CFCM

CENTRE FOR FINANCE, CREDIT AND MACROECONOMICS

Working Paper 14/15

The Importance of a Time-Varying Variance and Cross-Country Interactions in Forecast Models

Steven Trypsteen

Produced By:

Centre for Finance, Credit and Macroeconomics
School of Economics
Sir Clive Granger Building
University of Nottingham
University Park
Nottingham
NG7 2RD

Tel: +44(0) 115 951 4763
Fax: +44(0) 115 951 4159
suzanne.robey@nottingham.ac.uk

The Importance of a Time-Varying Variance and Cross-Country Interactions in Forecast Models*

Steven Trypsteen†

University of Nottingham

November 19, 2014

Abstract

This paper examines growth forecasts of models that allow for cross-country interactions and/or a time-varying variance plus feedback from volatility to growth. Allowing for these issues is done by augmenting an autoregressive model with cross-country weighted averages of growth and/or the GARCH-M framework. The models also allow for structural breaks in the mean and variance of growth. The obtained forecasts are then evaluated using statistical criteria, i.e. point and density forecasts, and an economic criterion, i.e. forecasting recessionary events. The results show that the two components are important to obtain improved point and density forecasts, but that forecasting recessionary events remains difficult.

*This article is based on a chapter of my PhD thesis. I would like to thank my advisor Kevin Lee for the guidance throughout my PhD. Financial support of the ESRC and School of Economics of the University of Nottingham is greatly acknowledged.

†Correspondence to: Steven Trypsteen, Sir Clive Granger building, University of Nottingham, University Park, NG7 2RD, Nottingham, United Kingdom.
Email: steven.trypsteen@nottingham.ac.uk

1 Introduction

Information on current and future economic activity is an important input in most economic decisions, ranging from investment decisions to the formulation of monetary policy. There is a large literature on how to obtain early estimates of economic activity discussing a wide range of approaches. One of the most popular tools is the use of bridge equations, which relate quarterly GDP growth with quarterly averages of a selection of monthly indicators[1]. Some observations of the monthly indicators, however, will not be available for a full quarter and so a quarterly average cannot be calculated. This is due to a publication lag and the fact that some months of the respective quarter lie in the future. Fortunately, these observations of the monthly indicators can be estimated, but an obvious drawback is that an inadequate model of the monthly indicators could increase the nowcast and forecast errors of the GDP output growth estimates. Thus, besides an adequate bridge equation, the bridge equation approach also requires a good forecasting model for the monthly indicators.

Commonly used models to extrapolate the monthly series used in the bridge equation approach is the random walk and the autoregressive model. Both these models, however, have two important drawbacks. *First*, these models do not take time-varying volatility into account. Time-varying volatility is potentially important to obtain good forecasts as volatility changes over the business cycle. Moreover, monthly indicators are generally more volatile than quarterly series. Monthly series for industrial production growth, for example, which is a crucial indicator to obtain good early estimates of GDP growth, is much more volatile than quarterly GDP growth. *Second*, the random walk and autoregressive model ignore cross-country

[1] Another popular approach is a basic statistical model where GDP growth depends linearly on its own lags. Mitchell (2009) shows that it is difficult to improve upon this type of model in periods of stable growth, but that its performance in more volatile periods is poor. More advanced methods are MIDAS models, factor models and dynamic factor models. See Banbura, Giannone, Modugno, and Riechlin (2013) for a review of the main approaches to obtain early estimates of economic activity.

interactions. As countries are linked though trade and financial links, forecasts could be improved by taking this into account.

In this paper, I investigate the forecast performance of models that take time-varying volatility and/or cross-country interactions into account. Comparing the forecasts obtained by a model that includes these two components and models where one of the two components, or both, is switched off, allows us to judge the relative importance of these components in forecasting performance.

A model that allows for time-varying volatility and cross-country interactions is easily obtained by augmenting a univariate autoregressive model with two components. First, I include the generalised autoregressive conditional heteroskedasticity (GARCH) framework, developed by R. F. Engle (1982), to accounts for a time-varying variance. In this framework, the variance of the residuals depends on past residuals and lags of the variance. Thus, a large shock in the previous period raises the volatility today. Second, I include cross-country weighted averages to accounts for cross-country interactions. Using cross-country weighted averages to allow for cross-country interactions is a technique developed in Pesaran, Schuermann, and Weiner (2004) and Dées, Di Mauro, Pesaran, and Smith (2007) in the context of vector autoregressive (VAR) modelling; the so-called Global VAR. It is also possible to allow for structural breaks in the mean and variance of the dependent variable and this is also done here.

In the context of forecast performance of bridge equations, an important monthly indicator is industrial production growth. Therefore, I investigate the forecast performance of the models in forecasting industrial production growth. As sample of countries I use the G7, which is a group of countries with highly developed economies[2] and arguably the most connected.

In the case of modelling industrial production, there is also empirical evidence that time-varying volatility and cross-country interactions are important. Trypsteen

[2]The G7 consist of Canada, France, Germany, Italy, Japan, the United Kingdom and the United States.

(2014) studies the importance of allowing for cross-country interactions in the relationship between volatility and growth. To that end, he estimates a GARCH-in-mean (GARCH-M) model, developed by R. Engle, Lilien, and Robins (1987) of industrial production growth augmented with cross-country weighted averages of growth to account for cross-country interactions. This model is the same as the one discussed above, except that it also allows for volatility to affect the mean, which in this case is industrial production growth. His results show that it is important to allow for a time-varying variance and cross-country interactions when modelling industrial production growth. Regarding the effect of volatility on industrial production growth, he finds evidence for a statistically significant effect of volatility on growth. It follows that, in the case of forecasting industrial production, allowing for this can also be important. Therefore, instead of using the GARCH model, I will use the GARCH-M model in the forecasting exercise.

The criteria used to evaluate the forecast performance of the various models are both statistical and economic and the benchmark model to which all models are compared to is the random walk model. As statistical criteria, I evaluate the forecast performance using point and density forecasts. Density forecasts take the complete distribution of the forecasts into account and so the spread of the density changes with volatility. As to judging forecast performance using economic criteria, I compare the ability of the various models to forecast recessionary events, which is defined as two consecutive negative growth realisations.

Four conclusions can be drawn from the forecasting exercise. *First*, the model without the GARCH-M framework and without the cross-country weighted averages is generally outperformed by a model with one or both of these components in terms of point and density forecasts. *Second*, the further in the future one wants to forecast, the more important the GARCH-M component becomes. *Third*, the GARCH-M framework is more important to obtain good density forecasts compared to good point forecasts. *Fourth*, with respect to forecasting recessionary events in the G7, the GARCH-M framework and/or the cross-country interactions do not improve the forecast performance much.

The paper is organised as follows. In Section 2, I describe the models used for forecasting G7 industrial production growth and the break detection procedure for the mean and variance of industrial production growth. In Section 3, I first describe how I obtain the forecasts, what criteria I use to judge the forecast performance and how to test if the forecast performance of some model is statistically better compared to the random walk model. Then, I discuss the forecast performance of the alternative models using the statistical and economic criteria. I also graphically illustrate the effect of the GARCH-M framework and the cross-country weighted averages on the forecasts of G7 industrial production growth. Section 4 concludes.

2 Modelling framework

The univariate GARCH-M model of industrial production (henceforth IP) growth augmented with cross-country weighted averages of growth and shift dummies in the mean and variance equation, denoted here as AR-Y*-GARCH-M, is given by

$$\Delta y_{it} = c_{i0} + \sum_{k=1}^{l} c_{ik} DM_{ikt} + \lambda_i \sigma_{it} + \tag{2.1}$$

$$\sum_{k=0}^{s} \beta_{ik} \Delta y^*_{it-k} + \sum_{k=1}^{p} \phi_{ik} \Delta y_{it-k} + \varepsilon_{it}$$

$$\sigma^2_{it} = \alpha_{i0} + \sum_{k=1}^{f} \alpha_{ik} DV_{ikt} + \eta_i \varepsilon^2_{it-1} + \gamma_i \sigma^2_{it-1} \tag{2.2}$$

where Δy_{it} is the growth rate for country i at time t, DM_{ikt} is a possible shift dummy k in the conditional mean which is equal to 0 before the break date and equal to 1 on and after the break date, σ_{it} is the conditional standard deviation of growth, which is the measure of volatility, Δy^*_{it-k} are the cross-country weighted averages of growth and ε_{it} is the error term.

