

FORECASTING INTERNATIONAL TRADE: A TIME SERIES APPROACH

Alexander Keck*

Economic Research and Statistics Division, World Trade Organization (WTO),
Rue de Lausanne 154, 1211 Geneva 21, Switzerland
email: alexander.keck@wto.org

Alexander Raubold

Trade Policy Unit, German Federal Ministry of Economics and Technology (BMWi),
Scharnhorststr. 34-37, 10115 Berlin, Germany
email: raubold@bmwi-bund.de

Alessandro Truppia

Economic Research and Statistics Division, World Trade Organization (WTO),
Rue de Lausanne 154, 1211 Geneva 21, Switzerland
email: alessandro.truppia@wto.org

* 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. None of the forecasts presented in this paper represents any official WTO forecast.

The authors would like to thank Hildegunn Kyvik Nordås, Patrick Low, George Rapsomanikis, Robert Teh, Stephen Tokarick, participants at the Swiss Institute for Business Cycle Research Seminar at the ETH Zurich and two anonymous referees for helpful comments on earlier drafts of this paper. All remaining errors and omissions are the fault of the authors.

**FORECASTING INTERNATIONAL TRADE:
A TIME SERIES APPROACH**

Abstract

This paper develops a time series model to forecast the growth in imports by major advanced economies in the current and following year (two to six quarters ahead). Both pure time series analysis and structural approaches that include additional predictors based on economic theory are used. Our results compare favourably to other trade forecasts, as measured by standard evaluation statistics, and can serve as a benchmark for more complex macroeconomic models.

JEL classification: F17, C53, C32, C22

Keywords: Forecasting, time series, international trade

1. INTRODUCTION

Short-term forecasting of key economic variables, such as GDP and inflation, has a long tradition, as many important business and investment decisions are based on forecasts for the outlook of the economy (Blix et al., 2001). In a "globalized" world characterized over the last several decades by increasing interdependency among nations involving intensive political, social and economic interaction the need for predictions of key economic parameters both at home and abroad has increased. With the important role that trade has played in this wider process of globalization (WTO, 2008), it comes as no surprise that forecasts of exports and imports of major trading nations and regional blocs have become a central feature of providers of economic forecasts. Especially, in the current climate of economic crisis, many of the open economies that have enjoyed decades of high economic growth and have been badly affected by the global downturn are expected to be better placed to stage a faster and stronger recovery when the expansion of trade resumes (WTO, 2009).

At the international level, the International Monetary Fund (IMF) and Organisation for Economic Co-operation and Development (OECD) are leading providers of macroeconomic data and forecasts, including on import and export performance.¹ Two aspects of the way in which such forecasts are produced call for particular attention and have largely motivated the approach that we develop in this paper: First, the OECD and IMF use large structural macroeconomic models of the world economy (called INTERLINK and MULTIMOD respectively). The goal is to establish consistency between the different economic variables and countries through international financial and trade linkages (OECD, 2004; IMF, 2005). The primary purpose of the predictions made on the basis of these models is not to produce mere short-term forecasts of individual variables, but to identify structural economic problems

¹ 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. For an extensive analysis of forecasting by these two organizations see Batchelor (2001). Of course, plenty of private forecasters also exist; however, official sources, such as the IMF and OECD, provide a backdrop for most forecasts, at the very least as far as the reliability, consistency and completeness of the data are concerned.

and possible policy responses by simulating a range of different scenarios (Lenain, 2002; IMF, 2005).² This implies that forecasts of individual variables are constrained by the modelling structure and the need for consistency imposed by it. As a consequence, if the principal interest was to obtain an as accurate forecast as possible of any individual variable, such as imports or exports, a better prediction may be possible if modelling restrictions were minimized.

Second, both institutions rely heavily on expert judgement in making predictions. For both the OECD Economic Outlook and the IMF World Economic Outlook reports, individual country projections emanating from expert consultations are fed into the INTERLINK and MULTIMOD models and are then adjusted in an iterative fashion to ensure consistency (OECD, 2004; IMF, 2005). In that regard, the approaches are similar to economic forecasting methods employed by private providers, such as "Consensus Economics",³ which state 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" (Consensus Economics, 2005). Proceeding in such a manner entails a substantial degree of intransparency. Even if public institutions use well documented models at some point in the process, it is not clear to what extent and in what manner expert knowledge is brought to bear, be it to generate baseline predictions or to adjust model outcomes. This is particularly worrying in light of recent research showing that forecasts by international institutions may be subject to political considerations (Dreher et al., 2008).⁴

² The OECD has also developed a structural "International Trade Model" for 24 countries. See Pain et al. (2005). The OECD notes that the results of its structural models only serve "as a starting point to help animate the early stages of the OECD's forecasting round" (Rae and Turner, 2001: 4). Similarly, the IMF states that its MULTIMOD model "has not been designed to be a forecasting tool" (IMF, 2003).

³ 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>.

⁴ The authors analyze and explain the bias in IMF forecasts. They note that while initial projections are based on an econometric model, subsequent adjustment of forecast estimates leave much discretionary leeway for political influence. They hypothesize, *inter alia*, that governments are interested in optimistic forecasts, as economic environments perceived to be "good" may increase approval by their citizens. The authors find that the greater a country's direct influence at the Fund, the more optimistic the IMF's forecasts are for that country. This is particularly relevant for the countries examined in our paper (US, Euro area, Japan) which are major IMF shareholders and which, according to Dreher et al. (2008), receive particularly optimistic forecasts at election times.

The present paper proposes an alternative approach that addresses these shortcomings. First, the time series models we develop estimate import demand directly on the basis of historical relationships and key influencing factors. This eliminates constraints arising in large macroeconomic models that simultaneously estimate and seek consistency with other variables of interest. Second, the proposed approaches are simple and transparent⁵ and, hence, can easily be replicated. Our main objective then must be to show that our models perform at least as good as a naïve forecast and a comparable forecast made by any of the institutions mentioned above.

