International Journal of Forecasting 31 (2015) 769–781

Contents lists available at ScienceDirect

International Journal of Forecasting

journal homepage: www.elsevier.com/locate/ijforecast

Interpretation and use of sensitivity in econometrics, illustrated with forecast combinations

Jan R. Magnus [a], Andrey L. Vasnev [b,*]

[a] Department of Econometrics & Operations Research, VU University Amsterdam, The Netherlands

[b] University of Sydney, New South Wales, Australia

ARTICLE INFO

Keywords:
Sensitivity analysis
Reconciling forecasts
Yield curve prediction

ABSTRACT

Sensitivity analysis is important both for its own sake and in combination with diagnostic testing. We consider the question of how to use sensitivity statistics in practice, and in particular, how to judge whether the sensitivity is large or small. For this purpose, we distinguish between absolute and relative sensitivity, and highlight the context-dependent nature of sensitivity analysis. The relative sensitivity is then applied to forecast combinations, and sensitivity-based weights are introduced. All of the concepts are illustrated using the European yield curve. In this context, it is natural to consider the sensitivity to autocorrelation and normality assumptions. Different forecasting models are combined using equal, fit-based and sensitivity-based weights, and compared with the multivariate and random walk benchmarks. We show that the fit-based and sensitivity-based weights are complementary, but that the sensitivity-based weights perform better than other weights for long-term maturities.

1. Introduction

The majority of papers in applied econometrics concentrate on the fit of the models and the statistical significance of the coefficients, while sensitivity analysis is often not reported at all, or is reported only tangentially. This is unfortunate, because sensitivity analysis is at least as important as diagnostic testing. While diagnostic testing attempts to answer the question: is it true? (for example, that a coefficient is zero), sensitivity analysis addresses the question: does it matter? (that we set the coefficient to zero). At first glance, the two questions seem to be closely related, but Magnus and Vasnev (2007) showed that such is not the case. In fact, the two concepts are essentially orthogonal.

Fig. 1 shows the potential danger of ignoring sensitivity. The sample is given by three points, (x_1, y_1), (x_2, y_2), and (x_3, y_3), and two models are fitted. The horizontal line, given by the average value of the dependent variable $\bar{y} = (y_1 + y_2 + y_3)/3$, provides a minimal fit, but it is not sensitive to autocorrelation, non-normality, or other model assumptions. The other model provides a perfect fit, but it can only be used in a very small neighborhood of the sample points. It is unstable outside the data range $[x_1, x_3]$, and even within this range it produces unjustified values that are bigger than the maximum in the observed data. In this situation, the simple non-sensitive model is more reliable.

There are also situations in which one might be interested in a model with a high rather than a low sensitivity. For example, if we are interested in detecting a crisis or abnormalities in the market, then we prefer a model which is maximally sensitive, even to small indications of a crisis.

Magnus and Vasnev (2007) provide an overview of the sensitivity literature, and prove the asymptotic independence of the commonly-used diagnostic tests and the sensitivity statistic formally. Diagnostic tests and sensitivity statistics are therefore complementary, and

* Corresponding author.
E-mail address: andrey.vasnev@sydney.edu.au (A.L. Vasnev).

0169-2070/$ – see front matter Crown Copyright © 2013 Published by Elsevier B.V. on behalf of International Institute of Forecasters. All rights reserved.
http://dx.doi.org/10.1016/j.ijforecast.2013.08.001

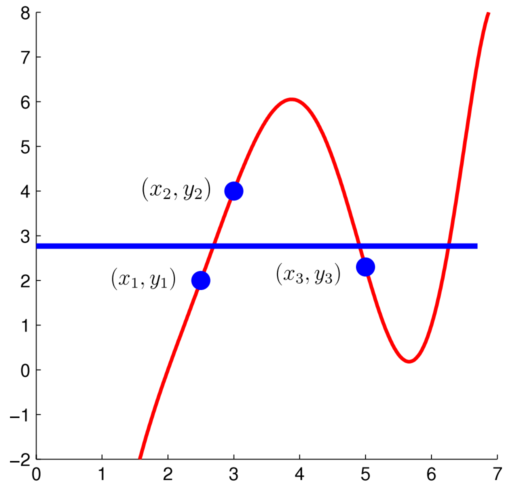


Fig. 1. The sample is given by three points. The horizontal line provides a minimal fit, but is not sensitive to model assumptions. The curve gives a perfect fit, but is very unstable.

both require our attention when analyzing a model. It is possible to derive sensitivity statistics, and several papers have suggested local and global sensitivity measures. However, it is more difficult to answer the question of when a sensitivity statistic is large or small, a question which is addressed in the current paper. The paper gives practical recommendations with regard to the way in which sensitivity statistics can be used. We shall see that the use of sensitivity is context-dependent, as is also emphasized by Severini (1996), so that we need to consider sensitivity in relation to the problem under consideration.

In some situations, the value of the sensitivity statistic is important, requiring a threshold in order to decide whether the model is sensitive or not; we call this case 'absolute sensitivity'. In other situations, only the relative magnitude is important, and we call this case 'relative sensitivity'. In both cases, it is essential to realize that sensitivity (unlike a diagnostic test) is context-dependent, and will be closely related to the estimator we are analyzing or the dependent variable we are modeling. To bring out this dependence, we illustrate all of the concepts introduced in this paper in a specific application, namely the forecasting of the Euro yield curve.

We show that, when several forecasts are available, the weights based on relative sensitivity perform well, and are complementary to the fit-based weights. The main purpose of combining forecasts is to improve the forecast accuracy (Bates & Granger, 1969), but the choice of weights is still an open question. Timmermann (2006) provides a thorough overview of the sizable forecast combination literature, but in practice the optimal weights have to be estimated, and this affects their actual performance. The adaptive weights seem to work well in many situations, but sometimes a simple alternative with equal weights gives better results, as was shown by Stock and Watson (2004). This fact is explained by Winkler and Clemen (1992) as being due to the instability of the estimated weights used in generating the combined forecast.

The paper is organized as follows. Section 2 introduces the practical aspects of sensitivity analysis and provides a brief overview of the sensitivity literature. It highlights the context-dependent nature of sensitivity analysis (Section 2.1), and distinguishes between absolute (Section 2.2) and relative (Section 2.3) sensitivity. Section 3 applies the concept of relative sensitivity to forecast combinations, and introduces sensitivity-based weights. The empirical Euro yield curve illustration is given in Section 4, and a detailed description of the data is given in the Data Appendix. Section 5 concludes.

2. Sensitivity analysis in practice

The concept of sensitivity is closely related to the concept of robustness, with which readers may be more familiar. Robustness was introduced by Hampel, Ronchetti, Rousseeuw, and Stahel (1986), and has since been studied extensively in the literature; see Kitamura, Otsu, and Evdokimov (2013) for a recent contribution. Robustness deals primarily with the effects of slight perturbations in the observed data, and often uses the influence function as a tool. Sensitivity analysis deals not only with data perturbations, but also with model perturbations. Sensitivity to the model specification studies the changes in the model output (often an estimator, test or predicted value) when one or more of the assumptions underlying the model are perturbed. The simplest example is given by Banerjee and Magnus (1999), who studied the sensitivity of the ordinary least squares estimator to autocorrelation in the regression errors. In this paper, we concentrate on model sensitivity, and refer to it simply as sensitivity.

Magnus and Vasnev (2007) introduced local sensitivity through a Taylor expansion. If the variable (or parameter) of interest, say y, depends on a nuisance parameter, say θ, then $\hat{y}(\theta)$ denotes the estimator of y for each given value of θ. Special cases are the 'restricted' estimator $\hat{y}(0)$, obtained by setting $\theta = 0$, and the 'unrestricted' estimator $\hat{y}(\hat{\theta})$, obtained by setting θ equal to its estimated value $\hat{\theta}$. The function $\hat{y}(\theta)$ provides not only these two special cases, but the whole *sensitivity curve*, given by the estimates of y for each given value of θ.

