CPB Background Document / *CPB Achtergronddocument*

Is world trade data useful when forecasting a small open economy?

Is wereldhandelsdata nuttig voor het voorspellen van een kleine open economie?

26 June 2012

Adam Elbourne

Niels Vermeer

Is world trade data useful when forecasting a small open economy?

Adam Elbourne and Niels Vermeer

June 26, 2012

Abstract

This paper uses pooled VAR forecasts to evaluate the contribution of world trade growth data to forecast accuracy for GDP growth and export growth for the Netherlands, a small-open economy. By using pooled VAR forecasts we have attempted to minimise a number of issues that make inference difficult regarding the importance of individual variables for forecast accuracy using traditional methods. We find that the most accurate forecasts for GDP growth can typically be made without using the world trade data that would have been available at the time the forecasts were made. Furthermore, this paper has shown that world trade data also doesn't improve forecast accuracy for export growth, the component of GDP most closely linked to world trade. As a robustness test we also repeated the forecasting competition using direct multi-period forecasts (see Pesaran et al. (2011) and Marcellino et al. (2006) for recent examples) but they were statistically no more accurate than the standard VAR approach. We examine why world trade data doesn't help and find a non-linear relationship between the accuracy of world trade forecasts and GDP forecasts. Whilst perfect foresight of world trade would improve forecast accuracy for both GDP and export growth, the improvement in world trade forecasts from the pooled VAR models relative to a 'no knowledge' benchmark does not.

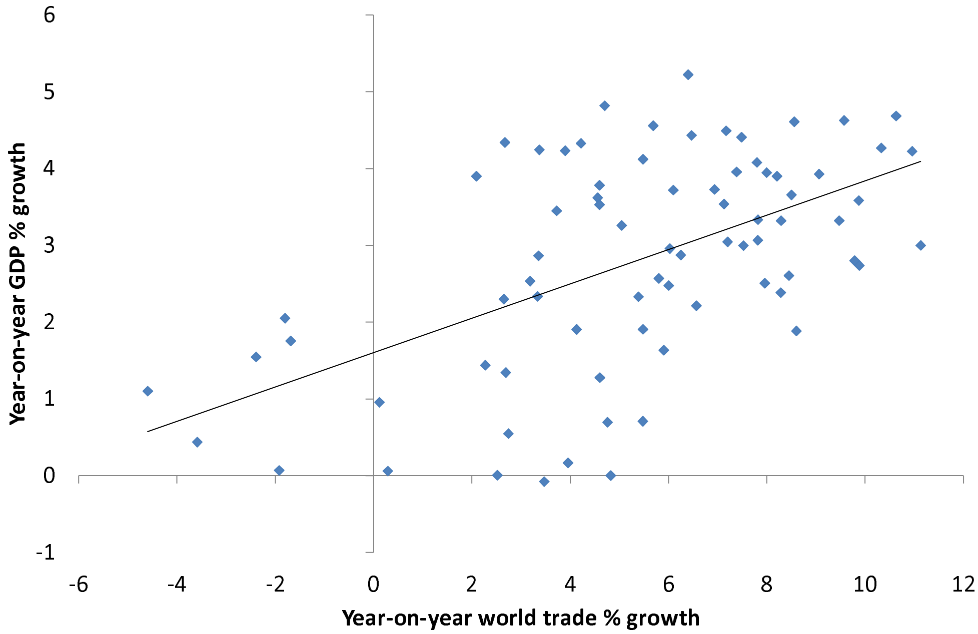
1 Introduction

The Netherlands is a small open economy where the sum of exports and imports is about 150% of GDP (Source: CBS) and the port of Rotterdam is the largest port in Europe in both container shipping and cargo (Source: Port of Rotterdam Authority). It seems fairly obvious that economic activity in the Netherlands is highly dependent on international trade; figure 1 shows a clear positive correlation between current GDP growth and the growth in world trade[1]. Traditional macro models for the Netherlands embody this close relationship. The current workhorse model for forecasting and policy analysis at CPB Netherlands Bureau

[1]In this study we use the releveant world trade for the Netherlands, which is reweighted to match the relative importance of particular trade flows.

for Economic Policy Analysis[2] is SAFFIER (see Kranendonk and Verbruggen (2007) and Kranendonk and Verbruggen (2010)), which is no exception to the rule that traditional small open economy models place a great deal of emphasis on world trade as a key driving factor behind economic activity.

Figure 1: Contemporaneous correlation between world trade and GDP growth: 1989-2007



However, forecasting concerns the relationship between what we know now and what will happen tomorrow, not the contemporaneous relationship between variables. The process of forecasting GDP using traditional models places great weight on world trade, largely because the model embodies the important contemporaneous relationship between world trade and GDP growth. To make a forecast for GDP growth requires that a forecast for world trade growth over the period in question is supplied to the model. But all we know at the time of making the forecast is current and past world trade. High contemporaneous correlation doesn't imply our knowledge of current world trade is useful for predicting GDP growth, even true causation does not imply that it is useful for forecasting[3]. Table 1 shows a simple example of this for the Netherlands.

[2]There is a long tradition in macro economic forecasting and policy evaluation at the CPB using macro models (see Don and Verbruggen (2006)).

