View metadata, citation and similar papers at core.ac.uk
brought to you by CORE
provided by Sydney eScholarship

The University of Sydney Business School
The University of Sydney

BUSINESS ANALYTICS WORKING PAPER SERIES

Practical use of sensitivity in econometrics with an illustration to forecast combinations

Jan R. Magnus
Tilburg University, The Netherlands
Andrey L. Vasnev
University of Sydney, New South Wales 2006, Australia

Abstract

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

March 6, 2013

BA Working Paper No: 04/2013
http://sydney.edu.au/business/business_analytics/research/working_papers

Practical use of sensitivity in econometrics with an illustration to forecast combinations

Jan R. Magnus

Tilburg University, The Netherlands

Andrey L. Vasnev

University of Sydney, New South Wales 2006, Australia

March 6, 2013

Abstract

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

1 Introduction

The majority of applied econometric papers concentrates on the fit of the models and the statistical significance of the coefficients. Sensitivity analysis is often not or only tangentially reported. 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 this is not the case. In fact, the two concepts are essentially orthogonal.

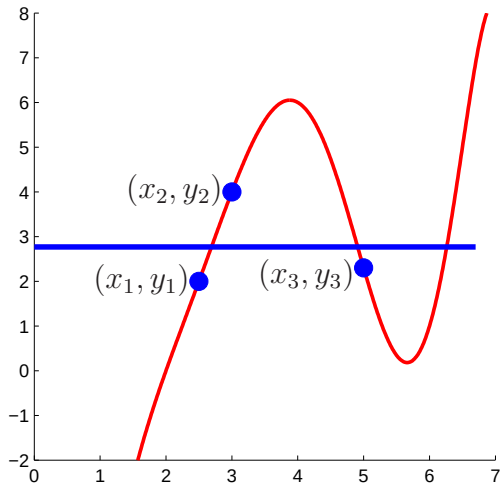


Figure 1: The sample is given by three points. The straight line provides minimal fit, but it is not sensitive to model assumptions. The curve gives perfect fit, but is very unstable.

A simple stylized example presented in Figure 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. A flat line, given by the average value of the dependent variable $\overline{y} = (y_1 + y_2 + y_3)/3$, provides minimal fit, but it is not sensitive to autocorrelation, non-normality, or any other model assumption. The other model provides 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 where one might be interested in a model with high rather than 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 to even small indications of a crisis.

Magnus and Vasnev (2007) provide an overview of the sensitivity literature, and prove formally the asymptotic independence of the commonly used diagnostic tests and the sensitivity statistic. Diagnostic tests and sensitivity statistics are therefore complementary, and both require our attention when analyzing a model. It is possible to derive sensitivity statistics, and several papers have suggested local or global sensitivity measures. It is, however, more difficult to answer the question when a sensitivity statistic is large or small. This question is addressed in the

current paper. The paper gives practical recommendations with regards to how the sensitivity statistics can be used, following the suggestion of Severini (1996) that the best approach is 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. We call this case 'relative sensitivity'. Essential for both cases is the realization that sensitivity (unlike a diagnostic test) is context-dependent, and will be closely related to the estimator we analyze or the dependent variable we are modeling. To bring out this dependence, we illustrate all concepts introduced in this paper in a specific context, namely forecasting 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 forecast accuracy, as first shown by Bates and Granger (1969). The choice of weights, however, is still an open question. Timmermann (2006) provides a thorough overview of the sizeable 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 shown by Stock and Watson (2004). This fact is explained by Winkler and Clemen (1992) as instability of 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 Practical aspects of sensitivity analysis

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.

One might think that the sensitivity statistic and the corresponding diagnostics would be highly correlated. If this were the case, then the Durbin-Watson statistic (diagnostic) should be highly correlated to the sensitivity statistic of the regression coefficients as a function of the autocorrelation parameter. Magnus and Vasnev (2007) showed that this 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 misspecification. In other words, sensitivity analysis answers an essentially different (and arguably more important) question than 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 further investigated. Wan et al. (2007) studied the sensitivity of the restricted least squares estimators. Qin et al. (2009) looked at the sensitivity of the one-sided t-test. Ashley (2009) assessed instrumental variable inference via sensitivity analysis. Sensitivity in panel data was studied in Vasnev (2010). Sensitivity analysis also attracted attention in quantitative finance; see Pospisil and Vecer (2010) and, in a somewhat different framework, the earlier work by Gourieroux et al. (2000).