The main results of this paper are the following: We find that time series estimation techniques commonly employed in finance can readily be used to forecast growth in international trade. Ex-post forecasting and averaging over different intervals shows that forecasts obtained in this way perform at least as good as the widely quoted forecasts of trade growth published by the IMF in its semi-annual World Economic Outlooks. The paper is structured as follows: In the next section (Section 2), the time series models are introduced. Section 3 presents the data. Section 4 consists of two parts: First, the estimating equations are presented and the forecasting procedures are explained. Second, the results obtained in terms of both ex post and actual forecasts are discussed at both the aggregated and regional levels. Section 5 presents evaluation statistics, including in comparison to relevant IMF forecasts as well as a naïve forecast. Section 6 concludes.

2. IDENTIFICATION OF THE APPROPRIATE TIME SERIES MODELS

This section begins by providing some arguments why trade may be particularly amenable to time series forecasting, perhaps more so than other economic variables. The first subsection also introduces the

⁵ Of course, the time series forecasts should be exposed to expert judgement. Expert judgement can then act as a complement to time series analysis 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. 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". In any event, the influence of expert assessments should be documented. For more on a transparent integration of econometric methods and expert judgement for time series forecasting see Armstrong and Collopy (1998).

theoretical framework and key variables used for our approach. In the following subsections, the variables are tested for stationarity, optimal lag structure as well as Granger causality and cointegration. Proceeding in this manner allows us to identify the appropriate models.

2.1. TIME SERIES MODELLING OF TRADE

Time series forecasting models use the past movements of variables in order to predict their future behaviour. Unlike macroeconomic 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). These types of models, especially those taking advantage of volatility information, are widely used in finance, but have rarely been used in trade.⁶ While unable to explain causation, a time series model can still produce quite accurate forecasts if the regression explains much of the variation and is stable over time. Hence, time series analysis may be a simple and effective way to make forecasts when causal relationships are less clear. This is certainly the case for trade flows, where causal relationships in the international economy are manifold and multidirectional.⁷ The difficulty to account for causal relationships of certain variables is reflected in the block structure commonly used in macroeconomic models. Only some variables are determined via estimated equations, while others simply follow from model links and strongly depend on the accuracy of the model structure and the quality of the estimated equations. In many macroeconomic models, the latter is often the case for trade flows, which are more of a "by-product" of models constructed principally to forecast other variables or conduct policy simulations.

⁶ See Engle (2001). An exception in trade analysis, albeit with a different focus from our work, is Mansfield and Reinhardt (2008).

⁷ Other variables, such as GDP or inflation, may be more amenable to forecasting in the context of structural macroeconomic models. See, for instance, Batchelor (2001). See Diebold (1998) and Hendry and Clements (2003) 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. Diebold (2004) comes to the conclusion that simple, parsimonious models tend to be best for out-of-sample forecasting in many areas of business, finance and economics. Granger (2004) provides a good overview of recent applications of time series analysis in different areas.

At the same time, time series analysis need not be void of theoretical considerations. Additional variables can be included based on a theoretical framework. We employ a simple import demand framework, whereby import demand depends on its own lags, GDP and relative import prices, similar to Fair (2004) and Meacci and Turner (2001).⁸ This approach is grounded in the larger framework of gravity models that was originally developed by Tinbergen (1962). Its compatibility with traditional and new trade theories has been shown by Evenett and Keller (2002). Tinbergen (1962) explains the value of bilateral trade between two countries as being an increasing function of GDP of both trading partners (reflecting the assumption of increasing import demand and export supply capacity with a country's economic size). At the same time, he observed that trade flows were influenced negatively by trade costs and the wedge they create between importer and world prices.

In order to characterize trade growth, we prefer to focus on imports rather than exports for two main reasons: First, import data are more reliably collected than export data with the former serving for the assessment of tariffs by customs authorities. Second, the estimation of export demand would require information on world demand and world prices, which are more difficult to obtain than the national data required for estimating import demand. Also for reasons of data reliability, we focus on advanced economies for the moment, more specifically on the OECD-25 group of countries for which internally consistent data are readily available. These countries collectively account for about three quarters of global imports over the past 45 years; however, given their increasing weight in international trade emerging markets, notably China, would need to be included in future forecasting exercises if truly global predictions were to be made.⁹

The relationship between the variables used in our import demand framework, in a first instance, can be characterized by some descriptive statistics that show how growth rates and volatility look like over

⁸ Meacci and Turner (2001) and Fair (2004) model import volumes using total final expenditure and relative prices. Given an increasing import content of exports also in advanced economies (Hummels et al., 2001), we find it preferable to use GDP. Alternatively, GDP less imports could be used, as in Pain et al. (2005). Results are not very sensitive to these differences.

⁹ For China certain data shortages persist, notably in regard to prices, for which time series data have only been collected in recent years.

different time segments, and notably during key events over the sample period. Figure 1 shows the development of quarterly growth rates of imports and GDP for the OECD-25.¹⁰ It can be seen that imports are more volatile than GDP, which is also confirmed by the descriptive statistics contained in Table 1. Wider swings in the observed values of a variable imply a higher degree of uncertainty about influencing factors. Table 1 also indicates that both GDP and imports have seen quite some variation in average growth rates and volatility with the latter being above average for instance in the 1970s or in most recent years. Major recessions in the years 1974/75, 1981/82, 1993, and most recently 2008 (shaded areas in Figure 1) have affected imports more strongly than GDP, but subsequent recoveries were also more pronounced for the former than the latter. As the trends in Figure 1 show, GDP "anticipates" the development of imports over certain phases of our sample. For instance during the 1981/82 recession GDP growth was negative in the fourth quarter of 1981 and the first quarter of 1982, while import growth rates were still positive. In the subsequent three quarters of 1982 the import growth rate was constantly negative, while GDP growth had already recovered. Furthermore, imports and GDP increase more or less continuously for the years 1960 to 2008, but the increase in imports has been more pronounced (by a factor of 14) than the increase in GDP (about fourfold), especially over the last 20 years, which to an important extent might be explained with the increased international fragmentation of production (Figure 2).¹¹ The pronounced growth of imports during the "boom" between 2003 and 2007 is particularly remarkable. In this time period imports surged by more than 40 per cent, while GDP increased by 16 per cent. Both a higher volatility of quarterly growth rates and a stronger absolute increase of imports over time suggest that forecasting may be more difficult for imports than for GDP, as noted above. Figure 2 also shows the oil price as a proxy for import price developments at the aggregate OECD-25 level, where

¹⁰ 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.

