cells in the row "current year" contain the forecast for the year in the column header, i.e. forecasts of annual import growth when the first and second quarters of the year are already known. Cells in the row "next year" refer to forecasts for the year in the column header made in the *previous* year, i.e. the "next year" forecast in the 1996 column is the six quarter forecast of annual import growth in 1996 made in 1995 on the basis of data up to the second quarter of 1995. RMSFEs are calculated on the basis of the years 1990-2004.

All forecasts presented so far are made using the iterative method. The inferiority of the multiperiod method in terms of RMSFEs for both the current and following year is exemplified for the ADL model with domestic demand as an additional predictor. Clearly, for identical models, the multiperiod forecasting method results in much higher RMSFEs (0.79 of forecast 7 in Table 3 versus 0.49 of forecast 2 in Table 1 for the current year and 3.59 versus 2.72 for the next).

[49] The changes in results are negligible when the model is estimated as a VECM.

Table 3: Annual import growth rates, OECD-25, 1995-2006, observed and forecasted (per cent)

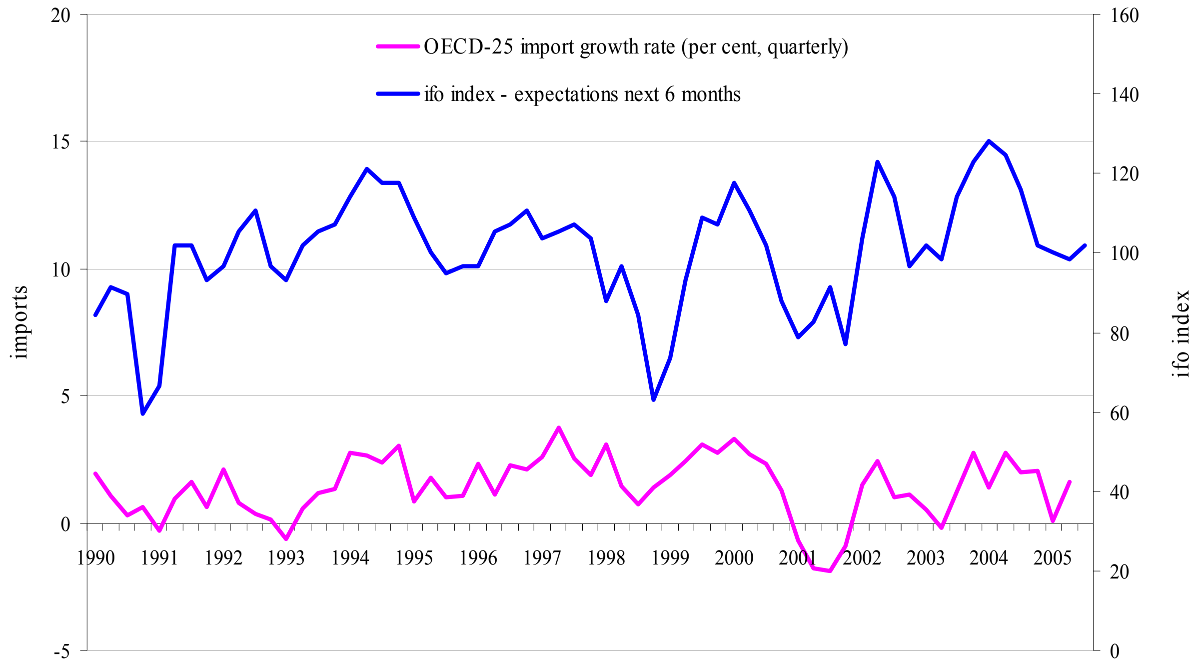
multiperiod forecast

		1995	1996	1997	1998	1999	2000	2001	2002	2003	2004	2005	2006	RMSFE
observed import growth		6.94	6.72	10.17	8.55	7.80	10.99	-0.43	2.04	3.78	8.11	-	-	-
forecast 7: IMP: 5 lags (3.-7. lag), DD: 6 lags (3.-8. lag), GARCH	current year	6.38	6.06	9.41	8.86	6.82	10.75	1.39	2.94	3.31	8.04	5.40	-	0.79
	next year	4.33	4.42	8.61	7.20	7.02	10.08	7.94	5.14	10.59	7.61	9.22	9.82	3.59