[3]Suppose GDP growth is determined solely by one-to-one causal relationship with world trade growth and that world trade growth follows a simple AR process. If we have time series of both GDP growth and world trade growth all of the information in the world trade series is contained in the GDP growth series, since they are identical. Therefore, knowledge of world trade growth will not improve forecast accuracy over and above a simple AR model of GDP

Table 1: MAE of GDP forecasts from 1999 to 2007 (% points)

	1 year ahead	2 years ahead
Baseline VAR	0.86	1.38
Plus world trade	0.82	1.37
Plus exogenous realised world trade	0.58	0.71

The first row shows the mean absolute error for forecasting GDP growth from a simple two-variable model of GDP growth and unemployment. Adding future realised world trade as is done in the third row dramatically improves the forecast accuracy from 0.86% to 0.58%. However, adding only the world trade data that was available at the time barely changes the forecast accuracy.

This paper investigates whether available world trade data improves GDP growth forecasts in a more systematic way than in table 1[4]. If the available world trade data is useful for forecasting GDP then the Netherlands is a good candidate for a country where this is likely to show up. In the Netherlands CPB produces forecasts for a range of variables 4 times per year, including GDP growth for the current and the next year; we base our analysis of the usefulness of world trade data on this forecasting framework. The paper finds no compelling evidence that the available world trade data improve GDP growth forecasts over and above that provided by a small group of other predictors. In an extension we also show that world trade data also doesn't help forecast exports, the component of GDP where one would expect that it would be most useful. We then look at various possible explanations for why the available world trade data doesn't help and conclude that there isn't a simple linear relationship between the accuracy of world trade forecasts and GDP and exports forecasts.

As a robustness excercise we also produce forecasts for both GDP growth and export growth using the direct multi-period method instead of the iterative method (see Pesaran et al. (2011) and Marcellino et al. (2006) for recent examples and Jordà (2005) for its use for impulse responses). For GDP growth forecasts direct multi-period forecasts are no more accurate than the standard VAR approach. When using direct forecasts there is some evidence that the available world trade data helps produce more accurate forecasts for GDP growth two-years-ahead. However, those forecasts are marginally less accurate than the corresponding forecasts made using the standard VAR approach that exclude world trade data. For exports, although there is some evidence that the available world trade date helps improve forecasts within particular classes of models, the

growth.

[4] It is important to note that this paper is concerned with directly forecasting GDP growth from data available when the forecast is made. It is not an evaluation of a two-step forecasting process where the first step involves forecasting world trade and the second step uses this forecast to produce a forecast for GDP growth. For that, a different set of explanatory variables may produce accurate forecasts for future world trade than those considered here. If the set of explanatory variables considered here were good predictors of world trade then they should also be good predictors of GDP growth due to the contemporaneous correlation between the two already highlighted.

most accurate forecasts were again made excluding world trade data.

The rest of this paper is structured as follows. Section 2 outlines the pooled VAR approach we use, whilst section 3 shows that world trade growth does not improve the forecast accuracy for GDP growth. Section 4 extends the analysis of section 3 to exports, the most likely component of GDP where world trade would improve forecast accuracy. In section 5 we investigate why world trade data doesn't improve forecast accuracy and finally section 6 summarises and offers some conclusions.

2 Our approach

To produce our forecasts we use simple stationary VAR models. We use VAR models because VARs are simple to estimate and are widely used in practice (Elliott and Timmermann (2008)). Linear univariate autoregressions and VAR models have also performed well in various comparisons. For example Stock and Watson (1998) find that linear autoregressions perform better than nonlinear models for a wide range of US macroeconomic series. Boero (1990) finds that VAR models outperform structural equations models for Italy. In a forecasting comparison for Norway, Eitrheim et al. (1999) found that a first difference VAR could produce more accurate forecasts in some cases than the large macro model used by the central bank of Norway. Another recent comparison of VAR based forecasts and published forecasts based on large macro models is reported in Edge et al. (2006). They find that, for certain macro variables, VAR based forecasts outperform the published forecasts from the Federal Reserve. Another reason for chosing VAR models is that a large class of models can be well approximated by a VAR model (see, for example, Fernandez-Villaverde et al. (2007)). A VAR(p) is shown if equation (1)

$$x_{t+1} = a + A_1 x_t + ... + A_p x_{t-p+1} + \varepsilon_{t+1} \tag{1}$$

where x_t is a vector of endogenous variables, a is a vector of constants, the A_is are matrices of regression coefficients and ε_t is a vector of error terms. Of course, the accuracy of the forecast produced by this model depends on a number of factors. Firstly, the variables included in x_t. Adding new variables to x_t increases the available information, which could improve accuracy, but it also increases the number of parameters to be estimated, which may result in reduced forecast accuracy. Secondly, increasing the lag length also allows a richer approximation of the dynamics of the true system, but again at the cost of having to estimate more parameters. Finally, sheer luck will play a part in the accuracy of each individual model, especially with the short out-of-sample forecasting periods available for GDP forecasting competitions such as this. Our approach tries to minimise the effect of each of these factors so that we can get a better picture of the usefulness of each individual variable rather than the accuracy of a specific model.