The cross-country weighted averages of growth are defined as

$$\Delta y_{it}^{*} = \sum_{j=1}^{7} w_{ij} \Delta y_{jt}, \qquad \sum_{j=1}^{7} w_{ij} = 1 \qquad \text{and} \qquad w_{ii} = 0, \tag{2.3}$$

where I have chosen to measure w_{ij} by the average share of total trade of country i with country j. Total trade is defined as the sum of exports and imports between country i and the other countries. Note that the weight w_{ij} is assumed to be constant over time, but Trypsteen (2014) estimates the AR-Y*-GARCH-M model with and without time-varying weights and shows that the estimates do not change much.

The conditional variance of country i at time t, σ_{it}^2, depends on a constant α_{i0}, possible shift dummies, DV_{ikt}, which are equal to 0 before the break date and equal to 1 on and after the break date, the squared lagged error term and a lag of the variance. Note that (2.2) reduces to an ARCH(1) model if γ_i equals zero.

As usual, the conditional mean of the error term is assumed to be equal to 0, so $\varepsilon_{it} \sim (0, \sigma_{it}^2)$. Assuming that the cross-country weighted averages account for all the cross-country interactions, it follows that the error term of each country only consist of an idiosyncratic component and so the error term across countries is uncorrelated, i.e. $COV[\varepsilon_{it}, \varepsilon_{jt}] = E[\varepsilon_{it}\varepsilon_{jt}] = 0$ for $i \neq j$.

The AR-Y*-GARCH-M model contains three components, namely autoregressive terms (AR), cross-country weighted averages (Y*) and a GARCH-M framework (GARCH-M). The other models that are used to obtain forecasts are nested to this general model. To obtain these models one or more components of the AR-Y*-GARCH-M model is switched off. Comparing the forecasts obtained using these models allows us to assess the relative importance of the GARCH-M framework and the cross-country weighted averages of growth for forecasting output growth. Imposing the following restrictions on the AR-Y*-GARCH-M model in (2.2)-(2.3) gives the nested models:

- AR: Setting $\lambda_i = \beta_{ik} = \alpha_{i0} = \alpha_{ik} = \eta_i = \gamma_i = 0 \quad \forall\ i$ and k.

- AR-Y*: Setting $\lambda_i = \alpha_{i0} = \alpha_{ik} = \eta_i = \gamma_i = 0 \quad \forall\ i$ and k.

- AR-GARCH-M: Setting $\beta_{ik} = 0 \quad \forall\ i$ and k.

2.1 Break detection procedure

A well-documented property of post-war output growth data is the occurrence of structural breaks in both the mean, as demonstrated by the literature on productivity slowdowns (see for example Nordhaus (2004)) and the variance, as demonstrated by the literature on the great moderation (see for example Bernanke (2004)) and so it is important to account for these. It is also important to account for structural breaks from an econometric point of view. As Perron (1989) shows, structural breaks biases the autoregressive parameters in the mean equation towards 1. In GARCH model, it is also important to allow for breaks in the variance as ignoring structural breaks here biases the autoregressive parameters in the variance equation towards 1 (see Lamoureux and Lastrapes (1990) and Hillebrand (2005)).

The number of breaks and the break dates, however, are unknown in practice and so they have to be estimated using a statistical procedure. Once the potential breaks are found, it is possible to create shift dummies and to include these in the relevant model.

To find the number of breaks and the break dates for the IP growth rates of the G7 countries, I use a procedure developed by Bai and Perron (2003, p.15-16). To find breaks in the mean of a series x_t, they recommend to first regress x_t on a constant and accounting for potential serial correlation via non-parametric adjustment[3]. Based on this regression, the $UDmax$ and $WDmax$ test statistics are used to see if at least one break is present. If the $UDmax$ test statistic, the $WDmax$ test statistic or both

[3]The non-parametric adjustment is to estimate the model using a quadratic spectral kernel based HAC covariance estimation using prewhitened residuals. The kernel bandwidth is determined automatically using the Andrews AR(1) method.

are larger than the relevant critical value, then we can reject the null hypothesis of 0 breaks in favour of at least one break. If there is evidence of at least one break, then the Sup$F(\ell{+}1|\ell)$ statistic, where we test the null hypothesis of $\ell{+}1$ breaks given ℓ breaks, is used sequentially to determine higher order breaks where the maximum number of possible breaks is 5[4,5].

The approach to find breaks in the variance of a series x_t, following Herrera and Pesavento (2005) and Fang and Miller (2009), is to first regress x_t on a constant and on the mean break dummy or dummies found previously, saving the estimated residuals and then using the $UDmax$, $WDmax$ and Sup$F(\ell{+}1|\ell)$ statistics in the same way as discussed above based on the regression of the absolute values of the estimated residuals on a constant.

The AR-Y*-GARCH-M and AR-Y* models discussed in Subsection **??**, however, allow for cross-country interactions and so the break detection procedure should also allow for this. To that end, I do not apply the procedure directly on the IP growth rates, but on the growth rates where the effects of cross-country interactions are taken out. I find the filtered growth rates by regressing the output growth rate of country i on a constant and cross-country weighted averages of growth, i.e.

$$\Delta y_{it} = c_{i0} + \sum_{k=0}^{s} \beta_{ik} \Delta y^*_{it-k} + \xi_{it} \tag{2.4}$$

where ξ_{it} is the error term and the estimated residuals, $\hat{\xi}_{it}$, are the filtered IP growth rates . The inclusion of the contemporaneous and the number of lagged cross-country weighted averages is based on the Akaike Information Criterion (AIC) and I allow for up to six lags. I then calculate the $UDmax$, $WDmax$ and Sup$F(\ell{+}1|\ell)$ statistics based on the regression of the filtered output growth rates, i.e. the estimated

[4] This procedure allows for serial correlation and different variances of the residuals across regimes.

[5] Throughout the procedure I use a trimming of 0.15. As will be discussed later, I estimate the models with a fixed sample window of 516 observations. Thus a trimming of 0.15 implies that each segment has a minimum of 77 observations, i.e. 6.45 years.

residuals of equation (2.4), $\hat{\xi}_{it}$, on a constant, i.e.

$$\hat{\xi}_{it} = \alpha_{i0} + e_{it} \tag{2.5}$$

where e_{it} is the error term. Note that the above procedure does not allow for volatility to effect the growth rate and therefore it is possible for the procedure to interpret a change in volatility as a break. It is, however, better to find too many breaks here, and to include potentially insignificant break dummies in the models, than to miss breaks that should be included.

To find breaks in the variance of the IP growth rate of country i, I first regress Δy_{it} on a constant, the mean break dummy or dummies found previously and the cross-country weighted averages of growth to filter out the effects of cross-country interactions, i.e.

$$\Delta y_{it} = c_{i0} + \sum_{k=1}^{l} c_{ik} DM_{ikt} + \sum_{k=0}^{s} \beta_{ik} \Delta y^{*}_{it-k} + \mu_{it} \tag{2.6}$$

where μ_{it} is the error term[6]. Then, I regress the absolute values of the estimated residuals of equation (2.6), $\hat{\mu}_{it}$, on a constant, i.e.

$$|\hat{\mu}_{it}| = \alpha_{i0} + \nu_{it} \tag{2.7}$$

where ν_{it} is the error term and I calculate the $UDmax$, $WDmax$ and $\text{Sup}F(\ell{+}1|\ell)$ statistics based on this regression in order to find the number of breaks and the break dates for the variance of output growth of country i.