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

In order to use sensitivity in practice, we need to decide when sensitivity is large and when it is small. Ideally we would like to have a threshold similar to the 5% significance level 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 a estimable variance, we can obtain an interval in which sensitivity lies with a probability of 5%. In other words, we can make statements as to the significance of our sensitivity statistic. But this is not quite what we want. We don't want to know the significance but the importance of the sensitivity, 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 control environments in medicine or chemistry where a fraction of one degree is already too much. Another example: in the stock market changes of around 1% are routine events, but yield fluctuations in fixed income of 1% would be considered colossal; here changes of 0.05% (five basis points) are considered normal.

The temperature example also depends on the scale of measurement (Celsius, Fahrenheit). This is true in many situations: if $\hat{y}$ represents personal expenditure then a change of 100 dollars is substantial, but if $\hat{y}$ represents national savings then a 100 dollar difference is negligible. This issue can be resolved by considering relative changes, except of course when $\hat{y}$ is close to zero.

Duan (1993) measured the sensitivity S relative to the estimated value $\hat{y}$. 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 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 absolute sensitivity or in 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 sensitive from non-sensitive cases.

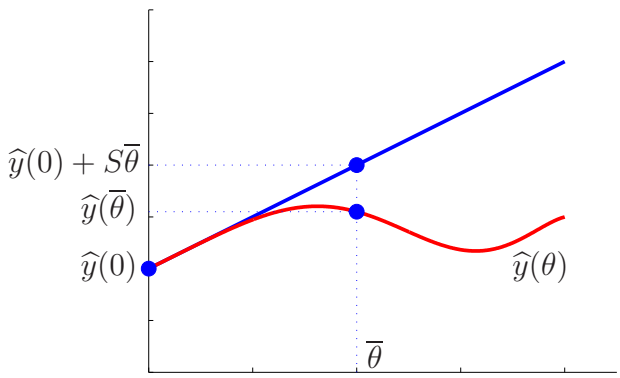


Figure 2: Sensitivity approximation.

Figure 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 $\overline{\theta}$ representing our worst-case scenario in the direction given by θ; and

2. a bound for the quantity of interest $\Delta y = \hat{y}(\overline{\theta}) - \hat{y}(0)$ that we are willing to tolerate.

These two bounds then lead to a sensitivity threshold

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

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

2.3 Relative sensitivity

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

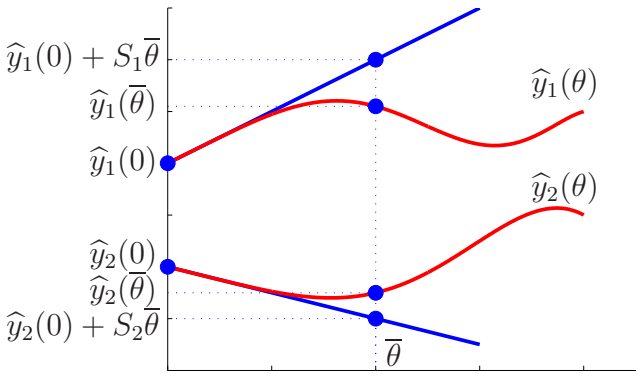


Figure 3: Sensitivity comparison of two models.

Figure 3 illustrates this idea. In this case, a comparison of a change in y approximated by $S_1\,\overline{\theta}$ and $S_2\,\overline{\theta}$ is equivalent to a comparison between 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 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:

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

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

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

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

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

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 illustrated in Figure 4.

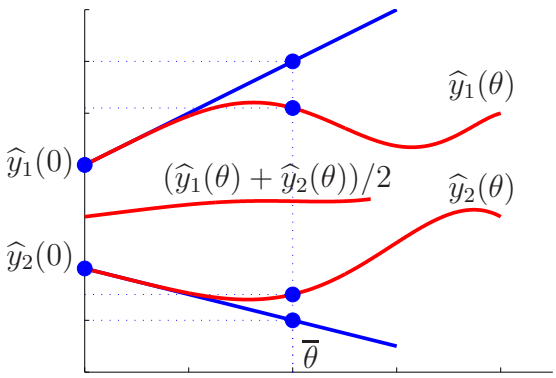


Figure 4: Sensitivity of the average of two models.