A researcher interested in testing the importance of a particular variable would typically estimate two versions of Equation (1): one including the variable in question and one without. Then the researcher would compare the accuracy of the forecasts produced by the two models. However, the conclusion drawn by following this approach does not allow the researcher to separate the usefulness of the variable from the negative effect of reducing the degrees of freedom in the regression.

Our approach attempts to minimise these effects by pooling forecasts from many different VAR models. Hendry and Clements (2003) conclude that simple forecasting methods often do best and that pooling forecasts helps. For our benchmark forecasts we estimate all combinations of lag and variables up to four lags and six variables from a selection of ten additional variables. That means we have 4 univariate models, 40 bivariate models, 180 three-variable models, 480 with four variables, 840 with five variables and 1008 with six variables for a total of 2552 individual models. Our benchmark forecast is the simple unweighted mean of all of the forecasts pooled together. We then compare what happens when one variable is removed from the list of variables where we again compute forecasts from all combinations of lags and variables up to four lags and six variables. Excluding a variable from selection reduces the number of individual models to 1528 models[5]. The degrees of freedom problem highlighted above still plays a role because there are less large models when one of the variables is removed, but it is less severe than the standard approach because there are still a considerable number of six-variable models, rather than none as there would be in the standard approach.

We use an expanding window for producing the forecasts. We make one forecast per year for up to 8 quarters ahead based on data up to and including the fourth quarter of the previous year. The eight quarterly forecasts are transformed into forecasts for the yearly growth rates for the first year and the second year. This is equivalent to how CPB produces forecasts for our CEP publication every March. We then move on four quarters and make a new set of forecasts.

2.1 Direct multi-period forecasts

The traditional method of producing forecasts and impulse responses for VAR models involves simulating the estimated VAR model the requisite number of periods forward. For multistep forecasts all forecasts for horizons greater than one rely on the forecasts for the intervening periods. For example, equation (1) shows the one-step-ahead forecast from a VAR(p). The corresponding two-step-ahead forecast relative to the information known at the time it is made is

$$x_{t+2} = a + A_1 \left(a + A_1 x_t + ... + A_p x_{t-p+1} + \varepsilon_{t+1} \right) + A_2 x_t + ... + A_p x_{t-p+2} + \varepsilon_{t+1} \tag{2}$$

[5]There are again 4 univariate models, 36 bivariate, 144 three-variable models, 336 with four variables, 504 with five variables and 504 more with six variables.

If the estimated coefficients in the A_i matrices deviate from their true value (as a result of a limited sample size, for example) the effects of this error will accumulate each extra period the model is simulated. Furthermore, since all of the elements of the A_i matrices are required to make the implied 1-step-ahead forecast for use in the two-step-ahead forecast, any errors in the rest of the system will impact the accuracy of the GDP growth forecast. To get around this problem, we use direct multi-period forecasts (see Pesaran et al. (2011) and Marcellino et al. (2006) for recent examples, and Jordà (2005) for their application to impulse responses).The direct method bypasses the need to simulate the model forward by regressing the impulse response or forecast for a given period directly onto a number of lags. The approach involves estimating a separate equation for each forecast horizon i:

$$x_{t+i} = a_i + A_{1,i}x_t + ... + A_{p,i}x_{t-p+1} + \varepsilon_{t+i}. \tag{3}$$

This results in forecasts or impulse responses that are more robust to mis-specification. The downside is that there are less degrees of freedom available for the multistep forecast regressions. For example, the latest right-hand-side data in the eight-quarter ahead regression is from two years ago.

3 Forecasting GDP growth

We perform an out-of-sample forecasting competition for the years 1998 to 2007 and compare forecasts with models including each particular candidate predictor with those excluding it. This excerise is performed on final, revised data, not real-time data.

3.1 Data

Table 2 lists the series used for forecasting GDP growth. The data span the period 1977Q2 to 2007Q4, which is the maximum length of available data. The ten series were selected based upon their leading correlations with GDP growth in the period ending 1992[6]. The first six series listed in table 2, i.e. real GDP to bankruptcies, enter the models in either log-levels or growth rates. When they are used in growth rates the models are designated as 'growth rate' models; when they enter in log-levels they are designated 'levels' models. The remaining four series enter all models in levels. Rather than use all ten series in a single VAR, which would run into degrees of freedom problems, we select combinations of variables and lag lengths for smaller VARs. We then pool all of the forecasts from the smaller VARs into a final pooled forecast. Ten series with all possible combinations of model sizes from 1 to 6 variables and from lags 1 to 4 gives a total of 2552 model combinations over which we pool.

We stop our forecasting competition in 2007 to avoid the competition being dominated by the severe recession caused by the credit crisis of 2008-2009.

[6]These are the same series used in Elbourne et al. (2008).