[6]Here too, the inclusion of the contemporaneous and the number of lagged cross-country weighted averages is based on the AIC and I allow for up to six lags

3 Forecast performance of alternative models

3.1 Data

The IP data for the G7 is from the OECD's Main Economic indicators (MEI) dataset. Figure 1 plots the G7 IP growth rates over the period 1962:01–2013:05[7]. Over the full sample period, the monthly growth rate of industrial production for the G7 are found to be stationary, serially correlated, conditional heteroscedastic and not normally distributed[8]. The data used to obtain the average trade weights used to calculate the cross-country weighted averages, w_{ij} in (2.3), is import and export data from the IMF's Direction of Trade Statistics.

3.2 Design of the out-of-sample forecast exercise and forecast evaluation

All forecasts of IP growth are based on a regression exercise[9] following a rolling scheme, starting with the estimation period of 1962:01-2004:12 and extending until all sample observations are exhausted. This implies that the estimation window is equal to 516 observation. Also, as the largest forecast horizon, h, is 12 and I have data up until 2013:05, this implies that the last rolling sample is 1969:06-2012:05 and so I have 90 forecasts. The trade weights used to calculate the cross-country interactions are also updated. The updating, however, only happens yearly as the data of the trade flows is yearly.

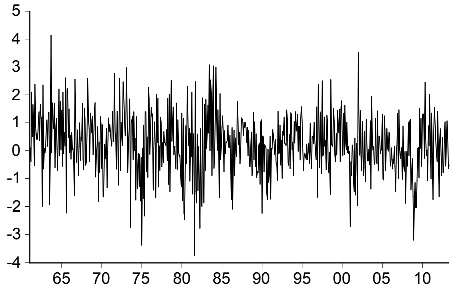
[7]The original data for France and Japan, however, has extreme observations and these are replaced by the median growth rate of the original data over the full sample. The extreme observations are March–April 1963 for France due to a miner strike, May–July 1968 for France due to the May '68 upraising and March–June 2011 for Japan due to an earthquake.

[8]Table 5 in the appendix presents all the results of the formal tests. Table 6 in the appendix presents the summary statistics of the output growth data over the full sample period.

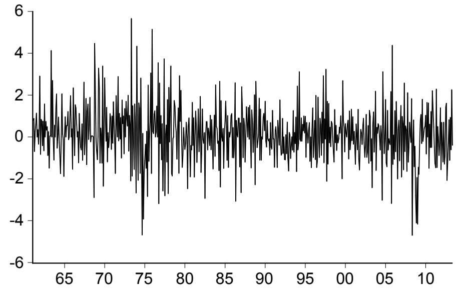
[9]The models are estimated with maximum likelihood using the Marquardt optimization algorithm. As with any iterative process, however, the algorithm could stop at a local maximum instead of the global maximum. To counter this problem, all the models are estimated with various initial values and choosing the initial value with the largest log likelihood. I use 10 different starting values. In particular, I estimate the models with the estimates of the OLS regression for the mean equation as starting values and various fractions of these OLS estimates. The fractions of the OLS estimates are 0.9, 0.8, . . ., 0.1.

Figure 1: Monthly growth rates of industrial production for the G7, 1961:02–2013:05

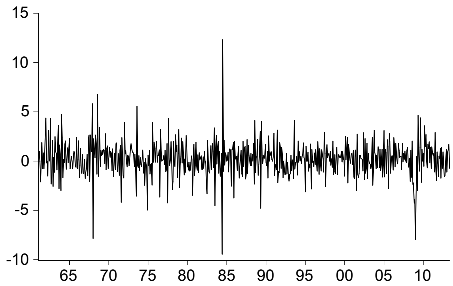
(a) Canada



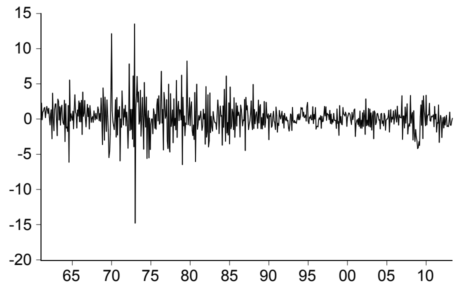
(b) France



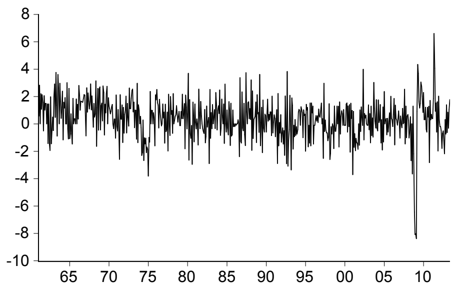
(c) Germany



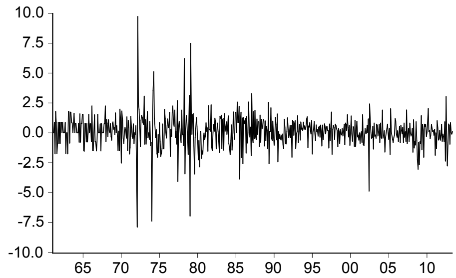
(d) Italy



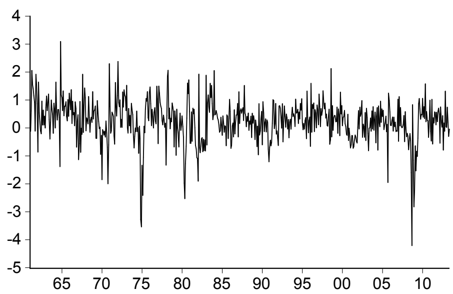
(e) Japan



(f) United Kingdom



(g) United States



The estimates of the four models are then used to simulate 20000 potential future output growth paths. I also simulate possible output growth paths using the random walk (RW) model as this will be the benchmark model. This simulation exercise gives the densities of the IP growth forecasts at horizon h. The mean of the density forecasts of the simulated IP growth paths at horizon h are then the point forecasts at horizon h.

The statistical criteria to evaluate the forecast performance of the models are point and density forecasts. The loss function for a point forecast is the quadratic loss function defined as $L(y_{t+h}, f_t^{(i)}) = (y_{t+h} - f_t^{(i)})^2$, where y_{t+h} is the realisation of IP growth at time $t+h$ and $f_t^{(i)}$ is the h step ahead point forecast of output growth of model i made at time t, where i = AR, AR-GARCH-M, AR-Y* or AR-Y*-GARCH-M. The average of $L(y_{t+h}, f_t^{(i)})$ for all forecasts is called the mean squared error (MSE) and the lower the better. The loss function for a density forecast is defined as $L(y_{t+h}, f_t^{(i)}) = log f_t^{(i)}(y_{t+h})$, where $f_t^{(i)}$ is now the probability density function of the forecast of model i made at time t. The average of $L(y_{t+h}, f_t^{(i)})$ for all forecasts is called the log score and the lower the log score, the better the density forecasts. This will always be a negative number as the logarithmic of a number in the unit interval is smaller than 0.

To judge the forecast performance of the various models using point and density forecasts, I apply the unconditional predictive ability test developed by Giacomini and White (2006)[10]. The first step is to regress the difference of the loss function of model i, $L_{t+h}^{(i)}(y_{t+h}, f_t^{(i)})$, and the loss function of the benchmark model, $L_{t+h}^{(b)}(y_{t+h}, f_t^{(b)})$, which in this case is the random walk model, on a constant, i.e.

[10]Other approaches to predictive ability testing are Diebold and Mariano (1995) and West (1996). Giacomini and White (2006) argue that their tests are a complement to these approaches, but also an extensions as their tests can be applied in all cases as Diebold and Mariano (1995) and West (1996) and in many more. The tests of West (1996), for example, are not applicable on nested models, whereas the Giacomini and White (2006) tests are. In addition, the Giacomini and White (2006) tests take estimation uncertainty into account, whereas the tests of Diebold and Mariano (1995) and West (1996) do not.

$$\Delta L_{t+h}^{(i)} = \mu + \varepsilon_{t+h}, \quad t = R, \ldots, T-h. \tag{3.1}$$

where $\Delta L_{t+h}^{(i)} = L_{t+h}^{(i)}(y_{t+h}, f_t^{(i)}) - L_{t+h}^{(b)}(y_{t+h}, f_t^{(b)})$.