The first-order Taylor expansion of the sensitivity curve at the restricted point is given by

$$\hat{y}(\theta) = \hat{y}(0) + S\,\theta + O(\theta^2), \tag{1}$$

where

$$S = \left.\frac{\partial \hat{y}(\theta)}{\partial \theta}\right|_{\theta=0} \tag{2}$$

is the first derivative at the restricted point $\theta = 0$, and is called the *local sensitivity statistic*, or simply the sensitivity.

Sensitivity is computed for maximum likelihood estimators by Magnus and Vasnev (2007), and, in general, it can be expressed in terms of the Hessian. In the cases of mean, variance, and distribution misspecification, the sensitivity statistics allow tractable representations. This is particularly the case for the B_s and D_s statistics of Banerjee and Magnus (1999), and the sensitivity of GLS estimators in panel data derived by Vasnev (2010).

One might think that the sensitivity statistic and the corresponding diagnostics would be highly correlated.

However, if this were the case, then the Durbin–Watson statistic (diagnostic) should be highly correlated with the sensitivity statistic of the regression coefficients as a function of the autocorrelation parameter, and Magnus and Vasnev (2007) showed that such is not the case. In fact, under general conditions, the sensitivity statistic and the most common diagnostic tests are asymptotically independent, and these general conditions are satisfied in the case of mean, variance, and distribution misspecifications. In other words, sensitivity analysis answers an essentially different (and arguably more important) question to that answered by diagnostic testing.

Magnus and Vasnev (2007) also provided an overview of the sensitivity area and the connection to related concepts in econometrics. Since then, the area has been investigated further. For example, Wan, Zou, and Qin (2007) studied the sensitivity of the restricted least squares estimators, while Qin, Wan, and Zou (2009) looked at the sensitivity of the one-sided t-test. Ashley (2009) assessed instrumental variable inference via sensitivity analysis, and sensitivity in panel data was studied by Vasnev (2010). Sensitivity analysis has also attracted attention in quantitative finance; see Pospisil and Vecer (2010) and, in a somewhat different framework, the earlier work by Gourieroux, Laurent, and Scaillet (2000).

We briefly introduce the sensitivity statistics used in the empirical illustration of Section 4. We shall consider the sensitivity of the forecast to AR(1) misspecification ($S^{\mathrm{AR}(1)}$) and to skewness (S^{sk}). One might argue that these are the two basic sensitivities which one should always compute. Sensitivity to AR(1) checks the importance of any remaining correlation in the model errors, and, in fact, captures more general forms of autocorrelation (Banerjee & Magnus, 1999). Sensitivity to skewness checks the importance of asymmetry in the error distribution. This is particularly relevant for applications where positive and negative deviations have different consequences, as is often the case in finance. For example, profit and loss announcements may have different effects on stock prices. In our empirical illustration, the interest rates cannot go below zero, which restricts the error distribution in a way which may or may not have an effect on the forecast.

2.1. When is sensitivity 'large' or 'small'?

In order to use sensitivity in practice, we need to decide when it is large and when it is small. Ideally, we would like to have a threshold similar to the 5% significance level which is typically used in diagnostic testing. If a sensitivity is below this threshold, then we call it not sensitive; otherwise, we call it sensitive.

Unfortunately, this is not easy; in fact, it is impossible. Of course, since sensitivity is a statistic with an estimable variance, we can obtain a 95% insensitivity interval; in other words, we can make statements as to the significance of our sensitivity statistic. However, this is not quite what we want. We want to know, not the significance of the sensitivity, but its *importance*, and the importance is not revealed by such intervals.

Sensitivity is and must be context-specific. For example, if the temperature outside changes by one degree, most of us hardly notice it, but it is easy to think of controlled environments in medicine or chemistry where a fraction of one degree is already too much. As another example, changes of around 1% are routine events in the stock market, but yield fluctuations in a fixed income of 1% would be considered colossal; here, changes of 0.05% (five basis points) are considered normal. Again, the temperature example also depends on the scale of measurement (Celsius, Fahrenheit). Such scale-dependence is found in many situations; for example, if $\hat{y}$ represents personal expenditure, then a change of \$100 is substantial, but if $\hat{y}$ represents national savings, \$100 difference is negligible. This issue can be resolved by considering relative changes, except of course when $\hat{y}$ is close to zero.

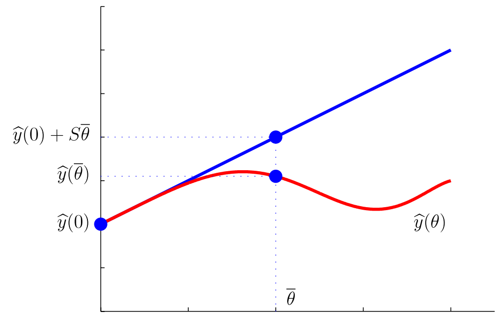


Fig. 2. Sensitivity approximation.

Duan (1993) measured the sensitivity S relative to the estimated value $\hat{y}$, while Severini (1996) suggested that the magnitude of a particular sensitivity value should be considered large if a change of this size would have an 'important' effect on the conclusion of the analysis. A less desirable, but more general, approach is based on comparing the sensitivity to the internal variability of the estimator, that is, the standard error of $\hat{y}(0)$.

2.2. Absolute sensitivity

One can be interested in either absolute or relative sensitivity, depending on the context. If one is interested in absolute sensitivity, then a sensitivity threshold, say δ, should be determined in advance (just like the significance level). Severini (1996) suggests using half the standard error of the estimator to distinguish between sensitive and non-sensitive cases.

Fig. 2 illustrates that there are two components of importance when the sensitivity curve $\hat{y}(\theta)$ is approximated by a first-order Taylor expansion, namely the direction S and the magnitude θ. Therefore, in order to determine δ, we need two bounds: (1) an upper bound for the nuisance parameter $\bar{\theta}$ representing our worst-case scenario in the direction given by θ; and (2) a bound for the quantity of interest $\Delta y = \hat{y}(\bar{\theta}) - \hat{y}(0)$ that we are willing to tolerate. These two bounds then lead to a sensitivity threshold

$$\delta = \Delta y/\bar{\theta}, \tag{3}$$

so that when $|S| < \delta$, we may call the sensitivity small, meaning that if the worst-case scenario $\bar{\theta}$ is realized, the change in $\hat{y}$ will be smaller than our tolerance Δy.

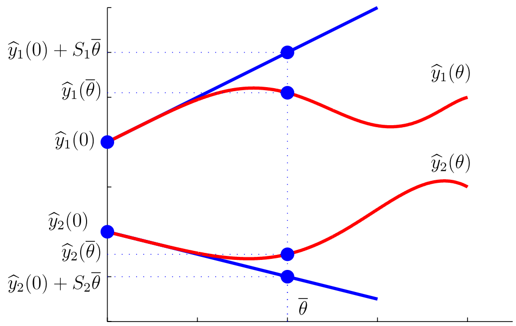


Fig. 3. Sensitivity comparison of two models.

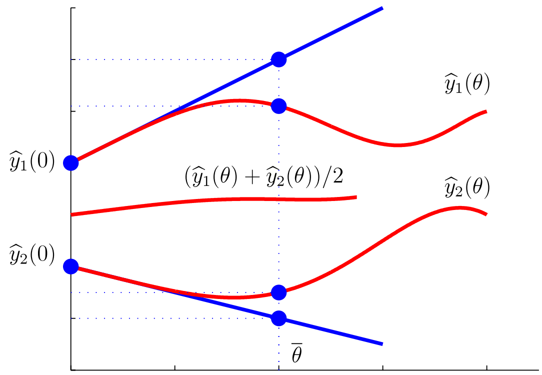


Fig. 4. Sensitivity of the average of two models.