Table 2: Quarterly data sources for GDP growth forecasts

Variable	Source
Real GDP	Statistics Netherlands
Relevant world trade	CPB
Real private consumption	Statistics Netherlands
Compensation per employee, market sector	Statistics Netherlands
CPI	Statistics Netherlands
Bankruptcies	Statistics Netherlands
3 month interest rates	DNB
Production expectations, manufacturing industry	Statistics Netherlands
Consumer confidence	Statistics Netherlands
German business climate (Ifo)	IFO
US 3 month interest rate	Datastream

3.2 Results

Table 3 shows the mean absolute error (MAE) of the pooled forecast of all combinations and for all combinations excluding a particular variable for forecasts from 1998 to 2007. To evaluate whether the difference in accuracy is statistically significant we use the test of unconditional predictive ability proposed by Giacomini and White (2006)[7]. We cannot use the more common Diebold and Mariano (1995) test since our comparison concerns nested models: excluding a variable from selection is equivalent to setting the weight to zero to all models containing that variable before making the pooled forecast from all variables.

For the models estimated in growth rates, including world trade data does not improve forecast accuracy at either forecast horizon. In fact, excluding world trade actually reduced the MAE marginally at both horizons (1-year-ahead was marginally lower but both round to 0.75). This becomes even more remarkable when one considers that the pooling excluding world trade is done over 1544 models rather than over 2552 for all variables. For 1 year ahead forecasts world trade only performed better than employee compensation and the business climate survey. The variable that contributes the most to forecast accuracy is the number of bankruptcies: we do not build macro models with such a prominent role for bankruptcies as we do for world trade.

In contrast, when using models estimated in levels, excluding world trade data lowered accuracy at both horizons: at the one-year horizon the difference is statistically significant. However, the levels based forecasts are marginally less accurate than the growth rate based forecasts. There is no evidence to conclude that you can do better than a growth rate based forecast excluding the available world trade data.

[7]Technically this test is only valid for rolling window forecasts, not the expanding window approach we take here. Nevertheless, the p-values give a useful, if not perfect, indication of the relative performance of two forecasting schemes.

Table 3: MAE of pooled VAR GDP forecasts from 1998 to 2007 (% points)

	Growth rates		Levels	
	1 year MAE	2 year MAE	1 year MAE	2 year MAE
All variables	0.75	1.26	0.76	1.27
Ex world trade	0.75	1.25	0.92*	1.31
	(0.72)	(0.29)	(0.07)	(0.60)
Ex consumption	0.76	1.27	0.67	1.29
	(0.72)	(0.48)	(0.23)	(0.66)
Ex CPI	0.73	1.24	0.74	1.22*
	(0.53)	(0.44)	(0.39)	(0.05)
Ex compensation	0.77	1.28	0.76	1.21*
	(0.43)	(0.35)	(0.77)	(0.03)
Ex bankruptcies	0.80*	1.30	0.80	1.31
	(0.07)	(0.34)	(0.24)	(0.45)
Ex short interest	0.70	1.27	0.73	1.38
	(0.45)	(0.97)	(0.42)	(0.16)
Ex production expectations	0.77	1.27	0.79	1.27
	(0.54)	(0.86)	(0.29)	(0.91)
Ex consumer confidence	0.78	1.24	0.76	1.28
	(0.59)	(0.61)	(0.94)	(0.80)
Ex Ifo	0.75	1.26	0.76	1.25
	(0.90)	(0.68)	(1.00)	(0.45)
Ex short US interest	0.79	1.28	0.79	1.29*
	(0.14)	(0.22)	(0.36)	(0.06)

Note: p-values for Giacomini and White (2006) test of equal predictive ability in parentheses.
* denotes a significant difference in forecast ability vis-a-vis all variables at the 10% level.

Table 4: MAE of pooled direct multi-period GDP forecasts from 1998 to 2007 (% points)

	Growth rates		Levels	
	1 year MAE	2 year MAE	1 year MAE	2 year MAE
All variables	0.71	1.28	0.92	1.27
Ex world trade	0.75	1.34	0.89	1.33*
	(0.58)	(0.39)	(0.22)	(0.05)
Ex consumption	0.67	1.29	0.89	1.58*
	(0.46)	(0.87)	(0.50)	(0.04)
Ex CPI	0.69	1.28	0.90	1.24
	(0.17)	(0.95)	(0.46)	(0.63)
Ex employee compensation	0.69*	1.27	0.87	1.23
	(0.08)	(0.62)	(0.23)	(0.25)
Ex bankruptcies	0.79*	1.29	1.05*	1.35*
	(0.08)	(0.89)	(0.05)	(0.09)
Ex short interest	0.70	1.16*	0.87	1.34
	(0.91)	(0.05)	(0.14)	(0.59)
Ex business climate survey	0.76	1.35	0.92	1.23
	(0.50)	(0.16)	(0.92)	(0.18)
Ex consumer confidence	0.64*	1.27	0.89	1.25
	(0.04)	(0.77)	(0.29)	(0.76)
Ex Ifo	0.67	1.31	0.90	1.29
	(0.19)	(0.83)	(0.70)	(0.49)
Ex short US interest	0.79	1.35	0.96	1.26
	(0.23)	(0.11)	(0.16)	(0.87)

Note: p-values for Giacomini and White (2006) test of equal predictive ability in parentheses.
* denotes a significant difference in forecast ability vis-a-vis all variables at the 10% level.