Notes: Imports refer to import volumes of goods and services. For forecasting, quarterly data are used starting from 1960. Cells in the row "current year" contain the forecast for the year in the column header, i.e. forecasts of annual import growth when the first and second quarters of the year are already known. Cells in the row "next year" refer to forecasts for the year in the column header made in the *previous* year, i.e. the "next year" forecast in the 1996 column is the six quarter forecast of annual import growth in 1996 made in 1995 on the basis of data up to the second quarter of 1995. RMSFEs are calculated on the basis of the years 1990-2004.

In order to maintain an ADL set-up while circumventing the problem of having to predict exogenous variables, we use the ifo World Economic Climate index as an additional predictor of imports. The graph in Figure 10 seems to suggest that as a leading indicator the ifo index could have some "explanatory" power for import growth. At the end of August 2005, the index number for the third quarter was published. This allows us to lead the indicator by two quarters and make a forecast for the first quarter 2006 using the multiperiod method. Unfortunately, in part also due to the inferior multiperiod forecasting method, the performance of the ifo index in forecasting import growth has been rather low both for the current and next year, when first quarter results are extrapolated.

Figure 10: Quarterly growth rates of OECD-25 imports and ifo World Economic Climate Index, 1990-2005



In Table 4, the best model of each individual country/region forecasting exercise is presented.[50] The RMSFEs of the forecasts for the United States and the EMU are of the same order as the ones for the OECD-25, whereas the six quarters ahead forecasts of Germany in terms of RMSFE is clearly worse and the one of Japan is practically unusable. It is noteworthy that all of the models are different. With the exception of Germany (see Table 5), the ADL model with domestic demand as an additional predictor performs best (see Table 4). However, due to nonstationarity for some of the time series even after differencing, second differences have to be used, including for imports in the case of Germany and Japan. The lag structures according to the usual information criteria also differ for otherwise similar models. For Germany, the ifo Business Climate Index, which may be considered more sophisticated than the ifo World Economic Climate Index used in the OECD-25 estimations, performs very well as an additional predictor.[51]

Several other specifications of the individual models are not presented in the table. For all countries and regions, we have included the real oil price and real effective exchange rates, whenever these variables appear to "Granger-cause" imports. However, the resulting RMSFEs are rather high. The simple AR(p) models do not improve the forecasts in any of the cases examined either. Since, for the United States, import data are available separately for goods and services, we re-run all of the specifications for goods only. Growth rates are generally more pronounced than for goods and services taken together. For the years 2005 and 2006, the model presented in Table 4 predicts 6.5 and 6 per cent growth in imports of goods and services versus, respectively, about 7 and 6.25 per cent for goods only. For Germany, we are able to select from a range of similar leading economic indicators. Using the ZEW index instead of the ifo Business Climate Index, however, do not improve our model.

[50] As stated in section II, due to data availability and the specific econometric test results, we run different sets of model specifications (explanatory variables, lags) for each country/region.

[51] Surveys for the Ifo World Economic Climate Index are conducted quarterly in numerous countries. The focus is on business cycle developments and other economic factors in the experts' home countries. The October 2005 survey received responses from 1,100 experts in 91 countries. Conversely, the Ifo Business Climate Index for Germany is based on a much larger number of responses (around 7,000), and surveys are conducted monthly. Firms give their assessments of the current business situation and their expectations for the next six months. There is also a sectoral disaggregation of firms in manufacturing, construction, wholesaling and retailing. Replies by firms are then weighted according to the importance of the industry and aggregated. For more information on how the Ifo Business Climate Index is calculated see http://www.cesifo-group.de/portal/page?_pageid=36,103089&_dad=portal&_schema=PORTAL&item_link=erlaeut_gk.htm, visited on 1 December 2005.