2.3. Relative sensitivity

Generally, the sensitivities of different models in different directions are not comparable, but one important exception is when the direction is the same. For example, if we have two models such that $\hat{y}_1$ and $\hat{y}_2$ predict (or estimate) the same thing, then their sensitivities in the same direction (θ) are comparable.

Fig. 3 illustrates this idea. In this case, a comparison of the changes in y approximated by $S_1\overline{\theta}$ and $S_2\overline{\theta}$ is equivalent to a comparison of the sensitivities S_1 and S_2 themselves. Therefore, if $|S_1| < |S_2|$, then model 1 is less sensitive than model 2 because it will produce a smaller change in y when θ deviates from zero. In some situations, the sign might be of importance as well.

3. Forecast combinations and relative sensitivity

3.1. Model/forecast combination

A natural application of the relative sensitivity discussed in Section 2.3, where many models are used for predicting the same thing, is the combination of forecasts. We refer the reader to Timmermann (2006) for an overview of this area; our focus is only on the sensitivity aspect. We simplify things by considering two models, but the generalization to more models is straightforward.

Suppose we consider the weighted average of two models with outputs $\hat{y}_1$ and $\hat{y}_2$ respectively:

$$\hat{y}_c = w\hat{y}_1 + (1-w)\hat{y}_2 \quad (0 \le w \le 1). \tag{4}$$

If the sensitivities of the individual models are S_1 and S_2, then the sensitivity of the combination is given by

$$S_c = wS_1 + (1-w)S_2. \tag{5}$$

The (absolute value of the) sensitivity of the combination is smaller than the average (absolute value of the) sensitivity of the individual models, because

$$|S_c| \le w|S_1| + (1-w)|S_2|. \tag{6}$$

Also, S_c will be in between S_1 and S_2: if $S_1 \le S_2$, then $S_1 \le S_c \le S_2$. More generally,

$$\min S_i \le S_c \le \max S_i \tag{7}$$

holds for a weighted average of any number of models.

Of particular importance is the case when the two sensitivities have opposite signs, say $S_2 < 0 < S_1$, as is illustrated in Fig. 4.

When the sensitivities of two models have opposite signs, they will compensate each other, and the combination becomes less sensitive than either of the underlying estimates. Regarding the absolute value of the sensitivities, we find that $|S_c| < \min(|S_1|, |S_2|)$ in the following cases:

$$\begin{cases} \text{for } 0 < w < -2S_2/(S_1-S_2) < 1 \\ \quad \text{when } S_1+S_2 > 0; \\ \text{for } 0 < w < 1 \quad \text{when } S_1+S_2 = 0; \\ \text{for } 0 < -(S_1+S_2)/(S_1-S_2) < w < 1 \\ \quad \text{when } S_1+S_2 < 0. \end{cases} \tag{8}$$

Hence, if $S_1 + S_2 = 0$, then *every* combination will be an improvement in terms of (absolute) sensitivity, but if $S_1 + S_2 \neq 0$, then only some choices of w will lead to an improvement. For example, if $S_1 = 2$ and $S_2 = -1$, then an improvement occurs when we choose $0 < w < 2/3$, but not when $2/3 < w < 1$. Similarly, if $S_1 = 2$ and $S_2 = -3$, then an improvement occurs when we choose $1/5 < w < 1$, but not when $0 < w < 1/5$.

In Fig. 4, we have chosen $w = 1/2$, and the sensitivity is obviously much reduced. This reduction in sensitivity provides a possible explanation of the good performance of forecast combinations in applications. The combination is often less sensitive than the individual models, and in fact, in the case of two models with sensitivities of opposite signs, we can reduce the combined sensitivity to zero by choosing the weight $w = S_2/(S_2 - S_1)$.

3.2. Sensitivity-dependent weights

In applications, the weight is typically determined by a measure of fit, but it might be better to introduce weights based on sensitivity. We now discuss how this can be achieved.

One possibility is given by solving the optimization problem

$$\min_w \left(w|S_1| - (1-w)|S_2|\right)^2, \tag{9}$$

which minimizes the difference in sensitivity between the two components of the combined forecast in Eq. (4). The optimal weight will make the sensitivities of the

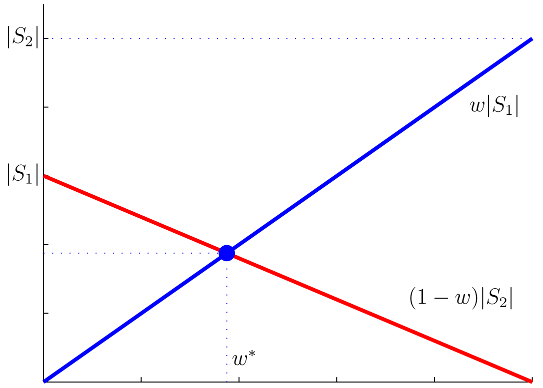


Fig. 5. Graphical representation of the optimization problem in Eq. (9) and the optimal solution w^*. The sensitivity contribution of $w\hat{y}_1$ in the combined forecast in Eq. (4) is given by the downward sloping red line. The sensitivity contribution of $(1-w)\hat{y}_2$ in the combined forecast in Eq. (4) is given by the upward sloping blue line. The sensitivities are equal at the intersection w^*. (For the interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

components $w\hat{y}_1$ and $(1-w)\hat{y}_2$ as close as possible. In fact, the optimal weight

$$w^* = \frac{|S_2|}{|S_1|+|S_2|} = \frac{1/|S_1|}{1/|S_1|+1/|S_2|} \tag{10}$$

takes the objective function to zero by setting the sensitivities of $w\hat{y}_1$ and $(1-w)\hat{y}_2$ equal to each other. Fig. 5 provides a graphical illustration of this fact.

The solution w^* to the optimization problem in Eq. (9) thus gives the model with the lower sensitivity a higher weight in the combination, and the sensitivity of the combined model is then

$$S_c = \frac{S_1/|S_1|+S_2/|S_2|}{1/|S_1|+1/|S_2|}. \tag{11}$$

If the sensitivities have opposite signs, then $S_c = 0$, irrespective of the size of the sensitivities. If they have the same sign, then

$$S_c = \frac{2S_1S_2}{S_1+S_2}. \tag{12}$$

Another possibility is to choose the weights to be proportional to the sensitivity,

$$w^{**} = \frac{|S_1|}{|S_1|+|S_2|}, \tag{13}$$

so that the model with the higher sensitivity gets a higher weight in the combination. In that case, the sensitivity of the combined model will be

$$S_c = \frac{S_1|S_1|+S_2|S_2|}{|S_1|+|S_2|}. \tag{14}$$

If the sensitivities have opposite signs, then $S_c = S_1 + S_2$. If the sensitivities have the same sign, then

$$S_c = S_1 + S_2 - \frac{2S_1S_2}{S_1+S_2}. \tag{15}$$

A third possibility is to combine fit and sensitivity. If the fit is measured by the root mean square forecast error (RMSFE), then one might define the weight as

$$w^{***} \sim 1/\text{RMSFE}_1 + 1/|S_1|. \tag{16}$$

More generally, the weight can be defined as a function $w(\text{RMSFE}_1, S_1)$, which is non-increasing in the first argument and non-increasing (or non-decreasing in some situations) in the second argument.

The weights w^* and w^{**} given by Eqs. (10) and (13) provide a simple way to bring the sensitivity into consideration; but what are the theoretical properties of those weights? Various different criteria can be used for this purpose. Bates and Granger (1969) minimize the variance of the combined forecast, while later contributions use different criteria: the in-sample mean squared error and the out-of-sample one-step-ahead mean squared forecast error (Hansen, 2008), the expected squared error, that is, the risk (Hansen & Racine, 2012), and the risk of the subset regression estimator (Elliott, Gargano, & Timmermann, in press). All of these criteria are closely related to the in-sample and out-of-sample fits of the model, that is, they are based on diagnostics, not on sensitivity statistics.