The second step is to test if the constant, μ, is statistically different from 0. To account for the time series properties of ε_{t+h} and the dependence of ΔL_{t+h} in (3.1), I apply heteroskedasticity and autocorrelation consistent standard errors. For both the point and density forecasts, if μ is smaller than 0, then this implies that the forecast performance of the model is better than the benchmark. Thus in both cases we want to test the null hypothesis of $H_0 : \mu = 0$ against the alternative of $H_1 : \mu < 0$. For the point forecasts, the model with the lowest MSE where we can reject the relevant null hypothesis is chosen as the model with the best point forecast performance. For the density forecasts, the model with the lowest log score where we can reject the relevant null hypothesis is chosen as the model with the best density forecast performance.

As to evaluate the forecast performance using economic criteria, I compare the ability of the various models to forecast recessionary events where recession is defined as two consecutive negative growth realisations. The recessionary event is predicted to occur if the forecast probability exceeds 0.5. Comparing the forecasts of the event with the actual data gives four possible outcomes. Table 1 presents all these possible outcomes in a standard contingency table.

To judge the forecast performance of the various models, I calculate the hit rate and the Kuipers score for each model i using the contingency table. The hit rate is

Table 1: Contingency table

	Event occurred	Event did not occur	Total
Event is forecasted	a (hit)	b (false alarm)	$a+b$
Event is not forecasted	c (miss)	d (correct rejection)	$c+d$
Total	$a+c$	$b+d$	

defined as the ratio between the number of correctly forecasted recessions and the total number of recessionary events, i.e. $a/(a+c)$. The Kuipers score is defined as the difference between the hit rate and the false alarm rate, where the false alarm rate is the ratio between the number of falsely forecasted recessions and the total number of non-recessionary events, i.e. $b/(b+d)$.

I also test if there is a relationship between outcome and prediction. To that end perform two tests described in Pesaran and Timmermann (2009). The first test is the 'reduced rank regression' test. This is a standard contingency table test where the null hypothesis states that the model is no more successful in predicting the event than would be achieved using the unconditional probabilities. The second test is the 'dynamically-augmented reduced rank regression' test. This test takes predictable runs in the data into account.

3.3 Model selection issues

In this subsection, I discuss the specification search to find the most parsimonious models of the ones discussed in Subsection **??** and the various issues that complicate this process. The specification search to find the most parsimonious model for the AR and the AR-Y* model is relatively simple. In particular, I estimate the models with up to six lags of the relevant lagged terms in the respective models. The model that minimises the AIC is then picked as the best representation of the data. I also check if there is no serial correlation left in the residuals and that the mean and variance of the standardized residuals is equal to 0 and 1, respectively. As these two models do not allow for a time-varying variance, the residuals are not conditional homoscedastic.

Adding the GARCH-M model to the AR and/or AR-Y* model, however, complicates the specification search. I apply a specification search in the same spirit as the one discussed above, but need to take a few more issues into account. A first issue is that the Hessian of the log-likelihood function of the GARCH-M model is not block diagonal and so the mean and variance parameters are correlated. Therefore, all parameters need to be estimated simultaneously. A second issue is that nonnormal errors lead to inconsistent standard errors[11]. As for most G7 countries the output growth rates are close to being t-distributed, I allow the error distribution to be the t-distribution. I also check if the model that minimises the AIC is well specified. As with the AR and AR-Y* model, I check if there is no serial correlation left in the residuals and that the mean and variance of the standardized residuals is equal to 0 and 1, respectively. I now also check if the errors are conditional homoscedastic and if the distribution of the errors corresponds to the one assumed in the estimation process.

A final, but important, issue with GARCH models is that the Zero-Information-Limit Condition (ZILC) can hold as highlighted by Nelson and Startz (2007). They noted that in many econometric models the asymptotic variance of a parameter estimate (say, $\hat{\theta}$) depends on the value of another structural parameter (say, υ) and that for values of the structural parameter in a particular range the asymptotic variance of the variable of interest is very large and the model is weakly identified. More formally, Nelson and Startz (2007, p.49) argue that ZILC holds for an estimator $\hat{\theta}$ if there is a value of υ, say υ_0, such that $\lim_{\upsilon \to \upsilon_0} I_{\hat{\theta}} = 0$, where $I_{\hat{\theta}}$ is the inverse of the variance of $\hat{\theta}$. Nelson and Startz (2007) introduce ZILC as a way to identify such models where the above leads to spurious inference.

Ma, Nelson, and Startz (2007) show that ZILC can hold for GARCH models. They show that if the true ARCH effect, η_i in (2.2), is small, then the GARCH(1,1) model is weakly identified. The effect of this is that the GARCH coefficient, γ_i in (2.2), is biased upward and the corresponding standard error is too small. Thus the results point to persistence of the conditional variance where in fact this is not the case. Ma, Nelson, and Startz (2007, p16-17) also propose a procedure to detect

[11] Previous studies remedied this by estimating GARCH models with the normal distribution but applying the consistent variance-covariance estimator developed by Bollerslev and Wooldridge (1992).

ZILC. To check for ZILC in the GARCH(1,1) model, the implied autocorrelation function of the conditional variance from the GARCH(1,1) should be compared with the one implied by an ARCH(q) model. If they differ a lot, then this is evidence that ZILC holds. Alternatively, if the estimated conditional variance of the GARCH(1,1) and an ARCH(q) models are very different, then this also gives evidence for ZILC. Ma, Nelson, and Startz (2007) propose to model the variance as an ARCH model if ZILC is detected and so in the case of ZILC, I replace equation (2.2) with

$$\sigma_{it}^2 = \alpha_{i0} + \sum_{k=1}^{f} \alpha_{ik} DV_{ikt} + \eta_i \varepsilon_{it-1}^2. \tag{3.2}$$

To sum up, for the AR-GARCH-M and the AR-Y*-GARCH-M models, I pick the well-specified model that minimses the AIC and where ZILC does not hold[12].

3.4 Forecast performance using statistical criteria: Point and density forecasts

Before discussing the forecast performance of the alternative models using statistical and economic criteria, it is useful to illustrate the effect of the GARCH-M framework and cross-country weighted averages of growth on the forecasts. As the GARCH-M framework allows for a time-varying variance, the variance of the density forecasts also changes over time and so a period of increased volatility should widen the density forecasts. A model with cross-country interactions takes the effects of other countries into account and so the dynamics captured by such a model should be richer compared to a model that does not allow for these cross-country effects. I will illustrate these effects with point and density forecasts around the time of the bankruptcy of Lehman Brothers in September 2008, the fourth largest investment bank in the United States at the time. This episode is a good vehicle to illustrate

[12]For some sample periods, the AR-GARCH-M and/or AR-Y*-GARCH-M model picked by the model selection procedure was misspecified and/or there was evidence for ZILC. If this was the case, then I adjusted the model accordingly. Table 7 in the appendix documents these adjustments.

these effects as the collapse lead to great volatility and spilled over to other countries.

Figure 2 illustrates the effect of allowing for a time-varying variance. It presents the density forecasts of the standard AR model (Panel A) and of the AR model augmented with the GARCH-M framework (Panel B) for the United States around the time of the Lehman collapse[13]. The density forecasts presented in Figure 2 are the 1-step ahead forecast made in August, September and October of 2008. The figure shows that for the AR model the width of the density forecast does not change during the Lehman episode. For the AR-GARCH-M model, in contrast, the widening of the density forecasts, due to the increased volatility, is apparent.