¹¹ 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 1960s can be explained by taking account of vertical specialization. Similar results are obtained by Baldone et al. (2001) for the European Union in relation to Central Eastern Europe. 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.

import prices are not available.¹² Given the strong fluctuations of oil price growth rates (with annual standard deviations peaking at over 80 percentage points in the 1970s, as shown in Table 1), no major insights about its relationship with imports can be derived. From the graphical representation of our time series it also appears that stochastic properties have changed over time. We will therefore need to test for unit roots and remove stochastic trends by differencing, if necessary. We will also determine the optimal lag structure of each time series variable. In a multivariate context, it is also possible that the variables share a common stochastic trend and that their linear combination is stationary. Hence, co-integration will also be tested for. We also conduct a Granger-causality test to verify whether the additional variables suggested by our import demand framework may indeed improve predictability.¹³

Figure 1: *Quarterly growth rates of OECD-25 imports and GDP, 1960-2008 (per cent)*

Figure 2: *Volumes of OECD-25 imports and GDP, and real petroleum price, 1960-2008, (1960 = 100)*

2.2. STATIONARITY

For univariate analyses, a Box-Jenkins graphical plot is a straightforward tool to examine the existence of a unit root. The time series of the logs of OECD-25 imports is non-random, since its autocorrelations are non-zero.¹⁴ The series has a rather high degree of autocorrelation between adjacent and near-adjacent observations. The slow decay of log imports for the structure under consideration (40 lags) indicates that the time series exhibits a unit root. The autocorrelation plot of the first differences of log OECD-25 imports 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. The differenced time series fulfils

¹² See Amano and van Norden (1998) for a similar approach. In recent times, the oil price also acted as an important indicator of demand growth expectations.

¹³ 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.

¹⁴ For all variables both the Ljung-Box portmanteau (Q) test and Bartlett's periodogram-based test reject at a highly significant level the hypothesis that the time series are generated by a white noise process.

the condition of weak dependence. 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. As a complement to Box-Jenkins, we apply the Augmented Dickey-Fuller test to check for stationarity as well as trend-stationarity by including an appropriate time trend in the Augmented Dickey-Fuller estimation. In both cases, we cannot reject the null hypothesis of the existence of a unit root for the logs of the OECD-25 import data. 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. This result is also confirmed by the Zivot-Andrews test, which considers the null hypothesis of a unit root with no break against the alternative of a stationary process with a break. The resulting import time series in log differences amounts to growth rates, i.e. our variable of interest. The same transformations have been made for the other variables, which are all integrated of order one.

2.3. LAG STRUCTURE

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. The autocorrelation function features an exponential decrease and then a sine wave-like pattern, i.e. a typical autoregressive process of an order higher than one and no moving average process (q=0). The significant spike at lag one of the partial autocorrelation (PAC) function of log OECD-25 imports in first differences and the lack of significant partial correlation at higher lags point to an autoregressive process of order one. 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). Besides the Box-Jenkins method for selecting the most appropriate lag structure we use the conventional information criteria. In the case of OECD-25 imports, the Akaike information criterion reaches its minimum for four lags. This number of lags is also confirmed by the Schwarz

Bayesian information criterion.¹⁵ In a similar fashion, we obtain the optimal number of lags for the other variables, e.g. one lag for OECD-25 GDP and five lags for the oil price.

2.4. GRANGER CAUSALITY

In order to know whether the inclusion of further variables, notably GDP and the oil price (alternatively relative import prices at the country/regional level) improves predictability, we apply a "Granger-causality" test. 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 GDP and the oil price both jointly and individually "Granger-cause" import growth, although the results for the GDP variable alone may be sensitive to the choice of lags. The test values, including for GDP 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.

2.5. COINTEGRATION

As discussed above, we also test for cointegration. If our variables were cointegrated, we could estimate a vector error correction model (VECM). Rather than making the nonstationary variables stationary through differencing, we could use the fact that their combination can produce a stationary error term. We apply the Augmented Dickey-Fuller and Johansen cointegration tests to the levels of OECD-25 imports, GDP and the real oil price. On the basis of these tests, for a variety of lag structures, the null hypothesis of no cointegration of imports and GDP is rejected at the 5 per cent significance level in most, but not all of the cases. Imports and the oil price do not appear to be cointegrated. Even though some of the test results show imports and GDP to be cointegrated, Stock and Watson (2003) warn that cointegration tests can be misleading, since they frequently improperly reject or fail to reject the null

¹⁵ Diebold (2004) notes that the Box-Jenkins method provides graphical aids to model selection that should be seen as a complement to the information criterion methods. Use of the latter results in a more systematic and replicable model selection, while the former provides important summary information of the dynamics of time series data, such as seasonality etc. However, the Box-Jenkins method is more art than science, which also explains why by mere visual inspection it is not unlikely that fewer lags are included in the model than what is suggested by the information criteria. As such, the Box-Jenkins method may also be seen as a backdrop on the prudent side

hypothesis of no cointegration. If variables that are not cointegrated are modelled using a VECM, the error term is assumed to have a 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 either. 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 (Abeyasinghe, 1998). In light of these caveats, we do not expect a VECM to be a preferred modelling choice; indeed, forecast performance turned out to be poor and the VECM approach is therefore not pursued further here.

3. DATA

Our analysis mainly relies on OECD Quarterly National Accounts (QNA) data and data from the IMF's International Financial Statistics (IFS). Since the data are observed at quarterly intervals, they might exhibit seasonality. Macroeconomic forecasting is geared towards projecting non-seasonal fluctuations, and seasonality should be removed to the extent possible. Therefore, all the data we use, have been seasonally adjusted.¹⁶

The dependent variable are quarterly imports of goods and services by the former OECD-25 countries. The maximum length of the time series we use stretches from 1960 to the second quarter of 2009. From this, we calculate volume indices (year 2000 = 100). Real GDP data are available from the OECD for the same time frame and in the same format as imports. Future values for the forecast time period are constructed from IMF predictions of GDP. Both imports and GDP are also estimated simultaneously using vector autoregressions (VAR), which obviates the need to obtain contemporaneous values for the

against which the information criteria and their respective penalization of the loss of degrees of freedom can be checked.