Table 4: Annual import growth rates in selected countries, 1995-2006, observed and forecasted (per cent)

		1995	1996	1997	1998	1999	2000	2001	2002	2003	2004	2005	2006	RMSFE
USA														
observed import growth		7.73	8.34	12.73	11.00	10.85	12.34	-2.75	3.34	4.48	10.19	-	-	-
1960 - , IMP: 2 lags, DD: contemp + 3 lags, GARCH	current year	8.10	8.07	12.46	12.50	10.58	11.68	-0.72	2.61	5.55	10.69	6.51	-	0.93
	next year	5.17	6.87	8.93	11.67	12.27	8.99	3.94	3.08	6.38	9.42	8.71	6.00	2.98
EMU														
observed import growth		7.33	3.59	8.76	9.64	7.27	10.40	1.98	0.41	2.89	6.06	-	-	-
1970 - , IMP: 3 lags, DD: contemp + 4 lags, GARCH	current year	7.71	2.98	8.58	10.42	6.82	9.41	2.84	1.19	2.49	5.65	3.33	-	0.58
	next year	7.91	4.66	4.04	9.80	7.46	6.25	2.24	0.39	6.67	4.11	5.60	5.36	2.27
Japan														
observed import growth		11.76	12.37	1.03	-6.85	3.20	8.85	0.16	1.91	4.89	9.63	-	-	-
1980 - , IMP: 9 lags, DD: contemp + 3 lags, GARCH	current year	10.61	14.38	2.91	-6.58	2.31	8.23	1.26	3.26	3.48	8.61	5.61	-	1.20
	next year	5.73	11.68	11.11	8.73	-4.73	4.93	5.17	-0.60	11.84	0.92	7.77	4.92	7.14

Notes: Imports refer to import volumes of goods and services. For forecasting, quarterly data are used starting from 1960. Cells in the row "current year" contain the forecast for the year in the column header, i.e. forecasts of annual import growth when the first and second quarters of the year are already known. Cells in the row "next year" refer to forecasts for the year in the column header made in the *previous* year, i.e. the "next year" forecast in the 1996 column is the six quarter forecast of annual import growth in 1996 made in 1995 on the basis of data up to the second quarter of 1995. RMSFEs are calculated on the basis of the years 1990-2004.

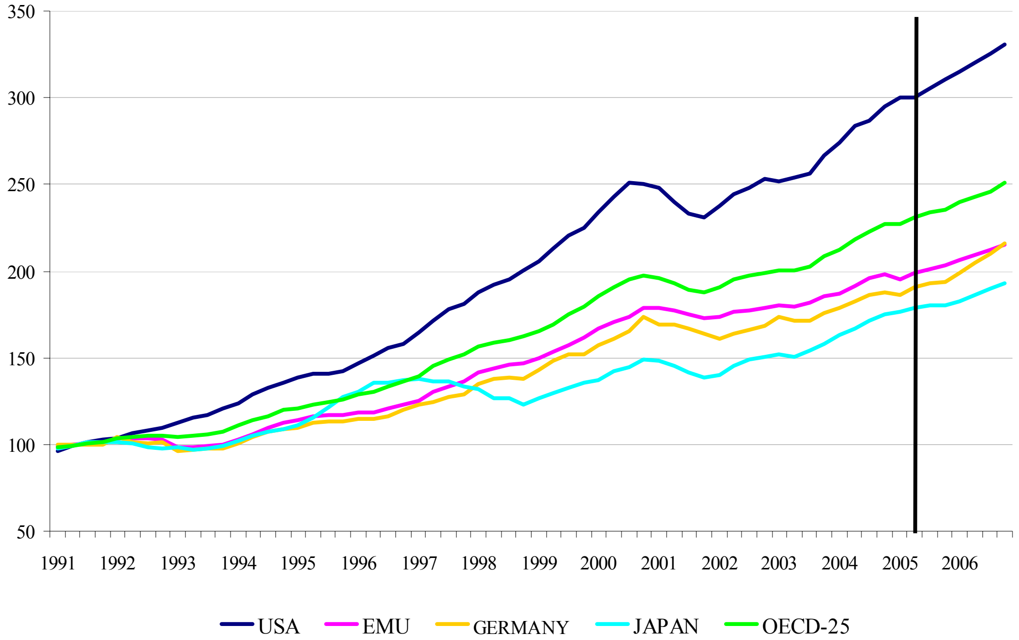
Table 5: Annual import growth rates in Germany, 2000-2005, observed and forecasted (per cent)