To illustrate the effect of allowing for cross-country interactions, it is useful to compare the point forecasts of the AR model with the ones obtained from an AR model augmented with the cross-country weighted averages. Figure 3 presents the point forecasts of the AR (Panel A) and the AR-Y^* model (Panel B) made on February 2009, a few months after the Lehman collapse, up until 24 months in the future for all the G7 countries. This shows that the point forecasts converge to the long run growth rate predicted by the respective model, but that, for most countries, it takes much more time for the AR-Y^* than the AR model to reach that point. Indeed, the inclusion of the cross-country weighted averages allows the model to capture richer dynamics.

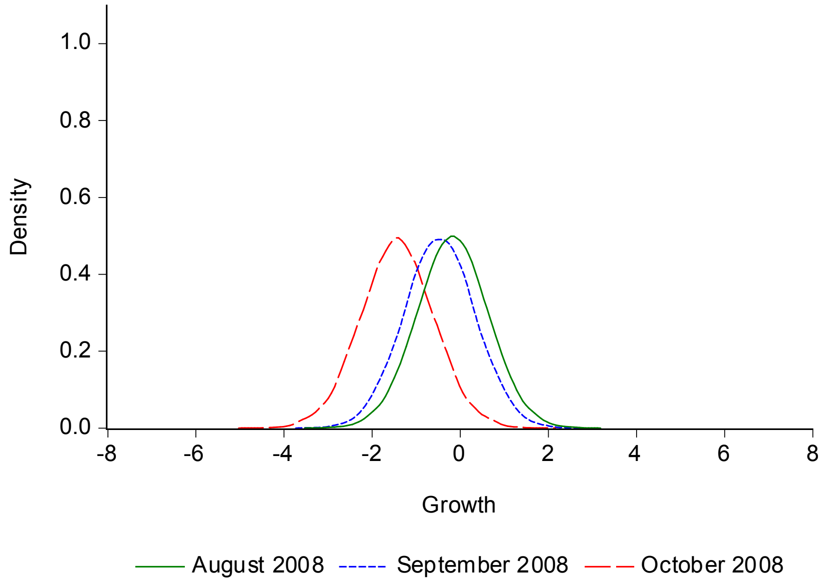
The above discussion illustrates the benefits of including the GARCH-M framework and cross-country weighted averages of growth to a model that is used to forecast G7 IP growth. In what follows I will formally investigate if these components also help improve the forecast performance in terms of point and density forecasts.

Table 2 shows the results of the point forecasts of the 1, 3 and 12-step ahead horizon in panel A, B and C, respectively. The numbers in italics are the average mean squared error (MSE) of the RW model. For the other models, the numbers

[13] The densities shown in Figure 2 are Kernel estimates of the true density forecasts.

Figure 2: 1-step ahead density forecasts of the AR and the AR-GARCH-M model made around the Lehman collapse for the United States

(a) AR



(b) AR-GARCH-M

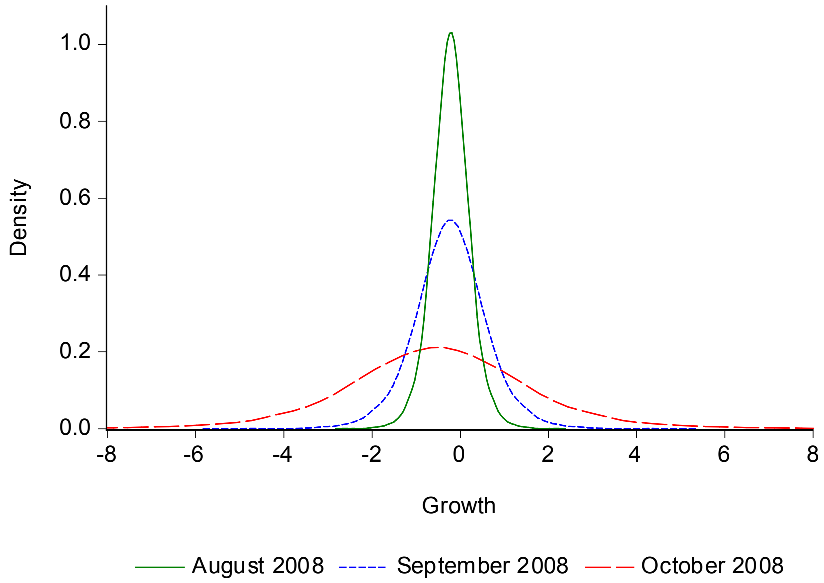
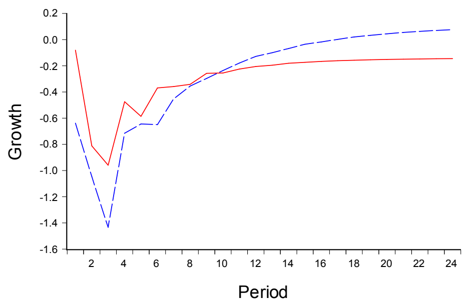
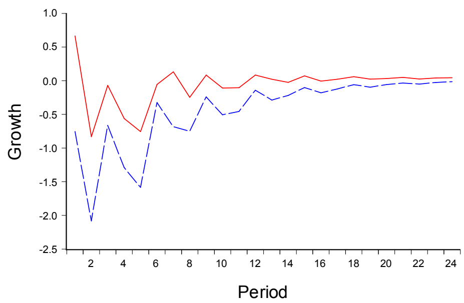


Figure 3: *h*-step ahead point forecasts made on February 2009 of the AR (solid) and the AR-Y* (dashed) model

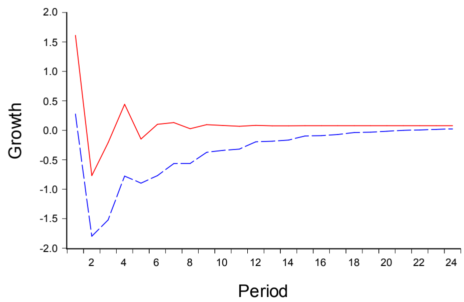
(a) Canada



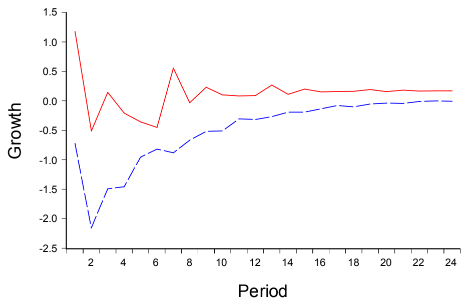
(b) France



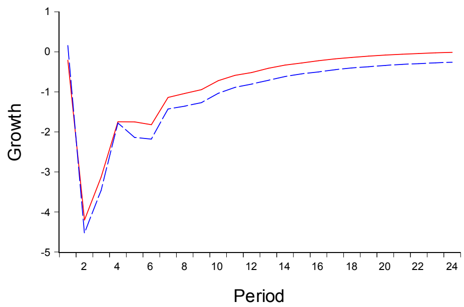
(c) Germany



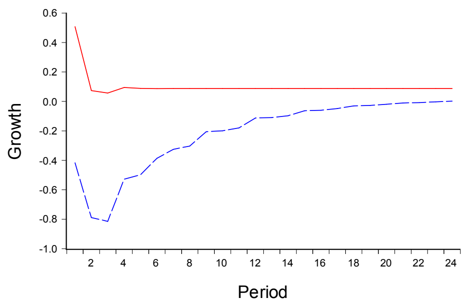
(d) Italy



(e) Japan



(f) United Kingdom



(g) United States

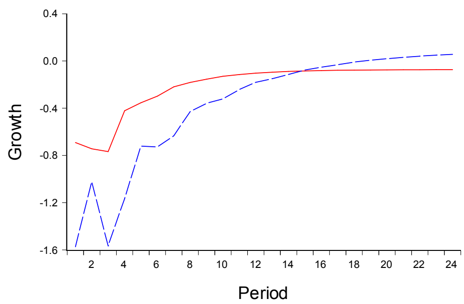


Table 2: Evaluation of point forecasts of IP growth

Panel A: MSE for 1-step ahead forecasts

	RW	AR	AR-GARCH-M	AR-Y*	AR-Y*-GARCH-M
CAN	*1.671*	0.613***	0.612***	0.537***	**0.532*****
FR	*5.459*	0.408***	0.454**	0.348***	**0.345*****
GER	*6.381*	0.656	0.622*	**0.560****	0.571**
ITA	*5.218*	0.658	0.640*	**0.476*****	0.495***
JAP	*4.954*	1.239	1.311	1.287	1.432
UK	*1.958*	0.548***	0.534***	0.475***	**0.461*****
US	*1.018*	0.598**	0.589**	0.560**	**0.509****