¹⁶ 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. Abeyasinghe (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. For instance, Box-Jenkins analysis requires seasonally adjusted time series. From the OECD, seasonal adjusted data are available. Our own seasonal adjustment proceeds as follows: 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. Remaining seasonality has been verified by a seasonal subseries plot. The quarterly means in our import data are almost identical, indicating the absence of seasonality.

exogenous variable. For the oil price, we use the petroleum average crude price, which is composed of the Dubai, UK Brent and West Texas Intermediate petroleum prices in dollars per barrel. The United States CPI is used to calculate real oil prices in constant 2000 dollars. These data are sourced from the IMF's IFS. It is common practice in many forecasting exercises to hold the real oil price constant for future time periods at the level of the last quarter for which data are available. However, owing to the record increase in nominal oil prices until mid-2007 and the steep reversal in subsequent quarters, we have used futures prices from the monthly oil market reports by the International Energy Agency (IEA) to make more realistic projections.¹⁷

For selected countries/regions, namely the United States, the European Monetary Union (EMU) (i.e. the original 12 countries of the Euro area) and Japan, a second set of data is constructed. Again, import and GDP data are in real terms (seasonally adjusted, constant dollars) expressed as volume indices (year 2000 = 100). For the EMU, we can use the ifo economic climate index for the euro area to predict GDP growth.¹⁸ Again, we also include the oil price. For the United States, we substitute relative import prices deflated with the consumer price index (CPI) for the real oil price. In the country/regional equations, we have alternatively introduced real exchange rates in order to take account of changes in the relative price indices of trading partners. 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. Regional/country import and demand data are from the OECD, oil price, relative import price and real effective exchange rate data are sourced from the IMF's IFS.

4. ESTIMATIONS AND FORECASTS

In the first part of this section, we describe the estimation of model parameters. The second part discusses the application of the models. This includes both ex post forecasts (in-sample forecasts) and the

¹⁷ The IEA reports are available at <http://omrpublic.iea.org/>. Alternatively, we could fit any exogenous variable with an autoregressive (AR) model determining lag length in the usual manner.

¹⁸ See <http://www.cesifo-group.de/>.

extrapolation of the time series beyond the sample period using the estimated parameters (out-of-sample forecasts).

4.1. THE MODELS

The two main types of models we employ are univariate pure time series and multivariate structural time series models. The univariate models are of an autoregressive ARIMA (p,1,0) nature of various orders p. Hence, in the case of OECD-25 imports (optimum four lags), they are of the following form:

$$M_t = \beta_0 + \beta_1 M_{t-1} + \dots + \beta_4 M_{t-4} + u_t \quad (1)$$

with M_t denoting first differences of log imports and u_t being white noise.¹⁹ The model is estimated using ordinary least-squares (OLS). The multivariate specifications are either autoregressive distributed lag (ADL), or VAR models. The ADL(p,q) forecast model for imports with GDP includes four own lags (p = 4) as well as the contemporaneous value and one lag of GDP (q=1). It is of the form:

$$M_t = \beta_0 + \beta_1 M_{t-1} + \dots + \beta_4 M_{t-4} + \delta_0 GDP_t + \delta_1 GDP_{t-1} + u_t \quad (2)$$

with M_t (GDP_t) denoting first differences of log imports (GDP) and u_t being white noise. 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 GDP and constant variance, we notice in the plot of United States import growth rates (see Figure 3) 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-25.²⁰ 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

¹⁹ 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. The same tests have been carried out for all other models in the paper.

²⁰ See also the descriptive statistics in Table 1.

ARCH effects. 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. 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. GARCH models are widely used in finance to forecast stock prices taking into account risk-return relationships. In our models, we only estimate GARCH (1,1) equations, i.e. we only consider the first lags of u_t and σ_t :²¹

$$\sigma_t^2 = \alpha_0 + \alpha_1 u_{t-1}^2 + \phi_1 \sigma_{t-1}^2 \quad (3)$$

The GARCH models are estimated using maximum likelihood.

Figure 3: *Quarterly import growth rates, United States and OECD-25, 1960-2008 (per cent)*

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

²¹ Instead of using GARCH (1,1) 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, the GARCH (1,1) model

examining the significance of the coefficients. With four lags, our VAR model for imports and GDP consists of the following two equations estimated by OLS:

$$\begin{aligned} M_t &= \beta_{10} + \beta_{11}M_{t-1} + \dots + \beta_{14}M_{t-4} + \gamma_{11}GDP_{t-1} + \dots + \gamma_{14}GDP_{t-4} + u_{1t} \\ GDP_t &= \beta_{20} + \beta_{21}M_{t-1} + \dots + \beta_{24}M_{t-4} + \gamma_{21}GDP_{t-1} + \dots + \gamma_{24}GDP_{t-4} + u_{2t} \end{aligned} \quad (4)$$

with M_t (GDP_t) denoting first differences of log imports (GDP) and u_t being white noise.

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. There are 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 recent date of the regressor being $t-h$. 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. 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 (Harvey, 1988; Stock and Watson, 2003). Both methods can be used for multivariate forecasts as well (ADL and VAR). In all cases, the iterative method results in a superior forecast performance and we only present this approach here. For instance, the two- and three-quarters ahead iterated forecasts for the ARIMA (4,1,0) model of imports would be:

$$\begin{aligned} \hat{M}_{t|t-2} &= \hat{\beta}_0 + \hat{\beta}_1 \hat{M}_{t-1|t-2} + \hat{\beta}_2 \hat{M}_{t-2} + \hat{\beta}_3 \hat{M}_{t-3} + \hat{\beta}_4 \hat{M}_{t-4} \\ \hat{M}_{t|t-3} &= \hat{\beta}_0 + \hat{\beta}_1 \hat{M}_{t-1|t-3} + \hat{\beta}_2 \hat{M}_{t-2|t-3} + \hat{\beta}_3 \hat{M}_{t-3} + \hat{\beta}_4 \hat{M}_{t-4} \end{aligned} \quad (5)$$

with $\hat{\beta}$ denoting OLS estimates of the coefficients, M_t denoting first differences of log imports and \hat{M} being forecasted values thereof.

is by far the most important case in practical applications. See Diebold (2004), in particular page 392 and footnote 7, as well as Engle (2001).