One should also keep in mind that theoretically optimal weights may not lead to a good performance in applications, because, in applications, these optimal weights have to be replaced by estimates. The classical example is given by Bates and Granger (1969), who show that the optimal weight that minimizes the variance of the combination (with a non-zero covariance between individual forecasts) performs worse than simple alternatives, even when the simple alternative is based on the incorrect assumption of zero covariance. More recently, Vasnev and Pauwels (2013) showed that the optimal weights in density forecast combinations are severely affected by how the weights are estimated in practice.

Returning to the question as to the theoretical properties of the weights given in Eqs. (10) and (13), we use the framework of Bates and Granger (1969). Let $\hat{y}_1$ and $\hat{y}_2$ be unbiased forecasts with variances of σ_1^2 and σ_2^2, respectively, and assume for simplicity that they are uncorrelated. The variance of the linear combination in Eq. (4) is given by

$$\sigma_c^2 = w^2\sigma_1^2 + (1-w)^2\sigma_2^2, \tag{17}$$

which is minimized for $w = \sigma_2^2/(\sigma_1^2+\sigma_2^2)$.

Forecasts typically come from different models with different sensitivities, say S_1 and S_2. This additional information can be used to adjust the criterion in Eq. (17) to

$$w^2\sigma_1^2|S_1| + (1-w)^2\sigma_2^2|S_2|, \tag{18}$$

so that the individual variances are sensitivity-adjusted. In this case, the optimal solution is

$$w^* = \frac{\sigma_2^2|S_2|}{\sigma_1^2|S_1|+\sigma_2^2|S_2|}, \tag{19}$$

which gives us the weights defined in Eq. (10) when the two variances σ_1^2 and σ_2^2 are equal.

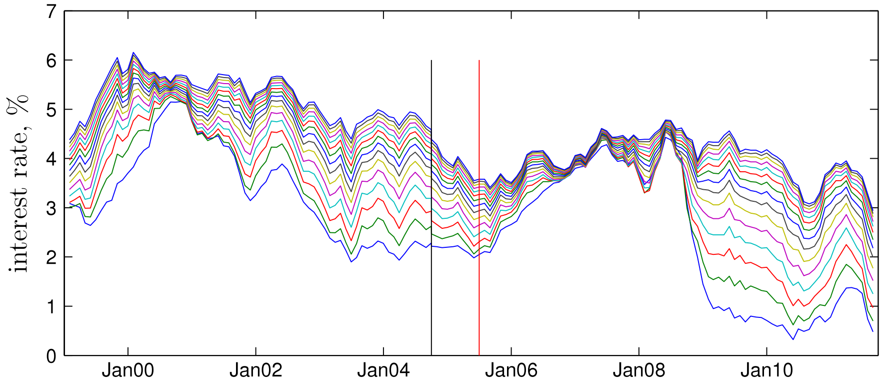


Fig. 6. The monthly zero-coupon yield curve spot rate for AAA-rated euro area (EA11-2000, EA12-2006, EA13-2007, EA15-2008, EA16-2010, EA17) central government bonds between January 1999 and September 2011. The bottom line gives the 1-year maturity, and the top line gives the 15-year maturity. The left vertical line (black) indicates the switch in methodology in October 2004. The right vertical line (red) indicates July 2005, the beginning of the out-of-sample forecast in the extended period. (For the interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

If the sensitivity adjustment is done with the inverse sensitivities, then the criterion becomes

$$w^2\sigma_1^2/|S_1| + (1-w)^2\sigma_2^2/|S_2|, \tag{20}$$

and the optimal solution is

$$w^{**} = \frac{\sigma_2^2/|S_2|}{\sigma_1^2/|S_1| + \sigma_2^2/|S_2|}, \tag{21}$$

which gives us the weights defined in Eq. (13) when the two variances are equal.

In our empirical illustration, we shall only consider the weights defined in Eqs. (10) and (13), as being the simplest and most intuitive cases. One could also use the weights defined in Eqs. (19) or (21), or introduce other modifications, for example to minimize $\sigma_c^2 \times S_c$ or σ_c^2/S_c.

4. Empirical illustration: Euro yield curve

Since sensitivity is context-dependent, the concepts introduced above should be illustrated with a concrete example. As our example, we have chosen the European yield curve.

4.1. Data

The yield of a zero coupon bond is the rate that equates the current price of the bond and the discounted principal repayment. Thus, the yield y solves the equation

$$Z = (1+y)^{-T}P, \tag{22}$$

where Z is the current price, P is the principal, and T is the maturity of the bond.

In reality, there are many types of bonds, which differ by origination date, maturity, payment structure, and embedded flexibility. Also, the creditworthiness of the issuer and the liquidity can vary substantially. A yield curve is a convenient way to aggregate all of this information into one object. The process of creating the yield curve from the original bonds is sometimes called 'distilling' or 'stripping' the yield curve. The final outcome depends on the methodology used to select the data and fit the curve.

In Europe, the centralized statistical office Eurostat collects the data and provides an estimate of the yield curve for the Eurozone area. Time series for maturities between 1 and 15 years are presented in Fig. 6. In October 2004, Eurostat changed their methodology, which led to a break in the series, indicated by the first black vertical line. This change has not affected the behavior of the yield curve, but has produced a shift in the level. In the empirical analysis, we therefore look at three periods:

1. historical period, January 1999–June 2005;
2. current methodology period, October 2004–September 2011; and
3. extended (combined) period, January 1999–September 2011, where the new methodology is used from October 2004.

To account for the difference created by the change in methodology, we add a dummy variable for the period starting in October 2004 when dealing with the extended period.

The European yield curve for the maturities from 1 to 15 years is forecasted with the help of macro and financial variables. We extend the dataset used by Magnus and Vasnev (2008) to include the latest available observations. A detailed description of the dataset is provided in the Data Appendix.

4.2. Absolute sensitivity

Fig. 7 shows the dynamics of the yield curve and highlights the fact that the curve can shift and can change its slope and curvature. The 1-year yield in December 2011 is 0.21%. In this context, the natural tolerance bounds would be $\Delta = 1$ bp (0.01%) or $\Delta = 10$ bp (0.1%). The reference interval for an absolute sensitivity analysis of the forecast/estimator $\hat{y}$ is given by $\hat{y} \pm \Delta$.

There are many directions in which one can look when analyzing a model that predicts yields. The natural directions for sensitivity analysis in this example are:

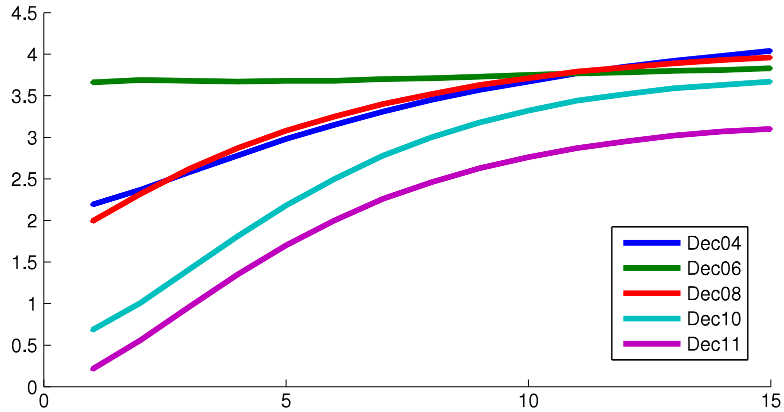


Fig. 7. Euro yield curve dynamics. The horizontal axis gives the maturity from 1 to 15 years. The vertical axis measures the yield. (For the interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

1. autocorrelation in the model error term, captured by sensitivity in the AR(1) direction, introduced by Banerjee and Magnus (1999); and
2. asymmetry in the error distribution, captured by the sensitivity to skewness, introduced by Magnus and Vasnev (2007).