3.3 Robust multi-period forecasts

The direct multi-period forecasts were made with 4 lags and the regression equations used the latest available data. For example, the one-quarter-ahead forecast for 1998Q1 was based on feeding the data up to 1997Q4 into a model with regression coefficients from a regression of GDP growth up to 1997Q4 on lags up to 1997Q3. For the eight-quarter-ahead forecasts the forecast for 1999Q4 was made from data up to 1997Q4 being fed into a regression model with coefficients from a regression of GDP growth up to 1997Q4 on lags up to 1995Q4. The results of the direct forecasts can be seen in table 4.

Although the MAE of one-year-ahead forecasts using all models in growth rates is slightly better using direct multi-period forecasts rather than the standard VAR approach, the difference is not statistically significant (p-value = 0.77). Likewise, the lower accuracy of the direct one-year-ahead forecast from models in levels is also not statistically significant (p-value = 0.18). Therefore we can conclude that there is no convincing evidence that direct multi-period forecasts benefit from their supposed extra robustness.

As for the importance of the world trade data, at both one-year and two-year horizons for growth rate models, excluding world trade results in lower accuracy, although the difference with the benchmark is not statistically significant. Once again, when the models estimated in levels are considered, there is a statistically significant decrease in accuracy by excluding world trade trade data, this time at the two-year horizon. Set against this finding, though, it must be remarked that once again there is no compelling evidence that direct forecasts in levels are any more accurate than standard VAR forecasts from growth rate models excluding world trade data. Therefore it is hard to conclude that the available world trade data is vital for forecasting GDP growth when the accuracy of a forecast made after excluding it is statistically indistinguishable.

4 Forecasting exports

Since there was only weak evidence that world trade data might help forecast GDP, it is useful to look whether it can help forecast the components of GDP. We focus on exports because this component of GDP is a direct link with international trade. Therefore, this section repeats the excercise above for forecasting exports.

4.1 Data

The list of series used to forecast exports is shown in table 5. These series were selected on theoretical grounds as displaying a mix of goods market and capital market determinants of exports. Once again, the first six series listed in table 5, i.e. real exports to competitor prices, enter the models in either log-level or in growth rates. Once again this choice is distinguished by the labels 'growth rates' and 'levels'. The remaining four series enter all models in levels. As with the GDP growth forecasts the data starts in the second quarter of 1977 and the forecast competition period is from 1998 to 2007.

4.2 Results

Table 6 shows the accuracy of standard VAR approach forecasts for export growth. For 1-year-ahead forecasts, growth rate models have an MAE of 2.15 and are more accurate than levels models that had an MAE of 2.28. For 2-year-ahead forecasts the accuracy is much closer, but growth rate models have marginally lower MAEs than the levels models. When looking at the importance of individual variables the conclusions are mostly robust to looking at growth rate models or levels models, so we will focus mainly on the growth rate models, which had lower MAEs anyway.

If we exclude the available world trade data from our models there is no deteritoration in forecast accuracy: the 1-year-ahead MAE stays the same at 2.15. For the 2-year-ahead forecasts the accuracy does deteriorate slightly, from

Table 5: Quarterly data sources for export growth forecasts

Variable	Source
Real goods exports (excluding energy)	SAFFIER Database
Relevant world trade	CPB
Euro area real GDP	SAFFIER Database
US real GDP	US Department of Commerce
Price of Dutch exports	SAFFIER Database
Competitor prices	SAFFIER Database
Dollar/euro exchange rate	DNB
3 month German interest rate	SAFFIER Database
10 year German interest rate	Datastream
Long US interest	Datastream
3 month US interest rate	Datastream

2.26 to 2.28 and the deteriration is borderline statistically significant at the 90% level. A similar pattern holds true for the levels models, although from a higher base MAE. If we compare the relative importance of world trade with the other variables, we can see that excluding either euro area GDP or the dollar/euro exchange rate leads to significantly less accurate forecasts. When we exclude euro area GDP the MAE increases all the way to 2.41. In contrast, both short and long US interest rates appear to have a negative effect of the accuracy of the pooled forecasts. Excluding them invariably lowers the MAE.

So there is some evidence here that the available world trade data does help to improve 2-year-ahead forecasts, although only marginally.

4.3 Robust multi-period forecasts

Table 7 shows the accuracy of direct multi-period forecasts for export growth. As with the GDP growth forecasts, none of the direct forecasts are statistically significantly more accurate than their standard VAR approach counterparts. The p-values for the growth rate based models were 0.83 and 0.53 for one-year and two-year forecasts. For the levels based forecasts the p-values were 0.83 and 0.98. Despite not being statistically significant the 2-year-ahead MAE for growth rate models is quite a lot lower than those from the standard approach: the MAE for dirct forecasts is 2.83 compared to 3.26 previously.