4.2. 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"). For instance, in 2009, the third and fourth quarter forecasts are combined with actual import growth rates in the first and second quarters to get the current year forecast. For 2010, all quarters must be forecasted to obtain the annual growth rate. Results of the annual import growth forecasts obtained from different models are compared on the basis of standard evaluation measures, most notably the root mean square forecast error (RMSFE).

Table 2 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.²² We begin with a simple ARIMA (4,1,0) model with a RMSFE for the current year of 0.63 and 3.31 for the next (forecast 1 in Table 2). This model has the advantage that no future values of exogenous variables are needed as inputs. The ADL model with GDP as an additional predictor (forecast 2 in Table 2) performs considerably better for both the current and following year with RMSFEs of 0.41 and 1.74, respectively.²³ However, in this case actual values of GDP growth have been employed. This is equivalent to assuming that GDP predictions as an input in the model have proved to be correct. In reality, this is not the case and RMSFEs may therefore be higher. To illustrate this divergence we re-run the ADL model with GDP as an

²² In the table, for space reasons, only the years 1995 to present are shown. However, as is standard practice, we select an estimation period, here 30 years, and make two to six quarter ahead forecasts on a rolling basis. For example, we first estimate our models for the 1960-1989 period, make two to six quarter ahead forecasts, then estimate and forecast again using the 1961-1990 period and so on until 1979-2008. The results are then averaged. Proceeding in this manner seeks to prevent that results are period-specific. By making forecasts from different forecast origins for the chosen forecast period (two to six quarters ahead) and averaging in order to determine standard evaluation criteria, the idea is to eliminate (or, average out) sample-specific concerns and especially the end-period problem in prediction. This way, better and stronger conclusions about the out-of-sample forecast performance of a model can be reached.

²³ We only exhibit the GARCH results in the paper, since the RMSFEs of the latter consistently outperform the OLS-based forecasts, as was to be expected given our discussion on heteroscedasticity above. We have also run an ADL model with the real oil price as the only additional predictor as well as one containing both GDP and the oil price. However, their performance is clearly inferior to our preferred ADL specification and results are therefore not presented.

additional predictor using the actual GDP forecasts by the IMF available at the time of forecasting (forecast 3 in Table 2). The results for the next year forecasts are still better than the outcomes obtained from the ARIMA (4,1,0) model, but the current year forecasts are slightly worse. This is perhaps not surprising for two reasons: First, ARIMA models are known to do well in very short-term forecasting, in this case two quarters ahead. Second, the IMF short-term forecasts of GDP that are required for the ADL model are not updated as frequently as forecasts made by other institutions and are of a lower quality on average (Batchelor, 2001). The ADL model evaluation would turn out better if historical time series of GDP forecasts, such as the ones by Bloomberg, Reuters or Consensus Economics were available to us, which would actually have been used at the time of forecasting. Unfortunately, we only have access to historical information of GDP predictions from the IMF, which we gathered from the archives of past World Economic Outlooks, and, hence, the RMSFE of the ADL GARCH model is likely to be overestimated.

Table 2: Annual import growth rates, OECD-25, 1995-2010, observed and forecasted (per cent)

Both the ARIMA and ADL models provide quite accurate forecasts for several years, for instance in the recent past 2006 and 2007. The two-digit import growth rates (11 and 12 per cent) in 1997 and 2000 are foreseen relatively well in the current year, but somewhat underestimated by both models in the preceding year. Yet, even with forecasts of 7 per cent both models still provide a strongly optimistic outlook for 2000 in the year before. 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, 1982 and, most likely, 2009) where import growth rates were negative in the last half century, both models fail to predict the actual trade reduction one year before. However, especially the ADL model, in 2001, accurately foresees the moderate pick-up in import growth of about 2 per cent in 2002 and the stronger recovery in 2003. Here, the ARIMA model tends to exaggerate the turnaround in 2002. The unusual developments in these years carry through to the 2003 forecast, which tends to be overoptimistic.

A closer look at the results of the VAR (forecast 4 in Table 2) reveals RMSFEs worse than those of the ARIMA and ADL models for both the current and next year. 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 4.5 and 6.0 per cent for most years. 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, with 4.8 and 5.0 per cent the VAR model considerably underestimates import growth in the years 1997 and 2000, whereas the ADL model (forecast 3 in Table 2) produces better results in those exceptionally good years, while suffering from higher deviations in others.

In an attempt to reduce the weight of extreme historical observations of imports and account for the possibility of structural breaks, we could include various time dummies in our models. The dummies would take the 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. The Quandt-Andrews breakpoint test, which tests for one or more unknown structural breaks, fails to reject the null hypothesis of no breaks. However, we run Chow tests on a few suspect dates, including the recessionary periods shown in Figure 1 (shaded areas), which, with the exception of the aftermath of the first oil crisis in 1974, do not indicate structural breaks. When we nevertheless include dummies (e.g. in the third quarter of 2001 to account for 9/11 and the following one or two quarters), the variation in the forecasts is reduced with no improvement of the RMSFE. Some years that are rather accurately forecasted by the model without dummies are underestimated by the dummy version. At best, the inclusion (separately or together) of time dummies, leaves the RMSFE about unchanged, as errors are reduced in the years that are difficult to forecast; in most cases, however, the RMSFE deteriorates.

In Table 3, the best forecasting model for each individual country/region is presented. 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 for Japan is practically unusable. For all countries/regions, the

ADL model with GDP and the oil price as additional predictors performs best, albeit obviously with different lag structures. Several other specifications of the individual models are not presented in the table. For all countries and regions, we have included real effective exchange rates or real import prices, but the resulting RMSFEs are rather high. The simple ARIMA (p,1,0) models do not improve the forecasts in any of the cases examined either.

From 1995 to 2004, 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; the opposite is the case in the years afterwards, and our forecasts appear to capture this tendency. 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, with the exception of the boom in 2006 and 2007, where actual import growth exceeded the OECD average. These developments are predicted quite well by our models. Japan features a lot more variation in import growth rates. The effects of double-digit import growth in 1995 and 1996 were strongly reversed by 1998 in the context of the Asian financial crisis, and Japan's import have grown less than the OECD average in practically all years since. The six quarter forecasts for Japan have an elevated RMSFE, heavily influenced by the strong variability in import performance of the second half of the 1990's, but in recent years have become better.