		2000	2001	2002	2003	2004	2005	RMSFE
Germany								
observed import growth		10.15	1.53	-1.33	4.87	5.93	-	-
1991 - , IMP: 3 lags, ifo business climate index: contemp + 3 lags, GARCH	current year	8.82	3.00	-0.50	4.66	5.94	3.86	0.96
	next year	-	4.11	5.51	7.18	5.35	2.49	3.84

Notes: Imports refer to import volumes of goods and services. Due to German re-unification in 1990, consistent data is only available for a relatively short period of time. Since an adequate sample size of at least 10 years is required for a time series estimation, forecasts can be made only for the most recent years. Cells in the row "current year" contain the forecast for the year in the column header, i.e. forecast of annual import growth when the first and second quarters of the year are already known. Cells in the row "next year" refer to forecasts for the year in the column header made in the *previous* year, i.e. the "next year" forecast in the 2001 column is the six quarter forecast of annual import growth in 2001 made in 2000 on the basis of data up to the second quarter of 2001.

Figure 11 collates the development of imports of both the OECD-25 and the individual countries/regions examined. Since 1995, annual United States import growth rates have been about 1 to 2 per cent higher (or lower in 2001) than for the OECD-25 as a whole. Roughly in line with this tendency, we predict 6.5 per cent import growth in the United States for 2005 (as opposed to about 5 per cent in the OECD-25) and 6 per cent in 2006 (5.5 per cent in the OECD-25). Import growth for the EMU as a group usually tracks quite closely developments in the OECD-25 group of countries, with a tendency to be slightly lower in a number of years. This tendency is also reflected in our forecasts of 3.3 per cent for 2005 and 5.4 per cent for 2006. Germany and Japan feature a lot more variation in growth rates over time resulting in more irregular developments of import levels as well. The effects of double-digit growth rates of imports in Japan in 1995 and 1996 were wiped out again by 1998. Apart from the second half of 1999, when it underperformed, and 2000, when it overperformed, Germany has clearly dominated import developments in the EMU. For both countries, we only predict the current year. The six quarter forecasts for Japan have a too elevated RMSFE, while data limitations for the ifo index prevent us from predicting more than the first quarter of the following year for Germany. Extrapolating this quarterly forecast over the whole year also results in an unacceptably high RMSFE. The 2005 forecasts are 3.9 per cent for Germany (about half a per cent higher than for the EMU) and 5.6 per cent for Japan.

Figure 11: Import volumes of goods and services, selected countries, 1991-2006, observed and forecasted (1991 = 100)