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

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

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 Figure 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 the forecast combination in applications. The combination is often less sensitive than the individual models. In the case of two models with sensitivities of opposite signs, we can in fact 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 usually determined by a measure of fit, but one can also introduce weights based on sensitivity, for example

$$w = \frac{1/|S_1|}{1/|S_1| + 1/|S_2|}, \tag{9}$$

giving the model with lower sensitivity a higher weight in the combination. The sensitivity of the combined model is then

$$S_c = \frac{S_1/|S_1| + S_2/|S_2|}{1/|S_1| + 1/|S_2|}. \tag{10}$$

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

$$S_c = \frac{2S_1S_2}{S_1 + S_2}. \tag{11}$$

A second possibility is to choose the weights proportional to the sensitivity,

$$w = \frac{|S_1|}{|S_1| + |S_2|}, \tag{12}$$

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

$$S_c = \frac{S_1|S_1| + S_2|S_2|}{|S_1| + |S_2|}. \tag{13}$$

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

$$S_c = S_1 + S_2 - \frac{2S_1S_2}{S_1 + S_2}. \tag{14}$$

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

$$w \sim 1/\mathrm{RMSFE}_1 + 1/|S_1|. \tag{15}$$

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

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. The yield y thus solves the equation

$$Z = (1 + y)^T P, \tag{16}$$

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 that 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 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 to 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 from 1 to 15 years are presented in Figure 5.

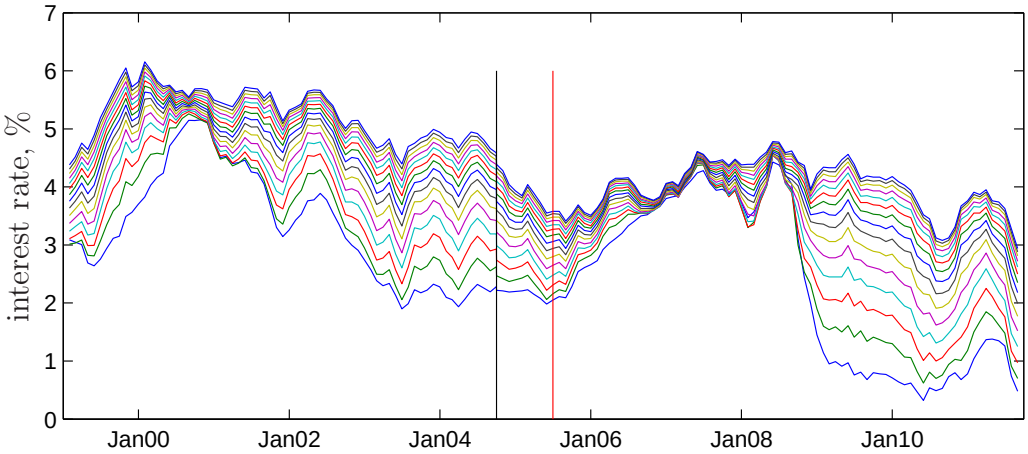


Figure 5: Monthly Zero-coupon yield curve spot rate for AAA rated euro area (EA11-2000, EA12-2006, EA13-2007, EA15-2008, EA16-2010, EA17) central government bonds between January 1999 and September 2011. The bottom line gives the 1-year maturity; the top line gives the 15-years maturity. The left vertical 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 period.

In October 2004, Eurostat changed their methodology, and this produces a break in the series indicated by the black vertical line. The change does not affect the behavior of the yield curve, but produces 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 use 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 data set used in Magnus and Vasnev (2008) to include the latest available observations. The detailed dataset description is provided in the data appendix.

4.2 Absolute sensitivity

Figure 6 shows the dynamics of the yield curve and highlights the fact the curve can shift and can change slope and curvature.

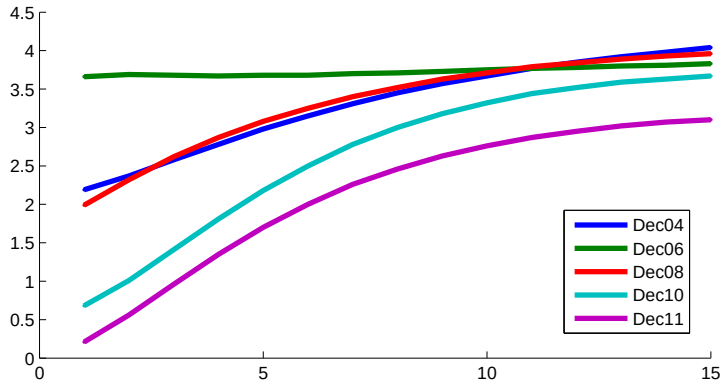


Figure 6: Euro yield curve dynamics. The horizontal axis gives maturity from 1 to 15 years. The vertical axis measures the yield.