Table 3: *Annual import growth rates in selected countries, 1995-2010, observed and forecasted (per cent)*

5. FORECAST EVALUATION

Which of the models should be used in making a benchmark forecast? And, should we give preference to our aggregate OECD-25 forecast or make region/country predictions for the US, EMU-12 and Japan and aggregate the results in order to obtain an overall forecast for OECD import growth? Finally, how does our preferred model compare to IMF forecasts of import growth and a naïve forecast? On the first question, 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 ARIMA (p,1,0) 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 the ADL model with GDP (using actual IMF forecasts), for example, only four ex post forecasts in the last fifteen years would have missed actual import growth by more than one per cent for the ongoing year (and always less than two per cent). The "next year" forecast with an RMSFE of around 3 seems acceptable given the long forecast horizon of six time periods and the fact that some of the more "difficult" years have been predicted reasonably well. In any event, results should be interpreted in the light of expert knowledge, for instance on expected policy developments in the year to come, such as demand-enhancing fiscal packages in the present situation.

Second, in order to decide on the merits of a forecast of aggregated OECD data versus an aggregation of region/country forecasts to get an overall OECD number, we have weighted the US, EMU and Japan ex-post forecasts between 1990 and 2008 with their relative share in real imports (using average annual exchange rates for currency conversion). We obtain an RMSFE of 0.52 and 1.85 for the current and next year forecasts respectively. This compares with a 0.41 and 1.74 RMSFE for our best aggregated OECD model (see forecast 2 in Table 2). We expected that an aggregation of regional forecasts could perform better, as it allows for the inclusion of additional (region specific) explanatory variables and divergences in the explanatory content of the same variable in different regions. Also, forward looking indicators, such as the ifo index for the EMU area, could be included that are based on business surveys and might prove a good alternative to institutional GDP forecasts. Apparently, however, for an assessment of advanced economies import growth, an aggregation of the three major country/regional forecasts is not necessarily better than an aggregate OECD forecast, although results are not widely different. One reason may be that the economic importance of other OECD economies, such as Australia, Mexico, the Republic of Korea or the United Kingdom, has increased in recent years and, hence, that import growth performance in the OECD can less well be approximated by the US, EMU and Japan alone. Also especially the Japanese data show quite some variation over the years, while the aggregated OECD data are inherently less volatile.

Finally, we need to compare the performance of our preferred model – the ADL model with GDP – to a naïve forecast benchmark as well as the IMF's performance of import growth predictions, as has been done elsewhere in the literature (Wu and Chen, 2001). It is noteworthy that our forecasts for most of time in the past 20 years diverge substantially from the long term trend of import growth of about 4.5 to 5.5 per cent. Such behaviour is closer to reality, which features quite some variation in annual trade growth rates from one year to the next. Judging from conventional error measures to assess relative model performance, it is not surprising, therefore, that our forecasts perform better than a naïve approach that projects the deterministic trend into the future for both the current and next year (see Table 4). The variation in our forecasts is also a distinguishing feature when compared to the predictions made by the IMF in past World Economic Outlooks, which usually lie between 5 and 6 per cent (see Table 2, last row). Again, the tendency of IMF predictions to stay close to the long-term trend, especially as far as six quarter ahead forecasts are concerned, may play a role in explaining why our models outperform the IMF predictions of import growth for both the current and next year (see Table 4).²⁴ Results that are broadly in line with the long-term trend are usually of limited value for short-term decision-making. 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 4.

Table 4: *Forecast evaluation statistics*

²⁴ It is important for our comparison with the IMF that we use the available information set at the time of forecasting and not the definitive values of the variables as known at the final data point. This is why we employ the GDP forecasts available at the time and not actual GDP values in our comparison with the IMF. Moreover, the time series we use, especially GDP data, are subject to revisions, and part of the forecasting error comes from such revisions. However, we find these revisions to be modest, namely 0.03 per cent on average for GDP and 0.1 per cent for imports (mean absolute revision for data published one year after; mean revision is practically zero implying that positive and negative revisions cancel out each other). McKenzie and Adam (2007) confirm that these revisions are generally statistically insignificant. We therefore believe that our results above remain unaffected by data revision.

6. CONCLUSIONS

In this paper, we have developed a time series approach in order to forecast growth in international trade, focussing on major advanced economies' annual import developments. Our preferred specification is an ADL model of imports with GDP as an additional predictor using a GARCH estimation. The model performs 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 at least as good as a naïve forecast and more elaborate IMF predictions. In contrast to large macroeconomic models combined with expert opinion our approach is both parsimonious and fully documented. In light of their comparatively strong performance, "mechanical" time series forecasts have a lot to offer compared to more information- and resource intensive approaches and may be used as a benchmark to either confirm the forecasts of larger models or, in case of divergence, encourage the modellers to explain the drivers of their results or the influence of expert opinion.

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. 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. 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.

References

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

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

Amano, R. A. and van Norden, S. (1998) 'Exchange Rates and Oil Prices', *Review of International Economics* 6, 4: 683-694.

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.

Baldone, S., Sdogati, F. and Tajoli, L. (2001) 'Patterns and Determinants of International Fragmentation of Production: Evidence from Outward Processing Trade between the EU and Central Eastern European Countries', *Weltwirtschaftliches Archiv* 137, 1: 80-104.

Batchelor, R. (2001) 'How useful are the forecasts of intergovernmental agencies? The IMF and OECD versus the consensus', *Applied Economics* 33, 2: 225-235.

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?*, London: Consensus Economics Inc.

Diebold, F. X. (1998) 'The past, present, and future of macroeconomic forecasting', *Journal of Economic Perspectives* 12, 2: 175-192..

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

Dreher, A., Marchesi, S. and Vreeland, J. (2008) 'The political economy of IMF forecasts', *Public Choice* 137, 1: 145-171.

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

Evenett, S. J. and Keller, W. (2002) 'On theories explaining the success of the gravity equation', *Journal of Political Economy* 110, 2: 281-316.

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

Granger, C. W. J. (2004) 'Time Series Analysis, Cointegration, and Applications', *American Economic Review* 94, 3: 421-425.

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*, Washington: IMF.

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

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

Mansfield E. D. and Reinhardt E. (2008) 'International institutions and the volatility of international trade', *International Organization* 62, 3: 621-652.