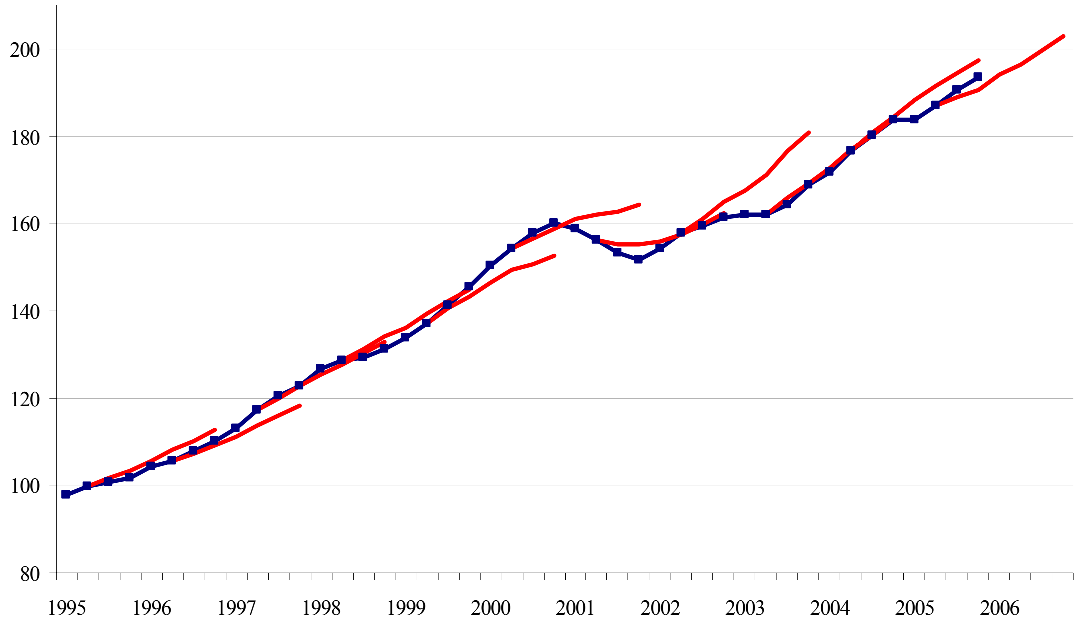


Which model should be followed in supporting WTO forecasts? In our view, in order for short-term forecasts to be useful, excessive weight should not be given to the long-term trend. We therefore have a preference for the ADL models, which have the potential to perform better than VAR or AR(p) if predictions of exogenous variables can be improved and if alternative specifications are run for comparison, which may include time dummies or other exogenous variables, such as leading indicators. In our models, current year forecasts of import growth rates are by and large very good with RMSFEs inferior to one. In the ADL model with domestic demand, for example, only five of fifteen ex post forecasts for the current year would have under- (one) or over-estimated (four) actual import growth by one per cent.[52] The "next year" forecasts with RMSFEs of around 2.5 to 3.5 seem

[52] The same is the case for the ADL model that includes both domestic demand and the oil price with two under- and three over-estimations of actual import growth.

acceptable given the long forecast horizon of six time periods. In any event, such results should be interpreted in the light of expert knowledge, for instance on expected policy developments in the year to come. Figure 12 portrays the six quarter ahead forecasts made in the third quarter of each year using the ADL model with domestic demand as an additional predictor. Ex post, it can be seen that our forecasts, except for 2001, would have captured upward or downward developments in imports quite well, in particular also the turnaround in 2002, albeit overestimating it to a certain extent. The chart also includes our forecast for the rest of 2005 and 2006. According to this model, we foresee a steady upward movement of imports by OECD-25 countries in the near future with a growth rate of 5 per cent in 2005 that accelerates slightly to 5.5 per cent in 2006.

Figure 12: OECD-25 import volumes of goods and services – ex post analysis, 1995-2006 (1995 = 100)



It is noteworthy that our forecast for 2006 is just slightly higher than the long term trend of import growth of 4.5 to 5.5 per cent.[53] Yet, for most of time in the past 15 years, our annual forecasts diverge substantially from that trend. Such behaviour is closer to reality, which features quite some variation in annual trade growth rates from one year to the next. This distinguishes our forecasts from, for instance, the predictions made by the IMF in past World Economic Outlooks, which usually lie between 5 and 6 per cent (see Appendix Table 1). In addition, judging from conventional error measures[54] to assess relative model performance, our models outperform the IMF predictions of trade growth in both the two and six quarter ahead forecasts on all accounts (see Table 6).[55]

Even a naïve approach obtained from projecting the deterministic trend into the future (see Appendix Figure 1) does surprisingly well by some of these statistics as compared to the IMF.[56] However, its apparent performance, at least in part, may be explained by the inherent deficiencies in the scale-

[53] The long-term trend is calculated as the average growth rate we obtain from regressing the first differences of logarithmized imports on a deterministic linear time trend.

[54] Besides the RMSFE, these are the mean absolute error (MAE), the mean absolute percentage error (MAPE) and Theil's U coefficient. The latter is a measure of the RMSFE in relative terms, standardized to values between 0 and 1, i.e. Theil's U statistic is scale-insensitive.

[55] Naturally, forecast error statistics increase for all models as the forecast horizon increases from two to six quarters.