If we look once more at the importance of the individual variables, we can conclude that excluding the available world trade data makes very little difference to forecast accuracy of both growth rate and levels models. The insignificance is highlighted again by comparing to euro area GDP and the dollar/euro exchange. Excluding euro area GDP would imply a significant deterioration in forecast accuracy for the 1-year-ahead forecast from both types of model. Forecast errors are also considerably higher when we exclude the dollar/euro exchange rate for both horizons; that is especially so for the growth rate models where the MAE rises by a statistically significant 0.7. Also, the conclusion that

Table 6: MAE of pooled VAR export forecasts from 1998 to 2007 (% points)

	Growth rates		Levels	
	1 year MAE	2 year MAE	1 year MAE	2 year MAE
All variables	2.15	3.26	2.28	3.30
Ex world trade	2.15	3.28*	2.30	3.41*
	(0.95)	(0.10)	(0.73)	(0.06)
Ex euro area GDP	2.41*	3.24	2.37	3.30
	(0.04)	(0.67)	(0.18)	(0.98)
Ex US GDP	2.14	3.27	2.32	3.35
	(0.86)	(0.59)	(0.71)	(0.71)
Ex price of Dutch exports	2.16	3.27	2.35	3.29
	(0.69)	(0.46)	(0.11)	(0.79)
Ex competitor prices	2.14	3.25	2.23	3.30
	(0.92)	(0.88)	(0.58)	(0.97)
Ex dollar/euro exchange rate	2.31*	3.28	2.32	3.34
	(0.09)	(0.84)	(0.66)	(0.72)
Ex long German interest	2.20	3.21	2.38*	3.47
	(0.59)	(0.81)	(0.04)	(0.11)
Ex short German interest	2.10	3.26	2.22	3.27
	(0.24)	(0.88)	(0.39)	(0.75)
Ex long US interest	2.13	3.19	2.32	3.29
	(0.62)	(0.38)	(0.17)	(0.14)
Ex short US interest	2.10*	3.22	2.27	3.24*
	(0.01)	(0.13)	(0.46)	(0.08)

Note: p-values for Giacomini and White (2006) test of equal predictive ability in parentheses.
* denotes a significant difference in forecast ability vis-a-vis all variables at the 10% level.

Table 7: MAE of pooled direct multi-period export forecasts from 1998 to 2007 (% points)

	Growth rates		Levels	
	1 year MAE	2 year MAE	1 year MAE	2 year MAE
All variables	2.27	2.83	2.19	3.29
Ex world trade	2.27	2.81	2.21	3.32
	(0.92)	(0.63)	(0.77)	(0.59)
Ex euro area GDP	2.65*	2.89	2.44*	3.23
	(0.03)	(0.39)	(0.09)	(0.61)
Ex US GDP	2.36	2.92	2.27	3.42*
	(0.34)	(0.28)	(0.31)	(0.00)
Ex price of Dutch exports	2.27	2.79	2.19	3.35
	(0.79)	(0.45)	(0.99)	(0.42)
Ex competitor prices	2.12	3.04	2.13	3.66*
	(0.32)	(0.27)	(0.66)	(0.01)
Ex dollar/euro exchange rate	2.90*	3.50*	2.35	3.40
	(0.01)	(0.04)	(0.25)	(0.61)
Ex long German interest	2.26	2.98*	2.33	3.51
	(0.81)	(0.04)	(0.32)	(0.25)
Ex short German interest	2.23	2.79	2.12	3.20
	(0.24)	(0.17)	(0.23)	(0.48)
Ex long US interest	2.27	2.71*	2.13	3.31
	(0.98)	(0.01)	(0.38)	(0.83)
Ex short US interest	2.19*	2.77	2.11*	3.25
	(0.07)	(0.15)	(0.07)	(0.59)

Note: p-values for Giacomini and White (2006) test of equal predictive ability in parentheses.
* denotes a significant difference in forecast ability vis-a-vis all variables at the 10% level.

we can safely ignore US interest rates in this forecasting excercise still holds since excluding improves forecast accuracy.

5 Why doesn't world trade data help?

In table 1 we showed that perfect knowledge of future world trade could improve GDP growth forecasts considerably. If the pooled VAR models in section 3 were producing perfectly accurate forecasts of world trade then we could expect a similar improvement in forecast accuracy for GDP growth as we saw in table 1. That wasn't the case, however. In this section we will examine why the available world trade data doesn't help. We first rule out two simple possible explanations before speculating what may actually be behind our results.

Suppose, for illustration, we have three variables, world trade (X_t), GDP (Y_t) and some other variable or variables (Z_t), and their data generating processes are given by the following:

$$Y_{t+1} = \alpha_X X_{t+1} + \alpha_Z Z_{t+1} \tag{4}$$

$$X_{t+1} = \beta_X X_t + e_{t+1} \tag{5}$$

$$Z_{t+1} = \rho_Z Z_t + \rho_X X_{t+1}. \tag{6}$$

If we know both X_{t+1} and Z_{t+1} we know Y_{t+1}. When we come to forecast time $t+1$ we only have information up to time t, which corresponds to the system after substituting out X_{t+1} and Z_{t+1}:

$$Y_{t+1} = (\alpha_X + \alpha_Z \rho_X)(\beta_X X_t + e_{t+1}) + \alpha_Z \rho_Z Z_t. \tag{7}$$

One reason why adding current and past world trade data wouldn't improve the accuracy of GDP growth forecasts is if world trade were a pure random walk. In that case world trade growth would be impossible to forecast and adding the available world trade data to the information set would add no information about future world trade growth. In the system above that corresponds to $\beta_X = 0$, in which case the X_t term disappears from equation (7).