McKenzie, R. and Adam, Z. (2007) 'Revisions in quarterly GDP of OECD Countries: An Update', Paper presented at the Working Party of National Accounts, October 2007, Paris: OECD.

Meacci, M. and Turner, D. (2001) 'Modelling import responsiveness for OECD manufactures trade', Organisation for Economic Co-operation and Development (OECD) Economics Department Working Paper No. 311, Paris: OECD.

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.

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.

Tinbergen, J. (1962) *Shaping the World Economy: Suggestions for an International Economic Policy*, New York: Twentieth Century Fund.

World Trade Organization (WTO) (2009) *Report to the Trade Policy Review Body from the Director-General on the Financial and Economic Crisis and Trade-Related Developments*, WTO Document WT/TPR/OV/W/2, Geneva: WTO.

World Trade Organization (WTO) (2008) *World Trade Report 2008: Trade in a globalizing world*, Geneva: WTO.

Wu, J.-L. and Chen, S.-L. (2001) 'Real Exchange-Rate Prediction over Short Horizons', *Review of International Economics* 9, 3: 401-413.

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

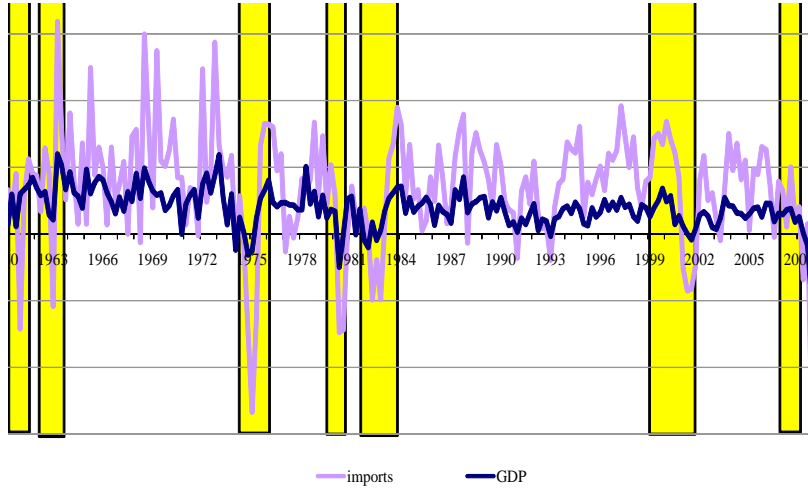
Tables and FiguresFigure 1: *Quarterly growth rates of OECD-25 imports and GDP, 1960-2008 (per cent)*

Figure 2: *Volumes of OECD-25 imports and GDP, and real petroleum price, 1960-2008, (1960 = 100)*

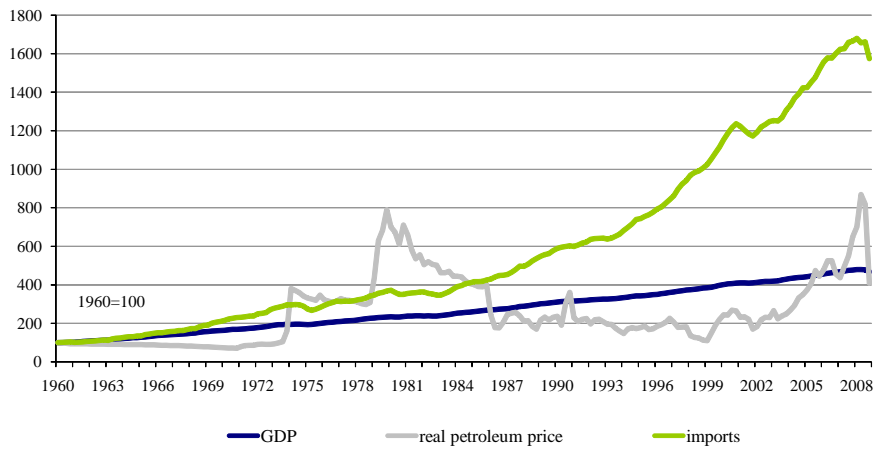


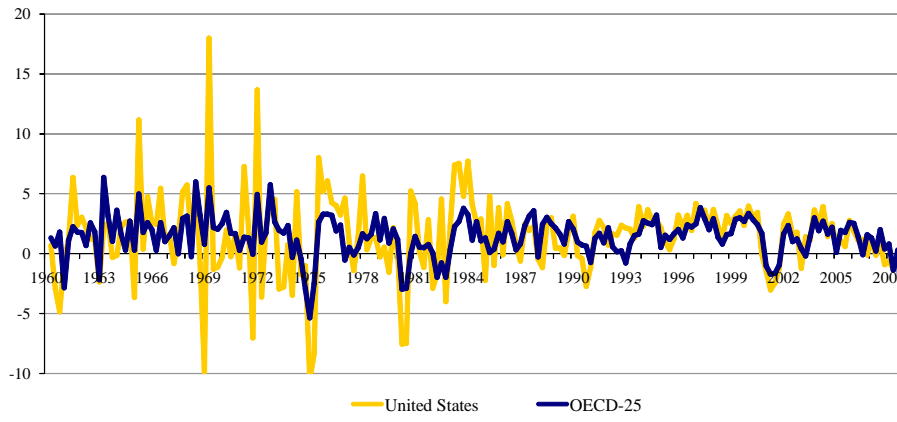
Figure 3: *Quarterly import growth rates, United States and OECD-25, 1960-2008 (per cent)*

Table 1: *Descriptive statistics (annual), OECD-25, 1960-2009 (per cent)*

	1960-1969	1970-1979	1980-1989	1990-1999	2000-2009	1960-2009
GDP						
Volatility (StDev)	1.1%	1.9%	1.5%	0.8%	2.0%	1.9%
Avg Growth Rate	5.3%	3.7%	2.8%	2.5%	1.9%	3.2%
IMPORTS						
Volatility (StDev)	3.7%	5.9%	4.9%	3.3%	6.6%	5.2%
Avg Growth Rate	8.0%	6.1%	4.7%	6.7%	3.8%	5.8%
OIL PRICE						
Volatility (StDev)	1.7%	80.5%	24.1%	27.5%	35.4%	45.8%
Avg Growth Rate	-2.7%	37.4%	-6.6%	0.0%	17.5%	9.3%