Autocorrelation is important, as we are dealing with time series, and asymmetry is important because of the positiveness of the yield curve.

If the sensitivity in a particular direction θ is given by S, then, in order to answer the question of whether the (absolute) sensitivity is 'large' or 'small', one has to compare $|S\overline{\theta}|$ and Δ (or equivalently $|S|$ and $\Delta/\overline{\theta}$), where $\overline{\theta}$ represents the worst-case scenario in the chosen direction.

For the yield curve, the sign of the change is important. For borrowers, '+' is bad and '−' is good; for lenders, the opposite holds.

4.3. Forecast combination based on relative sensitivity

Following Stock and Watson (2004), we consider univariate models for forecast combinations in order to study the performance of different weights. In total, fourteen univariate models are combined, where one of the explanatory variables described in the Data Appendix is used for each model. This corresponds to the special case of complete subset regressions from Elliott et al. (in press), with $k = 1$. We employ the following weights:

1. equal weights;
2. fit-based weights, that is, weights which are inversely proportional to the root mean square forecast error (RMSFE) computed from the previous periods and the smoothed AIC-based (SAIC) weights of Hjort and Claeskens (2003); and
3. sensitivity-based weights introduced in Section 3.2.

Multivariate and random walk models serve as benchmarks. The forecast evaluations are based on RMSFE comparisons for standard one-step-ahead out-of-sample forecasts with increasing windows. The initial window is chosen to cover the first half of the period under consideration.

4.4. Sensitivity to AR(1) misspecification

From Magnus and Vasnev (2007), the sensitivity statistic of the forecast $\hat{y}_f = x_f'\hat{\beta}$ computed at the point of interest x_f' is given by

$$S_{\hat{y}_f}^{\mathrm{AR}(1)} = x_f' S_{\hat{\beta}}^{\mathrm{AR}(1)} = x_f'(X'X)^{-1}X'T^{(1)}My, \tag{23}$$

where X is the matrix of the regressors, y contains the observations on the dependent variable used to compute the OLS estimator $\hat{\beta}$, $T^{(1)}$ denotes the Toeplitz matrix of order one (that is, it contains ones just above and below the diagonal and zeros elsewhere), and $M = I_n - X(X'X)^{-1}X'$.

Fig. 8(a) shows the sensitivities of the univariate models and of the equal-weight combination across the historical out-of-sample forecast period. As expected, the sensitivity of the combination is smoother than those of the individual models. The most extreme behavior is exhibited by the model with RTT.

The numerical results are contained in the left panel of Table 1. The panel shows the out-of-sample RMSFEs for the forecast combinations of univariate models with various weights: equal weights, weights based on the fit, weights which are proportional to the sensitivity of the univariate model, and weights which are inversely proportional to the sensitivity. In the historical period, the weights which are proportional to sensitivity perform best for short maturities, the fit appears to be important for the medium maturity, and the weights which are inversely proportional to sensitivity are better for the long-term. This shows again that weights are case-specific, and illustrates the value of sensitivity analysis and its complementarity to diagnostic testing and measures of fit.

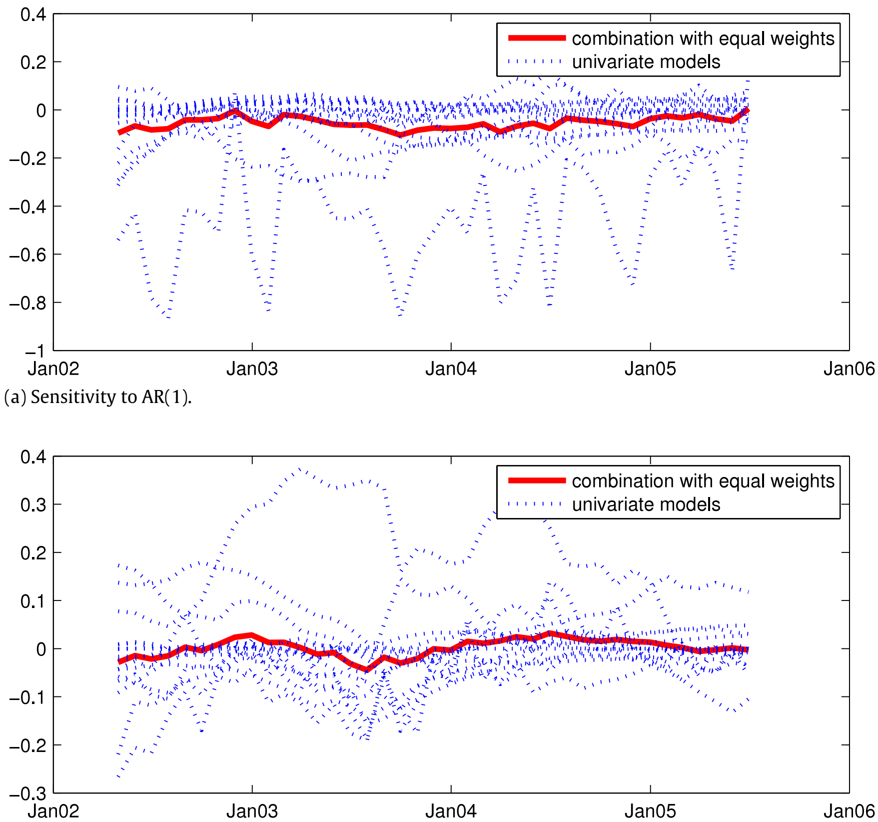
4.5. Sensitivity to skewness

Following Magnus and Vasnev (2007), the sensitivity statistic of the forecast $\hat{y}_f = x_f'\hat{\beta}$ computed at the point of interest x_f' is given by

$$S_{\hat{y}_f}^{\mathrm{sk}} = x_f' S_{\hat{\beta}}^{\mathrm{sk}} = -\frac{1}{2}x_f'\hat{\sigma}(X'X)^{-1}\sum_{i=1}^{n}(\hat{\epsilon}_i^2 - 1)x_i, \tag{24}$$

Table 1
RMSFEs of forecast combinations with different weights for the historical period. The best-performing weight for each maturity is shaded dark in the first panel. The weights which are proportional to skewness sensitivity are shaded light in the second panel, as they perform best across all weights.

Maturity	Equal weights	$w \sim 1/\text{RMSFE}$	$w \sim \lvert S^{\text{AR}(1)}\rvert$	$w \sim 1/\lvert S^{\text{AR}(1)}\rvert$	$w \sim \lvert S^{\text{sk}}\rvert$	$w \sim 1/\lvert S^{\text{sk}}\rvert$
1	0.739	0.632	0.533	0.860	0.478	1.003
2	0.764	0.690	0.597	0.813	0.567	0.930
3	0.748	0.687	0.618	0.747	0.597	0.910
4	0.725	0.671	0.625	0.699	0.605	0.939
5	0.701	0.652	0.625	0.687	0.612	0.821
6	0.678	0.632	0.622	0.627	0.613	0.795
7	0.658	0.615	0.616	0.614	0.608	0.745
8	0.641	0.600	0.608	0.615	0.597	0.731
9	0.628	0.589	0.601	0.614	0.584	0.721
10	0.618	0.581	0.596	0.604	0.572	0.717
11	0.611	0.576	0.592	0.598	0.561	0.702
12	0.605	0.572	0.588	0.580	0.551	0.702
13	0.601	0.571	0.586	0.550	0.543	0.699
14	0.598	0.570	0.583	0.553	0.538	0.693
15	0.596	0.570	0.581	0.552	0.534	0.687



(a) Sensitivity to AR(1).