[56] See, for instance, Ye et al. (2005) who follow a similar approach for forecast evaluation by comparing their model to two alternative models, a naïve as well as an established one.

sensitive evaluation statistics. As was said before, the RMSFE penalizes models with a few large errors more heavily than those with many small errors. Similarly, according to the MAPE the same absolute deviation is considered less serious if it occurs for a year with relatively high actual import growth.[57] The naïve approach results in annual predictions of import growth rates falling within a narrow range of values broadly in line with the long-term trend. Naturally, the usefulness of these results for short-term decision-making is limited. It seems preferable that short-term forecasts of import growth respond to sudden changes in economic conditions and use the information contained in the historical patterns of relevant variables in order to anticipate larger swings. With our forecasts being more volatile (which may increase the risk of committing some larger errors), they are also bound to give better directional signals. We are therefore confident that our models are an appropriate tool for short-term trade forecasting, as confirmed by the evaluation statistics shown in Table 6.[58]

Table 6: Forecast evaluation statistics

forecast current year			
	naïve	IMF	*forecast 2*
RMSFE	0.85	1.62	0.49
MAE	0.65	1.45	0.42
MAPE	0.44	0.40	0.20
Theil's U	0.07	0.13	0.04
forecast next year			
	naïve	IMF	*forecast 2*
RMSFE	3.76	4.10	2.72
MAE	3.35	3.34	2.32
MAPE	1.77	1.12	0.97
Theil's U	0.32	0.37	0.23

VI. CONCLUSIONS

In this paper, we have developed a time series approach to forecasting international trade. The models perform well, especially in making two quarter ahead forecasts. Import growth rates for the following year, most of the time, are also foreseen quite accurately. The major strength of the six quarter ahead forecasts lies in their tracking of turning points in trade developments. In relative terms, standard evaluation statistics confirm that the time series approaches presented in this paper perform better than both naïve forecasts and more elaborate IMF predictions. Both the IMF and OECD employ complex structural models combined with export opinion. The nature of the forecasting process is not completely known. By contrast, our models are both parsimonious and fully documented. We find that in light of their comparatively strong performance, "mechanical" time series forecasts have a lot to offer compared to more information- and resource intensive approaches. Our modelling strategy may also be useful in forecasting trade as an exogenous variable in large macroeconometric models.

[57] Hence, a few large deviations of the forecast from a relatively small base value of actual import growth may result in a larger MAPE than persistently wrong forecasts that deviate slightly from the actual values, especially if base rates are high. This effect is particularly evident in the IMF's "next year" forecast evaluation, where large average absolute deviations, as measured by the MAE, translate into a favourable MAPE score when deflated with actual annual growth rates.

[58] Our model even has smaller (and at least not larger) forecast errors compared to the naïve and IMF approaches, when the deficiencies in the forecasts of the exogenous domestic demand variables are taken into account, as in forecast 3 (see Table 1).

A number of areas merit further exploration for possible model improvement. On the data side, if the aim is to make forecasts of global trade growth, the OECD-25 number can only constitute a lower bound estimate, as trade growth in the developing world is usually much higher.[59] It would be desirable to extend the group of countries examined in our models to include other major traders, in particular China. Much will depend on the availability of quality data, especially at quarterly intervals, for a number of countries. If the regional/country coverage is sufficiently large, the individual forecasts could be aggregated and compared to the global forecast. It may also prove useful to separate goods from services data. With the latter growing in importance, our model set-up, which is geared towards merchandise trade, may need refinement. Future work on the econometric analysis could be targeted at improvements in the integration of expert judgement. This could be accomplished by mapping the assessment of the forecast variable by an independent expert/decision-maker possessing non-sample information into a guess on the parameters of the preferred econometric model. We could also make interval or density forecasts instead of point forecasts. Fan charts can be used to display confidence bands that widen over time and may not necessarily be symmetrical around the point forecast. This would allow us to reflect increasing uncertainty as well as the balance of risks that may be tilted towards the up- or downside.

[59] With expected growth rates for 2006 of about 9 per cent in the newly industrialized Asian economies and 12 per cent in the group of other emerging markets and developing countries, world import volumes of goods and services are bound to grow by about 7.5 per cent, i.e. 2 per cent more than our OECD-25 forecasts of 5.5 per cent. See IMF, World Economic Outlook Database, September 2005, available at http://www.imf.org/external/pubs/ft/weo/2005/02/data/index.htm, last visited on 25 February 2006.