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 absolute sensitivity analysis of the forecast/estimator $\hat{y}$ is given by

$$\hat{y} \pm \Delta. \tag{17}$$

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

1. autocorrelation in the model error term, which can be captured by sensitivity in AR(1) direction introduced by Banerjee and Magnus (1999),

2. asymmetry in the error distribution as the yield cannot be negative, which can be captured by sensitivity to skewness introduced by Magnus and Vasnev (2007).

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

If sensitivity in a particular direction θ is given by S, then to answer the question of absolute sensitivity, i.e. whether 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. We use the following weights

1. equal weights;
2. fit-based weights (weights inversely proportional to the root mean squared forecasting error, RMSFE, computed from the previous periods); and
3. sensitivity-based weights introduced in Section 3.2.

We use multivariate and random walk models as benchmarks. We also use standard one-step-ahead out-of-sample forecast with increasing windows and RMSFE for forecast evaluation.

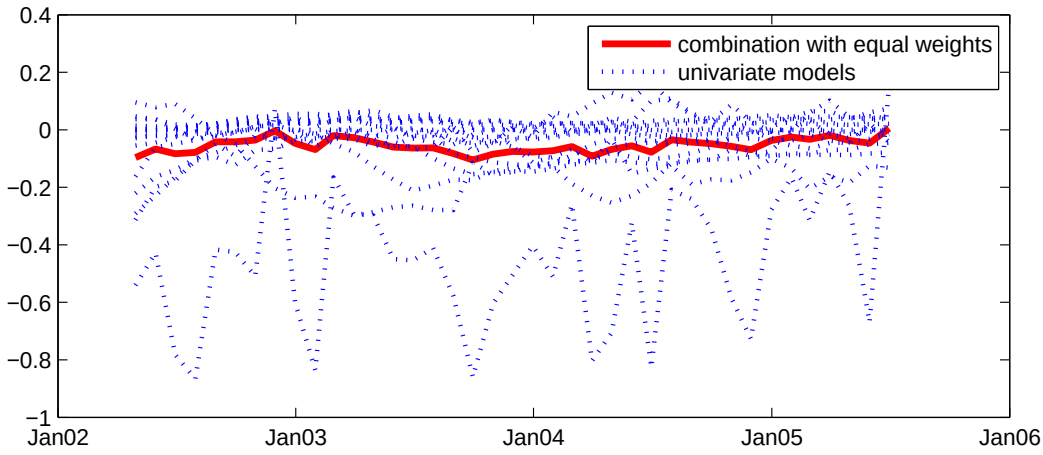
4.4 Sensitivity to AR(1) misspecification

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

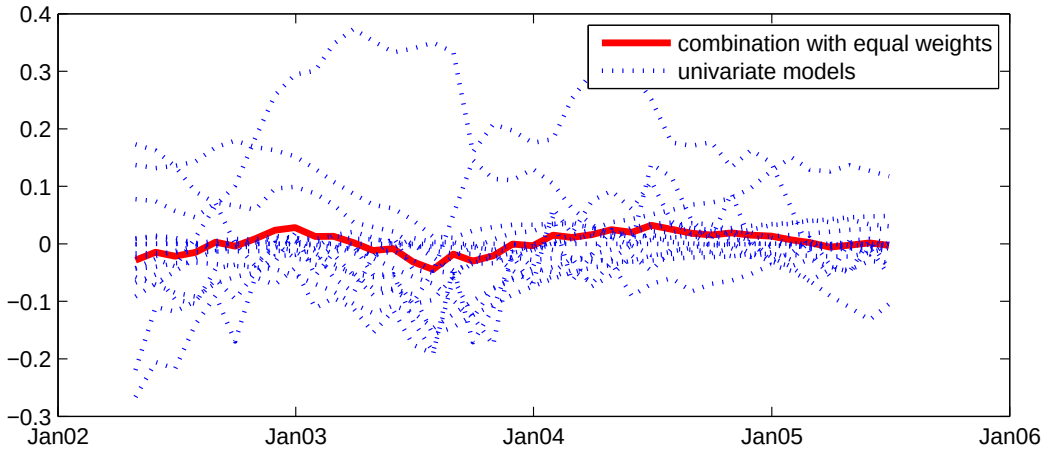
$$S_{\hat{y}_f}^{AR(1)} = x_f' S_{\hat{\beta}}^{AR(1)} = x_f'(X'X)^{-1}X'T^{(1)}My, \tag{18}$$

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 $M = I_n - X(X'X)^{-1}X'$.