(b) Sensitivity to skewness.

Fig. 8. Sensitivities of the univariate models and the equal-weight combination for a 3-year maturity during the historical period.

where $\hat{\sigma}$ is the OLS estimator of the standard deviation, $\hat{\epsilon}$ is the vector of normalized residuals $\hat{\epsilon} = My/\hat{\sigma}$, $\hat{\epsilon}_i$ is its ith component, and x_i' represents the ith row of the matrix X.

Fig. 8(b) shows the sensitivities of the univariate models and of the equal-weight combination across the historical out-of-sample forecast period. Again, the combination is smoother than the individual models.

The right panel in Table 1 shows that the weights which are proportional to skewness sensitivity outperform other weights. This indicates that skewness is an important feature of the forecasts during the historical period.

Table 2
RMSFEs of forecast combinations with different weights for the new methodology period. The best-performing weight for each maturity is shaded.

Maturity	Multivariate model	Random walk	Equal weights	$w \sim 1/\text{RMSFE}$	$w \sim \lvert S(\text{AR1})\rvert$	$w \sim 1/\lvert S(\text{AR1})\rvert$	$w \sim \lvert S(\text{sk})\rvert$	$w \sim 1/\lvert S(\text{sk})\rvert$
1	0.376	0.268	0.924	0.714	0.942	0.856	0.789	1.253
2	0.435	0.267	0.873	0.714	0.948	0.708	0.743	1.179
3	0.416	0.247	0.794	0.671	0.890	0.662	0.719	1.002
4	0.383	0.227	0.715	0.618	0.812	0.607	0.701	0.867
5	0.355	0.213	0.643	0.567	0.735	0.565	0.686	0.754
6	0.336	0.202	0.583	0.524	0.667	0.516	0.641	0.658
7	0.320	0.192	0.535	0.492	0.614	0.493	0.573	0.589
8	0.311	0.186	0.498	0.467	0.573	0.444	0.516	0.565
9	0.302	0.180	0.470	0.448	0.539	0.433	0.464	0.547
10	0.297	0.176	0.451	0.433	0.511	0.424	0.436	0.526
11	0.295	0.173	0.437	0.422	0.491	0.406	0.420	0.502
12	0.295	0.171	0.430	0.417	0.480	0.409	0.413	0.487
13	0.296	0.170	0.424	0.410	0.473	0.409	0.410	0.475
14	0.297	0.169	0.422	0.411	0.468	0.403	0.410	0.471
15	0.299	0.171	0.422	0.411	0.466	0.404	0.410	0.469

Table 3
RMSFEs of forecast combinations with different weights for the extended period (with a dummy variable). The best-performing weight for each maturity is shaded.

Maturity	Multivariate model	Random walk	Equal weights	$w \sim 1/\text{RMSFE}$	SAIC weights	$w \sim \lvert S(\text{AR1})\rvert$	$w \sim 1/\lvert S(\text{AR1})\rvert$	$w \sim \lvert S(\text{sk})\rvert$	$w \sim 1/\lvert S(\text{sk})\rvert$
1	0.283	0.217	0.769	0.565	0.520	0.776	0.876	0.800	0.731
2	0.297	0.223	0.730	0.576	0.433	0.766	0.793	0.718	0.858
3	0.286	0.211	0.670	0.548	0.445	0.723	0.710	0.641	0.762
4	0.274	0.197	0.608	0.509	0.466	0.672	0.568	0.571	0.715
5	0.264	0.187	0.553	0.472	0.473	0.621	0.506	0.517	0.641
6	0.255	0.178	0.507	0.442	0.450	0.576	0.470	0.478	0.580
7	0.247	0.171	0.473	0.421	0.435	0.540	0.446	0.453	0.540
8	0.241	0.166	0.448	0.406	0.428	0.513	0.424	0.440	0.494
9	0.237	0.162	0.429	0.397	0.422	0.493	0.415	0.432	0.482
10	0.234	0.159	0.418	0.392	0.421	0.478	0.392	0.421	0.450
11	0.232	0.155	0.410	0.389	0.420	0.468	0.381	0.409	0.440
12	0.232	0.154	0.407	0.390	0.421	0.462	0.381	0.403	0.446
13	0.232	0.152	0.405	0.391	0.420	0.458	0.380	0.399	0.443
14	0.233	0.152	0.405	0.393	0.419	0.455	0.374	0.398	0.440
15	0.234	0.152	0.406	0.396	0.419	0.454	0.379	0.398	0.433

4.6. Further analysis

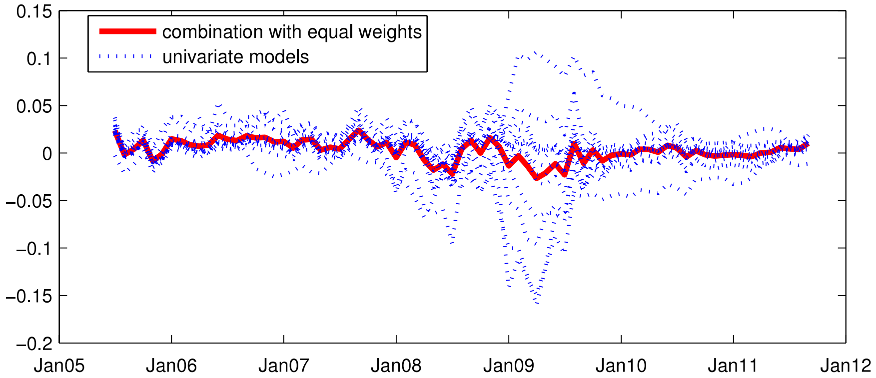
The results for the new methodology period (October 2004–September 2011) are given in Table 2. The weights that are inversely proportional to autocorrelation sensitivity perform the best in most cases. The table also provides results for the benchmark models, and shows that the forecast combination can be improved further.

The results for the extended period (January 1999–September 2011) are given in Table 3. Here, in addition to the weights already discussed, we also employ the smoothed AIC-based weights of Hjort and Claeskens (2003), which were used by Hansen (2008) as well. In the notation of Section 3.2, the smoothed AIC-based (SAIC) weight in the case of two models is given by

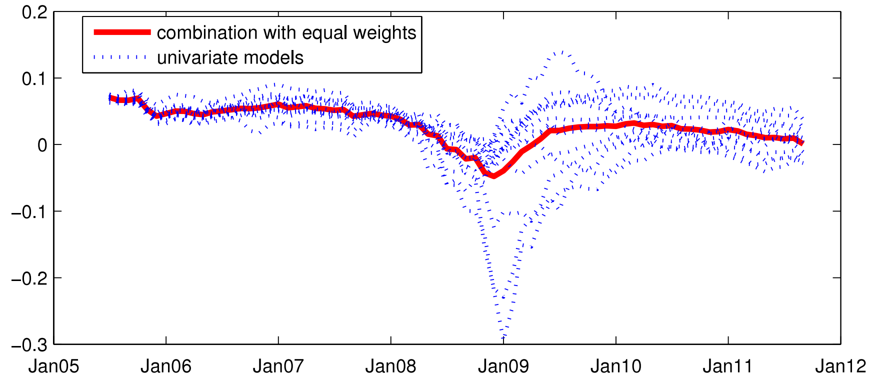
$$w = \frac{\exp\left(-\frac{1}{2}\text{AIC}_1\right)}{\exp\left(-\frac{1}{2}\text{AIC}_1\right) + \exp\left(-\frac{1}{2}\text{AIC}_2\right)}, \tag{25}$$

where AIC_1 and AIC_2 denote the Akaike information criterion computed for models 1 and 2, respectively.