Panel B: MSE for 3-step ahead forecasts

	RW	AR	AR-GARCH-M	AR-Y*	AR-Y*-GARCH-M
CAN	*1.465*	0.662***	0.662***	0.597***	**0.591*****
FR	*3.751*	0.598***	**0.594*****	0.644**	0.608***
GER	*4.992*	0.673**	**0.652****	0.698***	0.657***
ITA	*3.916*	0.760	0.699**	0.740*	**0.698****
JAP	*9.118*	0.695**	**0.601***	0.692*	0.607*
UK	*1.925*	0.574***	0.567***	0.578***	**0.565*****
US	*0.951*	0.668***	0.676***	0.675***	**0.638*****

Panel C: MSE for 12-step ahead forecasts

	RW	AR	AR-GARCH-M	AR-Y*	AR-Y*-GARCH-M
CAN	*2.377*	0.461***	**0.453*****	0.491***	0.459***
FR	*4.913*	0.428***	**0.428*****	0.450***	0.439***
GER	*7.447*	0.440***	**0.439*****	0.468***	0.459***
ITA	*6.822*	0.440***	**0.419*****	0.443***	0.429***
JAP	*12.520*	0.443***	**0.429****	0.455**	0.442**
UK	*2.226*	0.502***	**0.495*****	0.511***	0.498***
US	*1.669*	0.444**	**0.432****	0.481**	0.456**

Notes:
- For the RW, the numbers represent the average mean squared error. For the other models, the numbers represent the ratio of their average MSE relative to the average MSE of the RW. The bold numbers are the lowest relative average MSE for each country.
- $*$, $**$ and $***$ denote that the forecasts are statistically better than that from the RW model at the 10%, 5% and 1%, respectively, using the Giacomini and White (2006) test of equal forecast performance and the quadratic loss function.

represent the ratio of their MSE relative to the MSE of the RW model. Thus, a ratio less than 1 implies that the forecast performance of the model is better than the RW model. The *, ** and *** denote if the improvement in forecast performance of a particular model compared to the RW is statistically significant at the 10%, 5% and 1% level according to the unconditional Giacomini and White (2006) test. The model with the lowest ratio and that is statistically better than the RW model is the best model in terms of MSE. The ratio of these models are in bold.

The results for the 1-period ahead point forecasts of IP growth in panel A show that the ratio is less than 1 for all models and countries, except for Japan. For Japan, no model is better than the RW model as the ratio is always larger than 1. The table also shows that for all countries except Japan the model with the best forecast performance include the cross-country weighted averages. For Germany and Italy, the AR-Y^* model is the best one. Adding the GARCH-M framework to the AR-Y^* improves the forecast performance further in four countries, namely Canada, France, the United Kingdom and the United States. The unconditional Giacomini and White (2006) test shows that the forecast performance of these best models is also statistically better than the RW model at the 5% significance level.

For the 3-period ahead point forecasts, the ratio's of all models are lower than 1 for all countries. Compared to the results concerning the best models for the 1-period ahead forecasts, for the 3-period ahead forecasts the GARCH-M framework becomes more important and the cross-country weighted averages become less important. Now, for all countries the best model includes the GARCH-M framework. Adding the cross-country weighted averages of growth to the AR-GARCH-M model improves the forecast performance further for Canada, Italy, the United Kingdom and the United States. The unconditional Giacomini and White (2006) test shows that the improvement in forecast performance of these best models is statistically better than the RW model at conventional levels.

The results for the 12-period ahead forecast, shown in panel C, show that cross-country interactions are now unimportant. For all countries, the AR-GARCH-M model provides the best output growth point forecasts. The forecast performance of these best models is also statistically better than the RW model at the 5% level according to the unconditional Giacomini and White (2006) test.

Table 3 shows the results of forecast performance using density forecasts. Panel A, B and C show the results for the 1, 3 and 12-step ahead forecasts, respectively. The numbers in italics are the average log score of the RW model. The numbers for the other models are the difference between the log score of the RW model and the log score of the model, scaled by the log score of the RW model. Thus, positive numbers imply that the model outperforms the RW model. The *, ** and *** denote that the forecast performance of the model is statistically better than the RW model at the 10%, 5% and 1% level according to the unconditional Giacomini and White (2006) test. The best forecasting model is the model with the highest ratio for each country and where the forecast performance is statistically better than the RW model. The ratio's of these models are in bold.

The results for the 1-period ahead density forecasts show that the scaled ratio is positive for all models and for all countries. The best models include the GARCH-M framework for all countries, except for Canada where the best model is the AR-Y^*. Adding cross-country weighted averages to the AR-GARCH-M model improves the forecasting performance further for Japan, the United Kingdom and the United States. The forecast performance of these best models is also statistically better than the RW model at the 1% level according to the unconditional Giacomini and White (2006) test. Comparing the best models for the 1-period ahead density forecasts with the best models for the 1-period ahead point forecasts, shows that the GARCH-M framework in more prevalent for the density forecasts. For only four countries the best models for the 1-step ahead point forecasts include the GARCH-M framework, whereas for the 1-step ahead density forecasts this framework is included for six countries.

Table 3: Evaluation of density forecasts of IP growth

Panel A: Log score for 1-step ahead forecasts

	RW	AR	AR-GARCH-M	AR-Y*	AR-Y*-GARCH-M
CAN	*-2.796*	0.390***	0.355***	**0.415*****	0.385***
FR	*-3.015*	0.382***	**0.429*****	0.408***	0.408***
GER	*-2.913*	0.348***	**0.475*****	0.371***	0.463***
ITA	*-2.721*	0.322***	**0.385*****	0.366***	**0.385*****
JAP	*-2.940*	0.334***	0.407***	0.334***	**0.415*****
UK	*-2.714*	0.388***	0.414***	0.403***	**0.437*****
US	*-2.790*	0.403***	0.436***	0.418***	**0.448*****

Panel B: Log score for 3-step ahead forecasts

	RW	AR	AR-GARCH-M	AR-Y*	AR-Y*-GARCH-M
CAN	*-2.576*	0.341***	0.325***	**0.368*****	0.347***
FR	*-2.658*	0.308***	**0.360*****	0.326***	0.330***
GER	*-2.620*	0.300***	0.470***	0.319***	**0.473*****
ITA	*-2.556*	0.305***	**0.384*****	0.323***	0.370***
JAP	*-2.710*	0.269***	**0.376*****	0.282***	0.374***
UK	*-2.582*	0.357***	0.411***	0.365***	**0.416*****
US	*-2.631*	0.368***	0.415***	0.380***	**0.442*****

Panel C: Log score for 12-step ahead forecasts

	RW	AR	AR-GARCH-M	AR-Y*	AR-Y*-GARCH-M
CAN	*-2.577*	0.303***	0.327***	**0.361*****	0.345***
FR	*-2.600*	0.284***	**0.372*****	0.309***	0.351***
GER	*-2.576*	0.220***	0.464***	0.257***	**0.468*****
ITA	*-2.566*	0.297***	**0.391*****	0.315***	0.389***
JAP	*-2.684*	0.162***	**0.382*****	0.201***	**0.382*****
UK	*-2.553*	0.334***	**0.434*****	0.343***	0.426***
US	*-2.586*	0.349***	0.392***	0.370***	**0.413*****

Notes:

- For the RW, the numbers represent the average log score. For the other models, the numbers represent the difference between the average log scores of the RW and the average log score of the other model the scaled by the average log score of the RW. The bold numbers are the highest scaled difference for each country.
- $*$, $**$ and $***$ denote that the forecasts are statistically better than that from the RW model at the 10%, 5% and 1%, respectively, using the Giacomini and White (2006) test of equal forecast performance.