Table 2: Annual import growth rates, OECD-25, 1995-2010, observed and forecasted (per cent)

OECD		1995	1996	1997	1998	1999	2000	2001	2002	2003	2004	2005	2006	2007	2008	2009	2010	RMSFE
observed	import growth	7.2%	7.3%	11.1%	9.0%	8.3%	12.0%	-0.2%	2.2%	3.8%	8.6%	6.5%	7.5%	4.1%	0.0%	-	-	-
IMP: 4 lags, ARIMA (4, 1, 0)	current year	6.8%	6.3%	10.2%	8.8%	7.4%	11.1%	0.6%	3.3%	2.7%	8.2%	5.9%	7.9%	3.7%	1.2%	-17.0%	-	0.63
	next year	6.4%	5.1%	5.4%	7.1%	4.9%	7.0%	5.9%	3.1%	7.9%	4.0%	5.4%	6.0%	4.8%	4.7%	3.2%	1.7%	3.31
IMP: 4 lags, GDP: contemp + 1 lag, ADL (GARCH)	current year	6.8%	6.4%	10.1%	9.0%	7.7%	10.6%	0.0%	2.5%	3.5%	8.0%	5.8%	7.6%	3.9%	-0.5%	-16.0%	-	0.41
	next year	4.7%	5.5%	7.2%	6.7%	6.2%	9.4%	1.9%	0.8%	4.2%	6.6%	5.7%	6.8%	4.4%	0.9%	-17.8%	1.5%	1.74
IMP: 4 lags, GDP(IMF forecasts): contemp + 1 lag, ADL (GARCH)	current year	7.2%	5.5%	9.2%	7.7%	7.2%	11.2%	0.5%	2.9%	3.5%	8.9%	5.8%	7.8%	3.9%	0.9%	-16.0%	-	0.80
	next year	5.1%	4.5%	4.9%	6.5%	3.7%	7.2%	6.0%	2.1%	5.4%	6.3%	6.6%	6.3%	4.8%	3.6%	-1.4%	1.5%	3.10
IMP: 4, lags, GDP: 4 lags, VAR	current year	6.0%	6.2%	9.6%	8.2%	7.0%	11.1%	0.7%	2.8%	2.6%	8.0%	5.8%	8.0%	3.9%	1.2%	-16.1%	-	0.86
	next year	2.6%	1.6%	4.8%	4.4%	2.0%	5.0%	6.0%	3.5%	5.5%	3.4%	4.3%	5.6%	5.0%	5.2%	3.2%	3.7%	3.98
IMF 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%	7.5%	4.3%	1.9%	-13.6%	-	1.86
	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%	6.0%	5.0%	1.1%	0.6%	3.91

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 a rolling basis for thirty-year intervals for the years 1990-2008. For space constraints only the years 1995 and thereafter are shown in the table.

Table 3: Annual import growth rates in selected countries, 1995-2010, observed and forecasted (per cent)

	1995	1996	1997	1998	1999	2000	2001	2002	2003	2004	2005	2006	2007	2008	2009	2010	RMSFE	
USA																		
observed import growth	7.8%	8.3%	12.6%	11.1%	10.9%	12.3%	-2.9%	3.3%	4.3%	10.5%	6.0%	5.9%	2.0%	-3.2%	-	-	-	
IMP: 3 lags, GDP: contemp + 1 lag, Oil: contemp + 5 lags, ADL (GARCH)	current year	8.2%	8.0%	12.3%	12.5%	9.9%	11.4%	-1.1%	2.5%	4.9%	10.5%	5.6%	5.5%	2.9%	-2.6%	-16.2%	-	0.82
	next year	5.9%	7.8%	9.4%	10.6%	11.4%	8.3%	2.4%	2.7%	5.4%	8.1%	7.0%	5.4%	4.1%	1.8%	-8.7%	5.3%	2.62
Japan																		
observed import growth	13.2%	12.7%	0.6%	-7.0%	3.4%	8.8%	0.8%	0.9%	3.9%	7.8%	5.7%	4.2%	1.4%	0.8%	-	-	-	
IMP: 5 lags, GDP: contemp + 2 lag, Oil: contemp + 1 lags, ADL (GARCH)	current year	10.2%	13.3%	1.1%	-7.7%	2.6%	7.7%	1.2%	0.7%	2.1%	6.9%	5.3%	5.4%	3.2%	-1.7%	-20.0%	-	1.27
	next year	1.6%	4.7%	2.7%	-4.2%	-3.2%	6.7%	2.4%	-5.2%	6.5%	2.8%	4.6%	4.8%	4.9%	2.8%	-12.3%	-1.3%	4.95
EU																		
observed import growth	7.3%	3.5%	8.8%	9.6%	7.3%	11.1%	2.3%	0.4%	3.1%	6.2%	5.7%	8.0%	5.1%	0.9%	-	-	-	
IMP: 2 lags, GDP: contemp + 5 lag, Oil: contemp + 4 lags, ADL (GARCH)	current year	7.0%	2.9%	8.0%	9.3%	7.4%	9.7%	2.9%	0.2%	2.8%	5.8%	5.3%	7.9%	5.1%	0.1%	-13.7%	-	0.54
	next year	5.5%	2.0%	4.8%	7.2%	6.4%	9.7%	2.6%	0.8%	1.7%	5.4%	4.8%	8.4%	6.3%	1.4%	-14.2%	4.8%	1.60

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 a rolling basis for thirty-year intervals for the years 1990-2008. For space constraints only the years 1995 and thereafter are shown in the table.

Table 4: *Forecast evaluation statistics*

forecast current year			
	naïve	IMF	<i>ADL (GARCH) model</i>
RMSFE	0.98	1.86	0.80
Theil's U	0.08	0.15	0.06
forecast next year			
	naïve	IMF	<i>ADL (GARCH) model</i>
RMSFE	3.92	3.91	3.10
Theil's U	0.31	0.32	0.27

Notes: Besides the RMSFE, we show Theil's U coefficient, since it is a scale-insensitive measure of the RMSFE in relative terms, standardized to values between 0 and 1, and, therefore, allows for a comparison of forecasts from different origins.