Table 3 demonstrates the complementarity of sensitivity and fit-based weights. For maturities of up to 9 years, the fit is more important (the SAIC weights are better for maturities of 1 to 4 years, and the RMSFE-based weights are better for maturities of 5 to 9 years), while for maturi-



(a) Sensitivity to AR(1).



(b) Sensitivity to skewness.

Fig. 9. Sensitivities of the univariate models and the equal weight combination for a 3-year maturity during the extended period.

ties of over 9 years, autoregression sensitivity weights give the best results.

The sensitivities of the univariate models and the equal-weight combination are given in Fig. 9(a) and (b). The AR(1) sensitivities show more fluctuations in 2008, and destabilize in 2009. The skewness sensitivities declined during 2008 and destabilized dramatically in 2009. This highlights the forward-looking nature of sensitivity. The RMSFE only captures past performances, while sensitivity looks at the effects of additional features that might come into play. During and following the global financial crisis, the situation changed, and sensitivity was able (at least in part) to capture this change.

Nevertheless, the forecasts can be improved further. For this purpose, the random walk component is extracted first and the forecasts are produced for the remaining component. The results are given in Table 4. The distance between the benchmarks and the combinations is reduced. All of the combinations are better than the multivariate model. The best-weight combination is better than the random walk benchmark for the 1- and 2-year maturities, and close for the other maturities.

To compare the sensitivity- and fit-based weights in terms of their predictive accuracies, we test whether their differences are statistically significant. For this purpose, we employ the S_{2a} test statistic of Diebold and Mariano (1995) and compare the best-performing sensitivity-based weights with the best performing fit-based weights. In Table 4, the cases where the difference is statistically significant (at the 5% significance level) are labeled with an asterisk (*). In many cases, the test rejects the null hypothesis that the RMSFEs of the two methods are equal, even though the numerical values are close, especially for long-term maturities. Once again, there is strong evidence that the information provided by sensitivity is useful and is not contained in either diagnostic- or fit-measures.

4.7. *Macro model and financial model*

To complete the empirical part, we select two models and average them, in the spirit of Bates and Granger (1969). In our case, the obvious candidates are the model containing macro variables (HICP, IDOP, HUNE, RTT, DUR, IND, DR) and the model containing financial variables (EURIBOR, LIBOR, GBP, USD, $SPRD_{2,1}$, $SPRD_{5,2}$, $SPRD_{10,5}$). The results are given in Table 5.

In this situation, equal weights perform well for most of the maturities, but the sensitivity-based weights perform similarly to the fit-based weights and better than the individual models.

5. Concluding remarks

This paper has considered some practical aspects of sensitivity analysis, and has identified two feasible

Table 4
RMSFEs of forecast combinations with different weights for the extended period (with a dummy variable and a random walk component). The best-performing weight for each maturity is shaded. In addition, the shaded sensitivity-based weights are tested against the corresponding RMSFE-based weights using the Diebold and Mariano (1995) test.

Maturity	Multivariate model	Random walk	Equal weights	$w \sim 1/\text{RMSFE}$	SAIC weights	$w \sim \lvert S(\text{AR1})\rvert$	$w \sim 1/\lvert S(\text{AR1})\rvert$	$w \sim \lvert S(\text{sk})\rvert$	$w \sim 1/\lvert S(\text{sk})\rvert$
1	0.2145	0.2169	0.2092	0.2102	0.2190	0.2089*	0.2143	0.2043	0.2172
2	0.2518	0.2231	0.2218	0.2232	0.2276	0.2211	0.2250	0.2182	0.2251
3	0.2440	0.2109	0.2125	0.2139	0.2171	0.2128	0.2153	0.2114	0.2141
4	0.2288	0.1974	0.2006	0.2020	0.2044	0.2014	0.2027	0.2015	0.2002*
5	0.2159	0.1868	0.1911	0.1922	0.1948	0.1922	0.1940	0.1931	0.1910*
6	0.2045	0.1782	0.1832	0.1841	0.1870	0.1844	0.1850	0.1854	0.1825*
7	0.1955	0.1713	0.1768	0.1774	0.1807	0.1779	0.1780	0.1789	0.1772
8	0.1883	0.1659	0.1718	0.1722	0.1757	0.1726	0.1727	0.1738	0.1712*
9	0.1841	0.1620	0.1682	0.1684	0.1720	0.1686	0.1702	0.1699	0.1686
10	0.1799	0.1586	0.1650	0.1650	0.1687	0.1651	0.1677	0.1664	0.1652
11	0.1756	0.1551	0.1616	0.1615	0.1652	0.1615	0.1642	0.1627	0.1613*
12	0.1743	0.1536	0.1603	0.1599	0.1638	0.1599*	0.1623	0.1613	0.1610
13	0.1731	0.1523	0.1592	0.1587	0.1625	0.1586*	0.1608	0.1602	0.1581
14	0.1728	0.1516	0.1586	0.1579	0.1616	0.1578*	0.1596	0.1595	0.1575
15	0.1740	0.1521	0.1591	0.1584	0.1622	0.1583*	0.1603	0.1601	0.1593

* Indicates the cases where the difference is statistically significant (at the 5% significance level).

Table 5
RMSFEs of forecast combinations of macro and financial models with different weights for the extended period (with a dummy variable and a random walk component). The best-performing weight for each maturity is shaded.

Maturity	Random walk	Equal weights	$w \sim 1/\text{RMSFE}$	$w \sim \lvert S(\text{AR1})\rvert$	$w \sim 1/\lvert S(\text{AR1})\rvert$	$w \sim \lvert S(\text{sk})\rvert$	$w \sim 1/\lvert S(\text{sk})\rvert$	Macro model	Finance model
1	0.2169	0.1931	0.1937	0.1979	0.1953	0.2003	0.1945	0.1961	0.2070
2	0.2231	0.2195	0.2196	0.2241	0.2214	0.2280	0.2189	0.2150	0.2388
3	0.2109	0.2111	0.2114	0.2140	0.2156	0.2197	0.2113	0.2104	0.2313
4	0.1974	0.1983	0.1989	0.2006	0.2044	0.2073	0.1998	0.2015	0.2184
5	0.1868	0.1887	0.1895	0.1904	0.1956	0.2000	0.1911	0.1948	0.2082
6	0.1782	0.1812	0.1822	0.1823	0.1884	0.1908	0.1831	0.1887	0.1998
7	0.1713	0.1753	0.1764	0.1765	0.1823	0.1845	0.1772	0.1834	0.1931
8	0.1659	0.1711	0.1723	0.1722	0.1781	0.1786	0.1735	0.1795	0.1879
9	0.1620	0.1689	0.1701	0.1699	0.1758	0.1759	0.1711	0.1773	0.1847
10	0.1586	0.1665	0.1678	0.1676	0.1732	0.1726	0.1698	0.1747	0.1816
11	0.1551	0.1635	0.1649	0.1651	0.1701	0.1705	0.1662	0.1717	0.1784
12	0.1536	0.1631	0.1645	0.1653	0.1692	0.1698	0.1667	0.1712	0.1775
13	0.1523	0.1625	0.1640	0.1644	0.1687	0.1698	0.1663	0.1709	0.1763
14	0.1516	0.1624	0.1641	0.1644	0.1682	0.1693	0.1653	0.1707	0.1759
15	0.1521	0.1637	0.1653	0.1671	0.1681	0.1697	0.1659	0.1718	0.1768

approaches. First, if the magnitude is important, then context-dependent thresholds can be introduced that classify the models as either sensitive or non-sensitive. Second, in some situations when the models estimate the same parameter or forecast the same variable, their sensitivities in one common direction can be compared directly. When applied to forecast combinations, the relative sensitivity then provides an additional dimension for comparing the forecasts. The paper also introduces novel weights for forecast combinations. In our empirical illustration, these sensitivity-based weights often perform better than the fit-based weights, although the results vary across different periods.