Alternatively, suppose that $\rho_Z = 0$. In this case $Z_{t+1} = \rho_X X_{t+1}$ so if we add X_t to a regression of Y_{t+1} on Z_t we are adding no new information and it will not improve the forecasts[8]. This is true even if $\alpha_Z = 0$; that is, even if the only cause of Y_{t+1} is X_t because

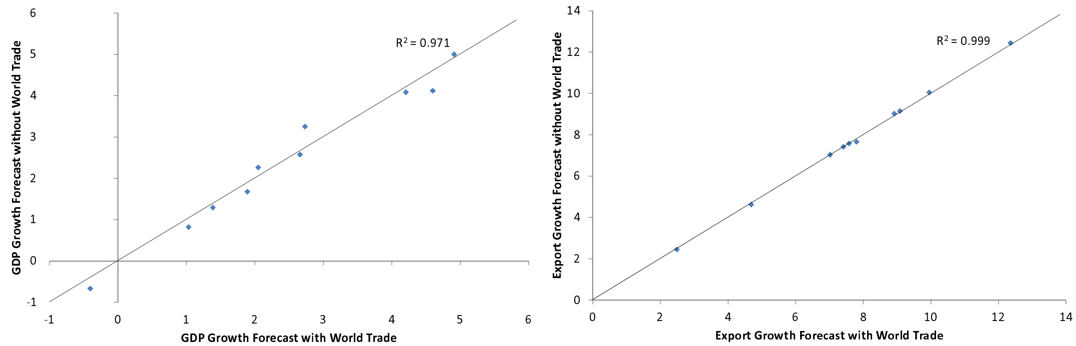
$$X_{t+1} = \frac{\beta_X}{\rho_X} Z_t + e_{t+1} \tag{8}$$

gives exactly the same forecast for X_{t+1} as does equation (5) above.

We now look at each of these two cases in the context of the forecasts in sections 3 and 4. In the system above, the usefulness of X_t for forecasting Y_{t+1} depends on how well it forecasts X_{t+1}. That depends on β_X and the variance

[8] We do not include an error term in equation (6) to get away from signal-to-noise issues.

Figure 2: Comparison of one-year-ahead forecasts with and without world trade: GDP (left) and exports (right)



of e_{t+1}. If $\beta_X = 0$ then X_{t+1} is just the mean plus noise and X_t adds no new information to what is already contained in the mean. If the variance of e_{t+1} is large then the information contained in X_t will be drowned out by the noise in e_{t+1}. Therefore, we compare a benchmark case where we only know the past mean of world trade growth with the forecasts from pooled VAR models for world trade. The MAE for the case where we use the past mean as our forecast was 2.6. If we use the set of variables previously used to forecast GDP growth (listed in table 2) to forecast world trade growth we get an MAE of 1.77. Likewise, if we repeat this exercise for the world trade variables in table 5 we get an MAE of 1.65. In other words, the set of variables used in this study explains about 35% of the one-year-ahead variation[9] in world trade growth. We could say that, in effect, we know about 35% of what there is to know about next years' world trade growth. Clearly, the variables used in this study contain useful information about future world trade growth and some part of future world trade growth is forecastable. The unhelpfulness of world trade data when forecasting GDP and exports is not due to world trade growth being unforecastable.

To evaluate whether the available world trade data is merely duplicating the information in the other series, we compare the direct multi-period forecast for world trade using all of the available series against forecasts excluding current and past world trade. If the world trade series is adding no new information to the system we would expect that the models excluding world trade should produce forecasts with MAEs for one-year-ahead world trade in the range 1.6-1.8 as above. However, when we try to forecast world trade without using current and past world trade data, the MAE rises to around 3. So clearly the information in the available world trade trade about future world trade is not duplicated well by the other series we have been using here.

Since we have shown that neither of these simple cases can account for the

[9] Of course, variation is not exactly the same thing as MAE, but we use the expression here as a short-hand for absolute uncertainty.

unhelpfulness of world trade data for forecasting GDP and exports, it may be instructive to look elsewhere for an explanation. Figure 2 plots one-year-ahead forecasts made including world trade against those made excluding world trade. The solid line is a 45-degree line and the R^2 is for the 45-degree line. For GDP growth the forecasts without world trade explain over 97% of the variation in the forecasts including world trade; for exports the figure rises to 99.9%. Clearly, adding world trade barely changes the resulting forecasts. What we can conclude from this is that the first 35% of knowledge about future world trade growth contains no extra information for forecasting GDP or export growth; the improvement in table 1 comes from the remaining 65%. In other words, there is a non-linear relationship between knowledge about future world trade and future GDP and exports growth: initial improvements in the accuracy of world trade forecasts does not improve the accuracy of GDP forecasts, only after some critical accuracy level does improved world trade forecasts improve GDP forecasts. Furthermore, it is also unclear where the critical level of knowledge about future world trade growth lies, would 50% knowledge be enough to improve forecasts for GDP and exports? All three of the series being discussed here are integrated series[10] and, as such, the intuition behind the Beveridge and Nelson (1981) decomposition and later multivariate versions thereof tells us that at least some part of any integrated series is an unforecastable pure random walk. It is possible that the improvement seen in table 1 arises solely from the unforecastable component of world trade, for example, because of future shocks affecting both series. In that case, the point where information about future world trade becomes useful for forecasting GDP and export growth already lies at an unobtainable level of accuracy for world trade forecasts.