Figure 7a 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 of the individual models. The most extreme behavior is exhibited by the model with RTT.



(a) Sensitivity to AR(1).



(b) Sensitivity to skewness.

Figure 7: Sensitivities of the univariate models and the equal weight combination for 3 year maturity during *the historical period*.

The numerical results are contained in the left panel of Table 1. The panel shows the out-of sample RMSFE of the forecast combinations of univariate models with equal weights, weights based on the fit, weights proportional to the sensitivity of the univariate model, and weights inversely proportional to the sensitivity. In the historical period for short maturities the weights proportional to sensitivity perform best, for the medium maturity the fit appears to be important, and for the long-term the inversely proportional weights to sensitivity are better. 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.

Maturity	Equal weights	$w \sim 1/\text{RMSFE}$	$w \sim \lvert S^{AR(1)} \rvert$	$w \sim 1/\lvert S^{AR(1)} \rvert$	$w \sim \lvert S^{sk} \rvert$	$w \sim 1/\lvert S^{sk} \rvert$
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

Table 1: RMSFE of forecast combinations with different weights for *the historical period*. The best performing weight for each maturity is dark shaded in the first pane. The weights proportional to skewness sensitivity are light shaded in the second pane as they are the best performing across all weights.

4.5 Sensitivity to skewness

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

$$S^{sk}_{\hat{y}_f} = x'_f S^{sk}_{\hat{\beta}} = -\frac{1}{2}x'_f\hat{\sigma}(X'X)^{-1}\sum_{i=1}^{n}(\hat{\epsilon}_i^2 - 1)x_i, \tag{19}$$

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

Figure 7b 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.

Maturity	Multivariate model	Random walk	Equal weight with const	w ∼ 1/RMSFE	w ∼ |S(AR1)|	w ∼ 1/|S(AR1)|	w ∼ |S(sk)|	w ∼ 1/|S(sk)|
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 2: RMSFE of forecast combinations with different weights for *the new methodology period*. The best performing weight for each maturity is shaded.

The right panel in Table 1 shows that the weights proportional to skewness sensitivity outperform other weights. This indicates that skewness is a very important feature for the forecast during the historical period.

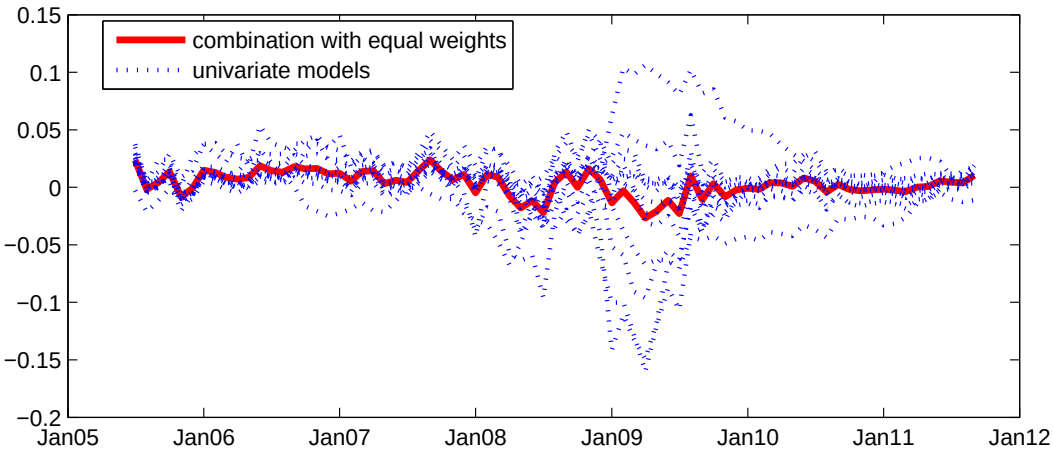
4.6 Further analysis

The results for the new methodology period (October 2004 – September 2011) are given in Table 2.