References

Abeysinghe, T. (1994) 'Forecasting performance of seasonal dummy models relative to some alternatives', *Economics Letters* 44: 365-370.

Abeysinghe, T. (1998) 'Forecasting Singapore's quarterly GDP with monthly external trade', *International Journal of Forecasting* 14: 505-513.

Armstrong, J. S. and Collopy, F. (1998) 'Integration of Statistical Methods and Judgement for Time Series Forecasting: Principles from Empirical Research', in: Wright, G. and Goodwin, P. (eds.) (1998) *Forecasting with Judgement*, Chichester: John Wiley & Sons Ltd.: 269-293.

Batchelor, R. (2000) 'The IMF and OECD *versus* Consensus Forecasts', mimeo, London: City University Business School.

Blix, M., Wadefjord, J., Wienecke, U. and Adahl, M (2001) 'How good is the forecasting performance of major institutions?', *Sveriges Riksbank Economic Review* 3/2001: 38-68.

Consensus Economics (2005) 'What are consensus forecasts?', mimeo, London: Consensus Economics Inc.

Diebold, F. X. (1997) 'The Past, Present, and Future of Macroeconomic Forecasting', Research Department Working Paper No. 97-20, Philadelphia, Penn.: Federal Reserve Bank of Philadelphia.

Diebold, F. X. (2004) *Elements of Forecasting*, 3rd edition, Mason, Ohio: Thomson, South-Western.

Engle, R. (2001) 'GARCH 101: The Use of ARCH/GARCH Models in Applied Econometrics', *Journal of Economic Perspectives* 15, 4: 157-168.

Fair, R. C. (2004) *Estimating how the macroeconomy works*, Cambridge, Mass.: Harvard University Press.

Harvey, A. C. (1988) *Time series models*, Oxford: Philip Allan Publishers.

Hendry, D. F. and Clements, M. P. (2003) 'Economic forecasting: some lessons from recent research', *Economic Modelling* 20: 301-329.

Hummels, D., Ishii, J. and Yi, K. M. (2001) 'The nature and growth of vertical specialization in world trade', *Journal of International Economics* 54, 1: 75-96.

International Monetary Fund (IMF) (2003) 'MULTIMOD Mark III Econometric Model', mimeo, Washington: IMF.

International Monetary Fund (IMF) (2005) 'About the World Economic Outlook Database', mimeo, Washington: IMF.

Lenain, P. (2002) 'What is the track record of OECD economic projections?', Organisation for Economic Co-operation and Development (OECD), mimeo, Paris: OECD.

Manganelli, S. (2006) 'A New Theory of Forecasting', European Central Bank (ECB) Working Paper No. 584, Frankfurt am Main: ECB.

Organisation for Economic Co-operation and Development (OECD) (2004) 'Sources & Methods of the OECD Economic Outlook', Paris: OECD.

Pain, N., Mourougane, A., Sédillot, F. and Le Fouler, L. (2005) 'The new OECD international trade model', Economics Department Working Paper No. 440, Organisation for Economic Co-operation and Development (OECD) document ECO/WKP(2005)27, Paris: OECD.

Pindyck, R. S. and Rubinfeld, D. L. (1998) *Econometric Models and Economic Forecasts*, 4th edition, Boston, MA: McGraw-Hill.

Rae, D. and Turner, D. (2001) 'A small global forecasting model', Economics Department Working Paper No. 286, Organisation for Economic Co-operation and Development (OECD) document ECO/WKP(2001)12, Paris: OECD.

Stock, J. H. and Watson, M. W. (2003) *Introduction to Econometrics*, Boston, MA: Pearson Education, Inc.

United Nations (UN) (2006) *World Economic Situation and Prospects 2006*, New York: UN.