6 Conclusion

This paper has used pooled VAR forecasts to evaluate the contribution of world trade growth data to forecast accuracy for GDP growth and export growth, the component of GDP most likely to be predicted by world trade. By using pooled VAR forecasts we have attempted to minimise a number of issues that make inference difficult regarding the importance of individual variables for forecast accuracy using traditional methods. We found no compelling evidence that the world trade data available at the time a forecast is made contributes to increased forecast accuracy for GDP growth over and above that provided by a relatively small group of explanatory variables. Furthermore, this paper has shown that world trade also doesn't improve forecast accuracy for export growth. The reason behind this result lies in a non-linear relationship between the accuracy of world trade forecasts and GDP or exports forecasts. Improving the forecast accuracy of world trade growth forecasts from an MAE of 2.6 to an MAE of 1.7 does not lead to corresponding improvements in the forecast accuracy for GDP or exports growth. However, somewhere between an MAE of 1.7 and perfect

[10]The augmented Dickey-Fuller tests for the null hypothesis of unit roots in GDP, exports and world trade have p-values of 0.90, 0.85 and 0.80, respectively.

foresight for world trade, increasing accuracy does start to improve GDP and exports forecasts.

As a robustness test we repeated the forecasting competition with direct multi-period forecasts. Despite the benefits claimed for direct forecasts the difference in accuracy when compared to the standard approach were statistically insignificant. This result is in line with the result found by Marcellino et al. (2006) who even found that iterated forecasts outperform the direct method especially as the horizon increases.

References

Beveridge, S. and C.R. Nelson, 1981, A new approach to decomposition of economic time series into permanent and transitory components with particular attention to measurement of the 'business cycle', *Journal of Monetary economics*, vol. 7, no. 2, pp. 151–174.

Boero, G., 1990, Comparing ex-ante forecasts from a SEM and VAR model: An application to the Italian economy, *Journal of Forecasting*, vol. 9, no. 1, pp. 13–24.

Diebold, F. and R. Mariano, 1995, Comparing predictive accuracy, *Journal of business and economic statistics*, vol. 13, pp. 253–263.

Don, F. and J. Verbruggen, 2006, Models and methods for economic policy: 60 years of evolution at CPB, *Statistica Neerlandica*, vol. 60, no. 2, pp. 145–170.

Edge, R.M., M.T. Kiley and J.P. Laforte, 2006, A comparison of forecast performance between federal reserve staff forecasts, simple reduced-form models, and a DSGE model, Tech. rep.

Eitrheim, O., T.A. Husebo and R. Nymoen, 1999, Equilibrium-correction vs. differencing in macroeconometric forecasting, *Economic Modelling*, vol. 16, no. 4, pp. 515–544.

Elbourne, A., H. Kranendonk, R. Luginbuhl, B. Smid and M. Vromans, 2008, Evaluating CPB's published GDP growth forecasts, CPB Documents 172, CPB Netherlands Bureau for Economic Policy Analysis.

Elliott, G. and A. Timmermann, 2008, Economic forecasting, *Journal of Economic Literature*, vol. 46, no. 1, pp. 3–56.

Fernandez-Villaverde, J., J. Rubio-Ramirez, T. Sargent and M. Watson, 2007, ABCs (and Ds) of understanding VARs, *The American Economic Review*, vol. 97, no. 3, pp. 1021–1026.

Giacomini, R. and H. White, 2006, Tests of conditional predictive ability, *Econometrica*, vol. 74, no. 6, pp. 1545–1578.

Hendry, D. and M. Clements, 2003, Economic forecasting: some lessons from recent research, *Economic Modelling*, vol. 20, no. 2, pp. 301–329.

Jordà, O., 2005, Estimation and inference of impulse responses by local projections, *American Economic Review*, vol. 95, no. 1, pp. 161–182.

Kranendonk, H. and J. Verbruggen, 2007, SAFFIER; a multi-purpose model of the Dutch economy for short-term and medium-term analyses, CPB Documents 144, CPB Netherlands Bureau for Economic Policy Analysis.

Kranendonk, H. and J. Verbruggen, 2010, SAFFIER II: 1 model voor de Nederlandse economie, in 2 hoedanigheden, voor 3 toepassingen, CPB Documents 217, CPB Netherlands Bureau for Economic Policy Analysis.

Marcellino, M., J.H. Stock and M.W. Watson, 2006, A comparison of direct and iterated multistep AR methods for forecasting macroeconomic time series, *Journal of Econometrics*, vol. 135, no. 1-2, pp. 499–526.

Pesaran, M.H., A. Pick and A. Timmermann, 2011, Variable selection, estimation and inference for multi-period forecasting problems, *Journal of Econometrics*, vol. 164, no. 1, pp. 173–187.

Stock, J.H. and M.W. Watson, 1998, A comparison of linear and nonlinear univariate models for forecasting macroeconomic time series, NBER Working Paper 6607.