The scaled ratio's for the 3-step and 12-step ahead density forecasts are also positive for all models and for all countries. Comparing the best models for the 3-step and 12-step ahead density forecasts with the best models for the 1-step ahead density forecasts shows that these do not change much. For Canada, France, Italy and the United States the best model is the same for all three forecast horizons. The best model for Germany is now the AR-Y^*-GARCH-M model compared to the AR-GARCH-M for the 1-step ahead density forecasts. Japan and the United Kingdom make the switch in the opposite direction. The unconditional Giacomini and White (2006) test shows that the forecast performance of the best models for the 3-step and 12-step ahead density forecast is also statistically better than the RW model at the 1% level.

3.5 Forecast performance using an economic criterion: Forecasting recessionary events

Table 4 shows the results of the models' performance of forecasting recessionary events. Panel A presents the percentage of actual occurrences of two consecutive negative growth realisations and the hit rates and panel B presents the Kuipers score together with the outcome of the reduced rank regression and the dynamically-augmented reduced rank regression tests.

Panel A shows that the recessionary event was the least frequent in Germany with 11% of the time, whereas Canada has the highest occurrence of 28%. Panel A also shows that the model with the highest hit rate, in bold, is the RW model for all countries

The hit rate, however, does not take the number of times a model falsely predicts the event into account. Panel B, therefore, presents the Kuiper score, which is defined as the difference between the hit rate and the false alarm rate. The bold numbers are the highest Kuipers score per country. This shows that the Kuipers scores for the RW model are much lower than the hit rates, implying that the

Table 4: Evaluation of probability forecasts: Forecasting recessionary events

Panel A: Actual probability and hit rate

	Actual	RW	AR	AR-GARCH-M	AR-Y*	AR-Y*-GARCH-M
CAN	0.28%	**0.44**	0.12	0.12	0.20	0.20
FR	0.21%	**0.37**	0.00	0.00	0.21	0.16
GER	0.11%	**0.40**	0.00	0.00	0.00	0.00
ITA	0.18%	**0.63**	0.00	0.00	0.13	0.13
JAP	0.12%	**0.36**	0.18	0.09	0.18	0.18
UK	0.24%	**0.41**	0.00	0.00	0.05	0.05
US	0.20%	**0.61**	0.39	0.44	0.33	0.39

Panel B: Kuipers Score

	RW	AR	AR-GARCH-M	AR-Y*	AR-Y*-GARCH-M
CAN	0.15	0.07	0.07	**0.20**	0.18
	(-,-)	(-,-)	(-,-)	(***,**)	(***,-)
FR	-0.01	-0.04	0.00	**0.18**	0.13
	(-,-)	(-,-)	(-,-)	(***,-)	(**,-)
GER	**0.15**	-0.03	-0.03	-0.03	-0.03
	(-,-)	(-,-)	(-,-)	(-,-)	(-,-)
ITA	**0.41**	-0.03	0.00	0.08	0.08
	(***,-)	(-,-)	(-,-)	(-,-)	(-,-)
JAP	**0.15**	0.14	0.05	0.14	0.14
	(-,-)	(*,*)	(-,-)	(*,-)	(*,-)
UK	**0.09**	0.00	0.00	0.02	0.00
	(-,-)	(-,-)	(-,-)	(-,-)	(-,-)
US	**0.46**	0.35	0.40	0.32	0.38
	(***,-)	(***,-)	(***,-)	(***,-)	(***,**)

Notes: - In panel B, the outcome of the reduced rank regression and the dynamically-augmented reduced rank regression test are respectively presented in the round brackets. '-' denotes that we cannot reject the null hypothesis at the 10% level and *,** and *** denote that we can reject the null hypothesis at the 10%, 5% and 1% level, respectively.

RW model has a high false alarm rate. In terms of Kuipers score, the RW model, however, still is the best model for Germany, Italy, Japan, the United Kingdom and the United States. For Canada and France the AR-Y* model is the best model.

The brackets under the Kuipers scores give the outcome of the reduced rank regression and the dynamically-augmented reduced rank regression tests, respectively. As discussed above, the null hypothesis of these tests is that there is no relationship between outcome and prediction. The reduced rank regression test assumes that there is no autocorrelation in the data, whereas the dynamically-augmented reduced rank regression test does take this into account. '-' denotes that we cannot reject the null hypothesis at the 10% level and *,** and *** denote that we can reject the null hypothesis at the 10%, 5% and 1% level, respectively.

According to the reduced rank regression test, we can reject the null hypothesis of no relationship between outcome and prediction for the models with the highest Kuipers score which is not the RW model for Canada and France at the 1% level. Using the dynamically-augmented reduced rank regression test we can reject the null hypothesis for Canada at the 5% level.

4 Conclusion

This paper examined the importance of allowing for a time-varying variance, feedback from the conditional standard deviation to growth and cross-country interaction for forecasting G7 IP growth by evaluating forecasts of the AR-GARCH-M-Y*, which incorporates all components, and nested versions of it where one of these components, or both, is switched off. The forecast performance of various models was done using statistical and economic criteria. The statistical criteria were point and density forecasts and the economic criterion was the ability of the models to forecast recessionary events.

As shown in the paper, there are important benefits to including the GARCH-

M framework and cross-country weighted averages of growth to a model used to forecast G7 output growth. The benefit of the GARCH-M framework is apparent when considering density forecasts. As the volatility of IP growth is time-varying, the density forecast widens with higher volatility. The inclusion of the cross-country weighted averages of growth allows the model to capture a richer dynamic structure compared to a model without these averages.

The forecasts exercise shows that the model without the GARCH-M framework and without the cross-country weighted averages is generally outperformed by a model with one or both of these components. Also, the further in the future one wants to forecast, the more important the GARCH-M component becomes. The forecast exercise also shows that the GARCH-M framework is more important to obtain good density forecasts compared to good point forecasts. Finally, including the GARCH-M framework and/or the cross-country weighted averages of growth to the forecast model does not improve the forecast performance of recessionary events in the G7 by much.

A Appendix A

Table 5: Unit root, distributional, serial correlation and conditional heteroskedasticity test results for the monthly growth rates of industrial production for the period 1961:02–2013:05

Panel A: Unit root test							
	CAN	FR	GER	ITA	JAP	UK	US
ERS DF-GLS	-8.25	-7.69	-2.45	-8.03	-4.02	-3.08	-6.59
C.V. at 5%	-1.94	-1.94	-1.94	-1.94	-1.94	-1.94	-1.94
Panel B: Distribution							
	CAN	FR	GER	ITA	JAP	UK	US
Skewness	-0.133	0.129	0.058	0.197	-0.856	0.026	-0.899
Kurtosis	3.496	4.391	9.431	10.704	7.538	12.882	7.357
Jarque-Bera	8.28	52.34	1082.62	1556.96	615.55	2555.16	581.39
	[0.0159]	[0.0000]	[0.0000]	[0.0000]	[0.0000]	[0.0000]	[0.0000]
Panel C: Tests for serial correlation and conditional heteroscedasticity							
	CAN	FR	GER	ITA	JAP	UK	US
Q(3)	48.23	59.65	67.52	49.14	92.04	22.04	170.51
	[0.0000]	[0.0000]	[0.0000]	[0.0000]	[0.0000]	[0.0000]	[0.0000]
Q(9)	81.39	100.63	76.37	89.99	117.89	35.94	250.32
	[0.0000]	[0.0000]	[0.0000]	[0.0000]	[0.0000]	[0.0000]	[0.0000]
$Q^2(3)$	14.49	25.11	87.63	109.95	377.60	109.25	128.05
	[0.0020]	[0.0000]	[0.0000]	[0.0000]	[0.0000]	[0.0000]	[0.0000]
$Q^2(9)$	27.68	66.56	89.46	138.60	399.19	113.75	140.66
	[0.0010]	[0.0000]	[0.0000]	[0.0000]	[0.0000]	[0.0000]	[0.0000]
ARCH(3)	12.36	23.11	104.95	104.45	235.24	114.04	96.84
	[0.0063]	[0.0000]	[0.0000]	[0.0000]	[0.0000]	[0.0000]	[0.0000]
ARCH(9)	21.33	48.49	105.51	114.38	241.40	118.48	100.05
	[0.0113]	[0.0000]	[0.0000]	[0.0000]	[0.0000]	[0.0000]	[0.0000]