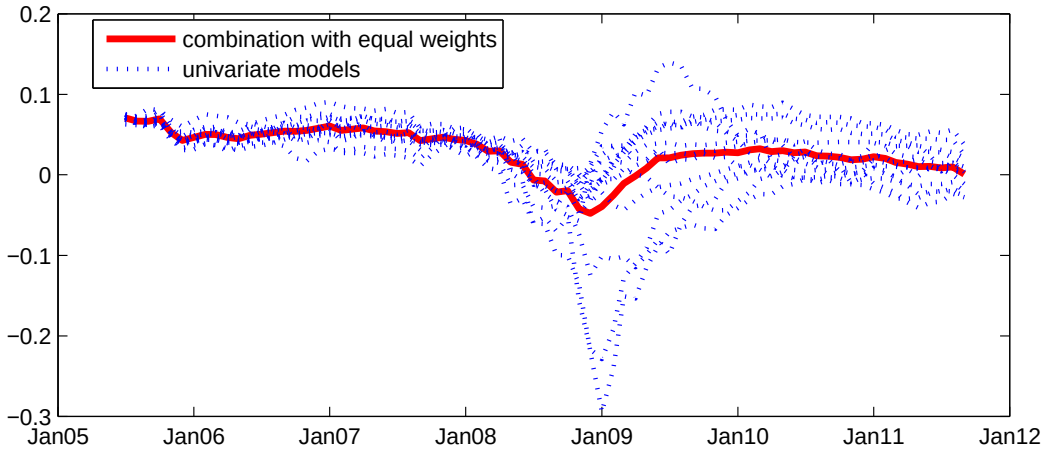
The weights inversely proportional to autocorrelation sensitivity perform best in most of the cases. The table also provides the benchmark models and shows that the forecast combination can be further improved.

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

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



(a) Sensitivity to AR(1).



(b) Sensitivity to skewness

Figure 8: Sensitivities of the univariate models and the equal weight combination for 3 year maturity during *the extended period*.

The results for the extended period (January 1999 – September 2011) are given in Table 3. They show the complementarity of sensitivity and fit-based weights. For maturities up to 9 years the fit is more important, while for maturities over 9 years autoregression sensitivity weights give the best result.

Sensitivities of univariate models and the equal-weight combination are given in Figures 8a and 8b. The AR(1) sensitivities show visibly more fluctuations in 2008 and destabilize in 2009. Skewness sensitivities declined during 2008 and destabilize dramatically in 2009. This highlights the forward-looking nature of sensitivity. The RMSFE only captures past performance, while sensitivity looks at the effect of additional features that might come into play. During the global financial crisis and after, the situation has changed and sensitivity was able to capture this change.

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

Table 4: RMSFE of forecast combinations with different weights for *the extended period* (with dummy variable and *random walk component*). The best performing weight for each maturity is shaded.

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

5 Concluding remarks

This paper has considered practical aspects of sensitivity analysis, and has identified two feasible approaches. First, if the magnitude is important, then context-dependent thresholds can be introduced that classify the models as 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, relative sensitivity gives an additional dimension for comparing the forecasts. It also gives a new possibility to choose the weights for forecast combinations. In our empirical illustration, the sensitivity-based weights often perform better than the fit-based weights. Although the results vary across different periods, we see that for long-term maturities the sensitivity-based weights perform better than other weights.

Acknowledgements

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 (MVTSF in Melbourne) for constructive comments, and Emma Smith for excellent research assistance at the initial stage of this project.

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 and Malta (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 in Magnus and Vasnev (2008) is excluded as it is no longer available. The unit root situation is addressed as in Magnus and Vasnev (2008) with year-on-year log differences.

HICP: Inflation is captured by the consumer price indices, which are measured for each country separately and further 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 males and females, which we denote as HUNE. Similar 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, which we denote 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: Stationary adjusted version of the industrial production index for the total industry excluding construction we denote as IND.

DR: The relation between the official deposit rate (DR), the official refinancing operation rate (REF), and the official lending rate (LOAN) is kept fixed (with the exception January–March 1999 when the gap between REF and LOAN was 1.5%) by the Central Bank with a gap of 1%. 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 high 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 spread between 2 and 1 year yields ($\text{SPRD}_{2,1}$), between 5 and 2 years ($\text{SPRD}_{5,2}$), and between 10 and 5 years ($\text{SPRD}_{10,5}$).

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. and J. R. Magnus (1999). The sensitivity of ols when the variance matrix is (partially) unknown. *Journal of Econometrics 92*, 295–323.

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

Duan, N. (1993). Sensitivity analysis for box-cox power transformation model: Contrast parameters. *Biometrika 80*(4), 885–897.

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

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

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

Piazzesi, M. and E. Swanson (2004). Futures prices as risk-adjusted forecast of monetary policy. *NBER Working paper 10547*.

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

Qin, H., A. T. K. Wan, and G. Zou (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. and M. W. Watson (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. Granger, and A. Timmermann (Eds.), *Handbook of Economic Forecasting, Volume 1. Handbook of Economics 24*, pp. 135–196. Elsevier, Horth-Holland.

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

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

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