DEUTSCHE
BUNDESBANK
EUROSYSTEM

Efficient estimation of forecast uncertainty based on recent forecast errors

Malte Knüppel

Discussion Paper
Series 1: Economic Studies
No 28/2009

Deutsche Bundesbank, Wilhelm-Epstein-Strasse 14, 60431 Frankfurt am Main,
Postfach 10 06 02, 60006 Frankfurt am Main

Tel +49 69 9566-0
Telex within Germany 41227, telex from abroad 414431

Please address all orders in writing to: Deutsche Bundesbank,
Press and Public Relations Division, at the above address or via fax +49 69 9566-3077

Internet http://www.bundesbank.de

ISBN 978-3–86558–564–6 (Printversion)
ISBN 978-3–86558–565–3 (Internetversion)

Abstract:

Multi-step-ahead forecasts of forecast uncertainty in practice are often based on the horizon-specific sample means of recent squared forecast errors, where the number of available past forecast errors decreases one-to-one with the forecast horizon. In this paper, the efficiency gains from the joint estimation of forecast uncertainty for all horizons in such samples are investigated. Considering optimal forecasts, the efficiency gains can be substantial if the sample is not too large. If forecast uncertainty is estimated by seemingly unrelated regressions, the covariance matrix of the squared forecast errors does not have to be estimated, but simply needs to have a certain structure. In Monte Carlo studies it is found that seemingly unrelated regressions mostly yield estimates which are more efficient than the sample means even if the forecasts are not optimal. Seemingly unrelated regressions are used to address questions concerning the inflation forecast uncertainty of the Bank of England.

Keywords: multi-step-ahead forecasts, forecast error variance, GLS, SUR

JEL-Classification: C13, C32, C53

Non-technical summary

In recent years, it has become more and more common to publish not only point forecasts for major economic variables, but also uncertainty forecasts. Examples are the fan charts of the Bank of England, the prediction intervals of the Eurosystem staff macroeconomic projections, or the uncertainty margins of the projections of the Deutsche Bundesbank. In all cases, the width of the published intervals conveys information about the probability that the future value of the forecast variable will lie within a certain range. The forecasts are often made for several periods ahead, and the forecast uncertainty typically increases with the forecast horizon. A reliable assessment concerning the precision of a forecast matters for example if decision makers like central banks want to avoid passing certain thresholds for inflation. One also needs to have knowledge about the forecast uncertainty of inflation to be able to determine, for example, risk premia of nominal bonds.

The future forecast uncertainty is often estimated based on past forecast errors. For this purpose, the squared (or absolute) values of the past errors are calculated, and for every forecast horizon the sample mean of these values is determined. These estimated values for the past forecast uncertainty can then be used to assess the future forecast uncertainty.

In this work, it is investigated whether it is possible to estimate the expected values of the past squared forecast errors more precisely than by their sample means. The idea that this might be possible is based on the fact that the forecast errors for different horizons are correlated with each other. If, for example, the inflation rate in a certain period attains an exceptionally high value, then generally the short- and the long-run forecasts will underestimate this value. Therefore, if the expected values of the past squared forecast errors are not estimated separately for each horizon, but jointly for all horizons, these correlations can be exploited to achieve a more precise estimation.

Typically, a prerequisite for such an improvement is a sufficiently exact estimation of the correlations mentioned. In this work, it is found that the estimation method based on seemingly unrelated regressions (SUR) often leads to improvements in the precision of the estimation, above all if the forecast horizon is large,

and if the available time series are relatively short. This result, however, is not due to the exact estimation of the correlations, but due to the surprising result that with SUR estimation, optimal forecasts, and the typical data structure of past forecast errors, the correlations can be determined arbitrarily. Also if the forecasts are not optimal, SUR estimation leads to improvements in most cases.

Using the inflation forecast errors of the Bank of England it is shown that SUR estimation can lead to markedly lower and more plausible values for the forecast uncertainty than the calculation of the horizon-specific sample means.

Nicht-technische Zusammenfassung

In zunehmendem Maße werden heute für zentrale gesamtwirtschaftliche Größen zusätzlich zu Punktprognosen auch Prognosen für deren Unsicherheit veröffentlicht. Beispiele dafür sind die sogenannten Fan Charts der Bank von England, die Prognoseintervalle der Stabsprognosen des Europäischen Systems der Zentralbanken oder die Unsicherheitsmargen der Prognosen der Deutschen Bundesbank. In allen Fällen vermittelt die Breite des veröffentlichten Intervalls eine Einschätzung darüber, mit welcher Wahrscheinlichkeit der zukünftige Wert der prognostizierten Variable innerhalb bestimmter Grenzen liegen wird. Dabei werden Prognosen oft für mehrere Perioden im Voraus erstellt, wobei die Unsicherheit üblicherweise mit dem Prognosehorizont ansteigt. Eine möglichst zuverlässige Einschätzung darüber, wie präzise eine Voraussage ist, ist zum Beispiel dann wichtig, wenn Entscheidungsträger wie Zentralbanken mit einiger Sicherheit ausschließen wollen, dass ein bestimmter Schwellenwert der Inflationsrate über- oder unterschritten wird. Außerdem ist eine Vorstellung über die Prognoseunsicherheit für die Inflation auch notwendig, um Risikoprämien zum Beispiel von nominalen Anleihen bestimmen zu können.