Notes:
- p-values are in square brackets.
- The unit root rest is the Elliott, Rothenberg, and Stock (1996) test. This test has more power than the original augmented Dickey-Fuller unit root test. This test has also another advantage. When a break is present in the data, the Dickey-Fuller test is biased towards non-rejection of a unit root (Perron, 1989). The ERS DF-GLS unit root test, in contrast, is asymptotically robust to level breaks.
- Ljung-Box Q-statistics at lag p for the data, Q(p), to test for serial correlation.
- Ljung-Box Q-statistics at lag p for the squared data, $Q^2(p)$, to test for conditional heteroscedasticity.
- ARCH LM test at lag p, ARCH(p), to test for conditional heteroscedasticity (R. F. Engle, 1982) .

Table 6: Summary statistics of the monthly growth rates of industrial production for the G7 (in %), 1961:02–2013:05

	N	Median	Mean	Sta. Dev.	Min	Max
CAN	628	0.19	0.24	1.10	-3.76	4.13
FR	628	0.04	0.14	1.35	-4.70	5.66
GER	628	0.23	0.20	1.75	-9.46	12.31
ITA	628	0.18	0.18	2.16	-14.78	13.51
JAP	628	0.38	0.34	1.54	-8.38	6.60
UK	628	0.10	0.09	1.36	-7.88	9.73
US	628	0.29	0.24	0.76	-4.21	3.09

Table 7: Adjustments to model selection procedure of the rolling estimations

Panel A: AR-GARCH-M

- *France*

  For estimation 60 there is evidence for ZILC in the GARCH model picked by the model selection. I set the variance model to ARCH.

- *Italy*

  For estimation 57-58, 62 and 65-69 there is evidence for ZILC in the GARCH model picked by the model selection. I set the variance model to ARCH. The same estimations, however, suffer now from serial correlation. To eliminate the serial correlation for these estimations, I set the number of autoregressive lags to six.

- *Japan*

  For estimation 48-52, 54-82, 85-87 and 89-90 there is conditional heteroskedasticity in the ARCH model picked by the model selection. I set the variance equation to GARCH. Estimation 48, however, then suffers from ZILC. So I set estimation 48 back to the ARCH model. These adjustments, however, do not eliminate the conditional heteroskedasticity for estimations 48-52 and 55-77, 79 and 81.

- *United States*

  For estimation 51 and 58 there is conditional heteroskedasticity in the ARCH model picked by the model selection. I set the variance equation to GARCH.

Table 7: Continued

Panel B: AR-Y*-GARCH-M

- *France*

  For estimation 42, 46-48, 48-69 there is evidence for ZILC in the GARCH model picked by the model selection. For these estimations I set the variance model to ARCH.

- *Italy*

  For estimation 83-84 and 87 there is evidence of serial correlation. To eliminate the serial correlation, I set the number of autoregressive and cross-country weighted average lags to six for these estimations.

- *Japan*

  For estimation 34-35 and 37 there is evidence of ZILC and for estimation 49-52 and 56-86 there was evidence of conditional heteroskedasticity. For the estimations where ZILC holds I set the variance to ARCH and for the estimations where there was still conditional heteroskedasticity in the errors, I set the variance model to GARCH. Estimation 49-52 and 56-86, however, still suffer from conditional heteroskedasticity.

References

Bai, J. S., & Perron, P. (2003). Computation and analysis of multiple structural change models. *Journal of Applied Econometrics*, *18*(1), 1–22.

Banbura, M., Giannone, D., Modugno, M., & Riechlin, L. (2013). Now-casting and the real-time data flow. *ECB Working Paper Series*(1564).

Bernanke, B. S. (2004). The great moderation. *At the meetings of the Eastern Economic Association, Washington, DC February 20, 2004*.

Bollerslev, T., & Wooldridge, J. (1992). Quasi-maximum likelihood estimation and inference in dynamic models with time-varying covariances. *Econometric reviews*, *11*(2), 143-172.

Dées, S., Di Mauro, F., Pesaran, M. H., & Smith, L. V. (2007). Exploring the international linkages of the euro area: A global var analysis. *Journal of Applied Econometrics*, *22*(1), 1–38.

Diebold, F., & Mariano, R. (1995). Comparing Predictive Accuracy. *Journal of Business & Economic Statistics*, *13*(3), 253-263.

Elliott, G., Rothenberg, T. J., & Stock, J. H. (1996). Efficient tests for an autoregressive unit root. *Econometrica*, *64*(4), 813–836.

Engle, R., Lilien, D., & Robins, R. (1987). Estimating Time-varying Risk Premia In the Term Structure - the Arch-m Model. *Econometrica*, *55*(2), 391–407.

Engle, R. F. (1982). Autoregressive Conditional Heteroscedasticity With Estimates of the Variance of United-kingdom Inflation. *Econometrica*, *50*(4), 987–1007.

Fang, W. S., & Miller, S. M. (2009). Modeling the volatility of real GDP growth: The case of Japan revisited. *Japan and the World Economy*, *21*(3), 312–324.

Giacomini, R., & White, H. (2006). Tests of conditional predictive ability. *Econometrica*, *74*, 1545-1578.

Herrera, A. M., & Pesavento, E. (2005). The decline in US output volatility: Structural changes and inventory investment. *Journal of Business & Economic Statistics*, *23*(4), 462–472.

Hillebrand, E. (2005). Neglecting parameter changes in GARCH models. *Journal of Econometrics*, *129*(1-2), 121–138.

Lamoureux, C. G., & Lastrapes, W. D. (1990). Persistence In Variance, Structural-change, and the Garch Model. *Journal of Business & Economic Statistics*, *8*(2), 225–234.

Ma, J., Nelson, C. R., & Startz, R. (2007). Spurious inference in the GARCH (1,1) model when it is weakly identified. *Studies In Nonlinear Dynamics and Econometrics*, *11*(1), 1081-1826.

Mitchell, J. (2009). Where are we now? the uk recession and nowcasting gdp growth using statistical models. , *209*(1), 60-69.

Nelson, C., & Startz, R. (2007). The zero-information-limit condition and spurious inference in weakly identified models. *Journal of Econometrics*, *138*(1), 47–62.

Nordhaus, W. (2004). Retrospective on the 1970s productivity slowdown. *NBER working paper no. 10950*.

Perron, P. (1989). The Great Crash, the Oil Price Shock, and the Unit-root Hypothesis. *Econometrica*, *57*(6), 1361–1401.

Pesaran, M. H., Schuermann, T., & Weiner, S. M. (2004). Modeling regional interdependencies using a global error-correcting macroeconometric model. *Journal of Business & Economic Statistics*, *22*(2), 129–162.

Pesaran, M. H., & Timmermann, A. (2009). Testing Dependence Among Serially Correlated Multicategory Variables. *Journal of the American Statistical Association*, *104*(485), 325-337.

Trypsteen, S. (2014). Cross-Country Interactions, the Great Moderation and the Role of Volatility in Growth. *CFCM working paper (University of Nottingham)*, *14*(14).

West, K. (1996). Asymptotic inference about predictive ability. *Econometrica*, *64*(5), 1067-1084.