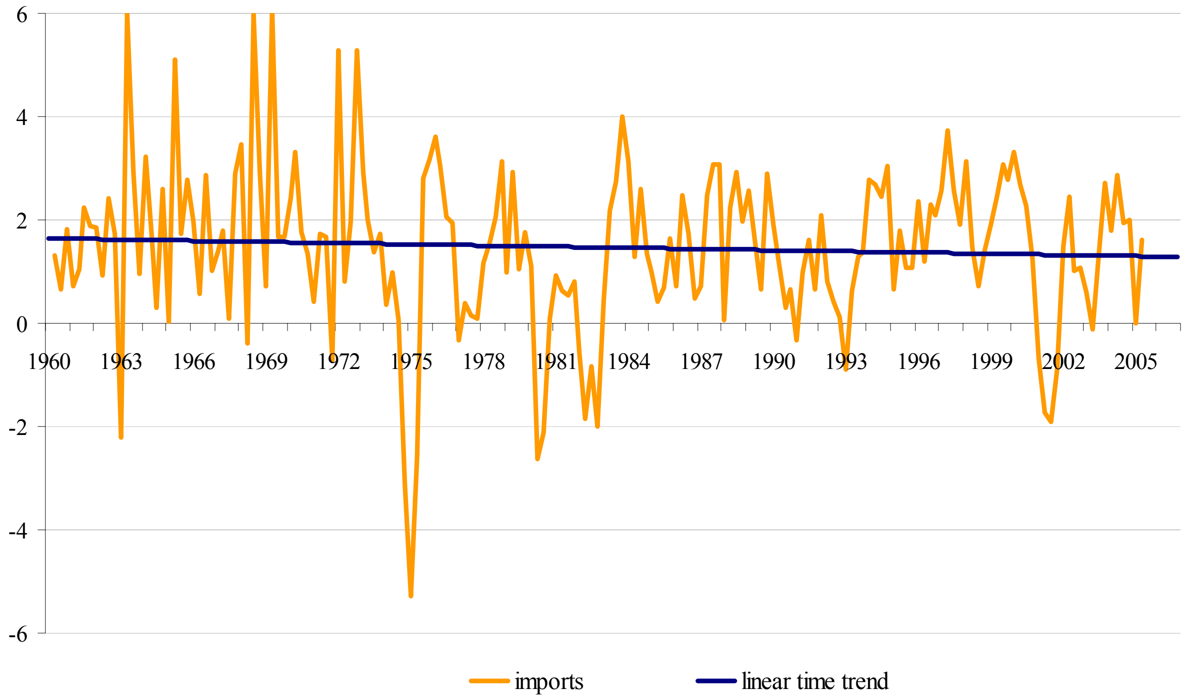
World Trade Organization (WTO) (2006) "Trade picks up in mid-2005, but 2006 picture is uncertain", Press Release PRESS/437, 11 April 2006, Geneva: WTO.

Ye, M., Zyren, J. and Shore, J. (2005) 'A monthly crude oil spot price forecasting model using relative inventories', *International Journal of Forecasting* 21: 491-501.

Yi, K. M. (2003) 'Can vertical specialization explain the growth of world trade', *Journal of Political Econom*y 111, 1: 52-102.

Appendix

Appendix Figure 1:
A "deterministic" trend in quarterly import growth rates, OECD-25, 1960-2005 (per cent)



Appendix Table 1:
IMF fall forecasts of annual import growth rates and observed values, advanced economies, 1995-2006, (per cent)

	1995	1996	1997	1998	1999	2000	2001	2002	2003	2004	2005	2006	RMSFE
observed import growth	9.0	6.4	9.3	5.9	8.0	11.7	-1.0	2.6	4.1	8.8	-	-	-
forecast current year	7.1	5.3	7.1	4.5	5.9	10.3	1.7	1.7	2.8	7.6	5.4	-	1.62
forecast next year	4.8	5.5	5.5	6.4	4.7	5.9	7.9	4.7	6.2	4.8	5.6	5.6	4.10

Notes: Imports refer to import volumes of goods and services. The numbers show forecasts of the IMF reported in the respective annual issues of the *World Economic Outlook*. Cells in the row "current year" contain the forecast for the year in the column header, i.e. forecasts for annual import growth when the first half of the year is already known. Cells in the row "next year" refer to forecasts for the year in the column header made in the *previous* year. RMSFEs are calculated on the basis of the years 1990-2004.