Acknowledgments

The authors thank Tommaso Proietti for helpful discussions; participants of several workshops and conferences in 2012 (MEAFA in Sydney, ISF in Boston, ESEM in Malaga, and SET in Sydney) and in 2013 (Melbourne and Tilburg) for constructive comments; Emma Smith for excellent research assistance in the initial stage of this project; and the associate editor and two anonymous reviewers for their detailed and constructive suggestions.

Appendix. Data appendix

The data used in the empirical example of Section 4 have been obtained from the Eurostat website http://epp.eurostat.ec.europa.eu/portal/page/portal/eurostat/home, and cover the period from January 1999 until September 2011. We analyze the Eurozone countries Austria, Belgium, Finland, France, Germany, Ireland, Italy, Luxembourg, The Netherlands, Portugal, Spain, Greece (from 1 January 2001), Slovenia (from 1 January 2007), Cyprus, Malta (both from 1 January 2008), Slovakia (from 1 January 2009), and Estonia (from 1 January 2011).

The dataset includes the 14 monthly predictors described below and a constant. The gross wages and salaries (GWS) variable used by Magnus and Vasnev (2008) is excluded, as it is no longer available. The unit root situation is addressed by using year-on-year log differences as per Magnus and Vasnev (2008).

HICP: Inflation is captured by the consumer price indices, which are measured for each country separately and then combined in one harmonized index. We use the harmonized index provided by Eurostat. The series contains a unit root, so we use its year-on-year log change, and denote it as HICP.

IDOP: Producer prices are reflected in the index of industrial domestic output prices. It contains a unit root, so we use its year-on-year log change, which we denote as IDOP.

HUNE: Unemployment is measured by the seasonally-adjusted harmonized unemployment index of all age classes including both males and females, which we denote as HUNE. Similarly to the price index, it is measured for each country separately and then combined into one harmonized measure using the weighted sum transformation.

RTT: The condition of industry and services is reflected by the seasonally-adjusted retail trade turnover index. The unit root in the series is removed in the standard way using the year-on-year log change, and we denote the result as RTT.

DUR: The seasonally adjusted index of the industrial production of consumer durables is differentiated in order to remove the unit root and denoted DUR.

IND: The stationary adjusted version of the industrial production index for the total industry excluding construction is denoted IND.

DR: The relationship between the official deposit rate (DR), the official refinancing operation rate (REF), and the official lending rate (LOAN) is kept fixed by the Central Bank, with a gap of 1% (with the exception of the period January–March 1999, when the gap between REF and LOAN was 1.5%). Therefore, we arbitrarily choose DR for our analysis.

EURIBOR, LIBOR: To account for other possibilities for investment, we include money market short-term interest rates EURIBOR for euro contracts and LIBOR for interbank loans in London. The LIBOR is taken because of the big influence this market has on the whole of Europe.

GBP, USD: We also include the exchange rate for the most important currencies: Pound Sterling (GBP) and United States Dollar (USD). In this way, we take into account international competition for investments.

SPRD: Piazzesi and Swanson (2004) find the yield spreads particularly useful for the analysis. We include three of them: the spreads between 2 and 1 year yields ($\mathrm{SPRD}_{2,1}$), between 5 and 2 year yields ($\mathrm{SPRD}_{5,2}$), and between 10 and 5 year yields ($\mathrm{SPRD}_{10,5}$).

References

Ashley, R. (2009). Assessing the credibility of instrumental variables inference with imperfect instruments via sensitivity analysis. *Journal of Applied Econometrics*, *24*, 325–337.

Banerjee, A. N., & Magnus, J. R. (1999). The sensitivity of OLS when the variance matrix is (partially) unknown. *Journal of Econometrics*, *92*, 295–323.

Bates, J. M., & Granger, C. W. J. (1969). The combination of forecasts. *Operational Research Quarterly*, *20*, 451–468.

Diebold, F. X., & Mariano, R. S. (1995). Comparing predictive accuracy. *Journal of Business and Economic Statistics*, *13*(3), 253–263.

Duan, N. (1993). Sensitivity analysis for Box–Cox power transformation model: contrast parameters. *Biometrika*, *80*(4), 885–897.

Elliott, G., Gargano, A., & Timmermann, A. (2013). Complete subset regressions. *Journal of Econometrics*, in press.

Gourieroux, C., Laurent, J. P., & Scaillet, O. (2000). Sensitivity analysis of values at risk. *Journal of Empirical Finance*, *7*, 225–245.

Hampel, F. R., Ronchetti, E. M., Rousseeuw, P. J., & Stahel, W. A. (1986). *Robust statistics. The approach based on influence functions*. John Wiley & Sons, Inc.

Hansen, B. E. (2008). Least-squares forecast averaging. *Journal of Econometrics*, *146*, 342–350.

Hansen, B. E., & Racine, J. S. (2012). Jackknife model averaging. *Journal of Econometrics*, *167*, 38–46.

Hjort, N. L., & Claeskens, G. (2003). Frequentist model average estimators. *Journal of the American Statistical Association*, *98*(464), 879–899.

Kitamura, Y., Otsu, T., & Evdokimov, K. (2013). Robustness, infinitesimal neighborhoods, and moment restrictions. *Econometrica*, *81*(3), 1185–1201.

Magnus, J. R., & Vasnev, A. L. (2007). Local sensitivity and diagnostic tests. *Econometrics Journal*, *10*, 166–192.

Magnus, J. R., & Vasnev, A. L. (2008). Using macro data to obtain better micro forecasts. *Econometric Theory*, *24*, 553–579.

Piazzesi, M., & Swanson, E. (2004). *Futures prices as risk-adjusted forecasts of monetary policy*. NBER working paper: vol. 10547.

Pospisil, L., & Vecer, J. (2010). Portfolio sensitivity to changes in the maximum and the maximum drawdown. *Quantitative Finance*, *10*, 617–627.

Qin, H., Wan, A. T. K., & Zou, G. (2009). On the sensitivity of the one-sided *t* test to covariance misspecification. *Journal of Multivariate Analysis*, *100*, 1593–1609.

Severini, T. A. (1996). Measures of the sensitivity of regression estimates to the choice of estimator. *Journal of the American Statistical Association*, *91*(436), 1651–1658.

Stock, J. H., & Watson, M. W. (2004). Combination forecasts of output growth in a seven-country data set. *Journal of Forecasting*, *23*, 405–430.

Timmermann, A. (2006). Forecast combinations. In G. Elliott, C. W. J. Granger, & A. Timmermann (Eds.), *Handbook of economic forecasting* (pp. 135–196). Elsevier.

Vasnev, A. L. (2010). Sensitivity of GLS estimators in random effects models. *Journal of Multivariate Analysis*, *101*, 1252–1262.

Vasnev, A. L., & Pauwels, L. (2013). *Practical considerations for optimal weights in density forecast combination*. Available at SSRN: http://ssrn.com/abstract=2210148 or http://dx.doi.org/10.2139/ssrn.2210148.

Wan, A. T. K., Zou, G., & Qin, H. (2007). On the sensitivity of the restricted least squares estimators to covariance misspecification. *The Econometrics Journal*, *10*, 471–487.

Winkler, R. L., & Clemen, R. T. (1992). Sensitivity of weights in combining forecasts. *Operations Research*, *40*, 609–614.

Jan R. Magnus is emeritus professor of econometrics at Tilburg University and visiting professor at the VU University Amsterdam. He is (co)author of eight books and more than one hundred scientific papers. His current interests are mainly model averaging, sensitivity analysis, catastrophe theory, and sport statistics.

Andrey L. Vasnev is a lecturer in Econometrics and Business Statistics at the University of Sydney Business School. He received his Ph.D. from Tilburg University, the Netherlands. His research interests are in applied econometrics, forecast combination and sensitivity analysis.