Die zukünftige Prognoseunsicherheit wird oft auf der Basis vergangener Prognosefehler geschätzt. Dazu werden die quadrierten (oder absoluten) Werte der vergangenen Fehler gebildet, und für jeden Prognosehorizont wird der Mittelwert aus diesen Werten errechnet. Auf der Grundlage dieser Schätzwerte für die Prognoseunsicherheit in der Vergangenheit kann dann die künftige Unsicherheit geschätzt werden.

In dieser Arbeit wird der Frage nachgegangen, ob es möglich ist, die Erwartungswerte der quadrierten vergangenen Prognosefehler präziser als durch die Mittelwerte zu schätzen. Ausgangspunkt dieser Überlegung ist dabei die Beobachtung, dass die Prognosefehler für verschiedene Prognosehorizonte miteinander korreliert sind. Wenn zum Beispiel die Inflationsrate in einer Periode einen ungewöhnlich hohen Wert annimmt, so werden üblicherweise sowohl die kurzfristigen, als auch die langfristigen Prognosen diesen Wert unterschätzen. Wenn man dementsprechend die Erwartungswerte der quadrierten Prognosefehler nicht - wie üblich - für jeden Prognosehorizont einzeln, sondern für alle Horizonte gemeinsam schätzt, so kann man diese Korrelation ausnutzen, um zu einer präziseren Schätzung zu gelangen.

Bedingung für eine solche Verbesserung ist allerdings gewöhnlich eine hinreichend genaue Schätzung der erwähnten Korrelationen. Diese Arbeit kommt zu dem Ergebnis, dass das Schätzverfahren der sogenannten scheinbar unverbundenen Regressionen (SUR) in vielen Fällen zu Verbesserungen der Schätzgenauigkeit führt, vor allem wenn der Prognosehorizont groß und die Zeitreihen, auf die man sich stützen kann, vergleichsweise kurz sind. Dies liegt allerdings nicht an der genauen Schätzung der Korrelationen, sondern an der überraschenden Tatsache, dass die Korrelationen für die SUR-Schätzung bei optimalen Prognosen und der üblicherweise gegebenen Datenstruktur vergangener Prognosefehler arbiträr bestimmt werden dürfen. Auch bei nicht optimalen Prognosen ergibt sich vielfach eine Überlegenheit der SUR-Schätzung.

An einem Beispiel der Inflationsprognosefehler der Bank von England wird gezeigt, dass die SUR-Schätzung zu deutlich geringeren und plausibleren Werten für die Prognoseunsicherheit führen kann als die einfache Berechnung der Mittelwerte für die einzelnen Prognosehorizonte.

Contents

1 Introduction **1**

2 Optimal Forecasts **4**

2.1 The Covariances of Squared Forecast Errors 5

2.2 The Data Structure of Recent Forecast Errors 7

3 Efficient Estimation of Forecast Uncertainty **8**

4 SUR Estimation of Forecast Uncertainty **13**

5 Properties of the GLS and SUR Estimator **16**

5.1 The Parameter-(In)dependence of the GLS and SUR Estimator . . 18

5.2 The Recursiveness of the GLS and SUR Estimator 20

6 Problems in Practice **24**

6.1 Unknown Covariance Matrix 24

6.2 Non-Optimal Forecasts 26

6.2.1 Bias 27

6.2.2 AR(1)-Processes 29

6.2.3 AR(2)-Processes 30

6.2.4 Structural Breaks 31

7 Applications to the Bank of England's Inflation Forecasts **32**

7.1 The Uncertainty About the 2- to 3-Year-Ahead Forecasts 32

7.2 The Width of the RPIX Inflation Fan Charts 34

8 Conclusion **36**

A Appendix **41**

A.1 Covariance of Squared Forecast Errors 41

A.2 Parameter Independence of the SUR Estimator 45

A.3 FGLS Estimation 47

Efficient Estimation of Forecast Uncertainty Based on Recent Forecast Errors[1]

1 Introduction

In recent years, many forecasting institutions have supplemented their point forecasts with measures of forecast uncertainty. That is, they have not only forecast the central tendency but also some measure of dispersion of the forecast density which is communicated, for example, by the width of fan charts. Examples of such institutions include the Bank of England (BoE), the Bank of Canada, the International Monetary Fund, the Sveriges Riksbank, the Norges Bank, the United States Congressional Budget Office and the Deutsche Bundesbank. The European Central Bank only reports a forecast range, so that actually only the exact forecast uncertainty, but not the exact central tendency is presented. All institutions mentioned publish forecasts for several periods ahead.

As stated by Wallis (1989, p. 56), "Estimating the future margin of error is itself a forecasting problem". When investigating uncertainty forecasts, researchers typically start by considering a general forecasting model. Within this model, they identify different sources of forecast uncertainty like estimation uncertainty and the accumulation of future errors. Then the uncertainty of the forecasts can be determined as the aggregate impact of these sources. Examples of this approach for assessing forecast uncertainty can be found e.g. in Clements and Hendry (1998, ch. 7) and Ericsson (2002).

However, as noted by Wallis (1989, pp. 55-56),"This approach is of little help to the practitioner. It neglects the contribution of the forecaster's subjective adjustments [...]. More fundamentally, the model's specification is uncertain. At any point in time competing models coexist, over time model specifications evolve, and there is no way of assessing this uncertainty. *Thus, the only practical indication of*

[1]The views expressed in this paper are those of the author and do not necessarily represent those of the Deutsche Bundesbank. The paper has benefited from valuable comments by Jörg Breitung, Karl-Heinz Tödter, Katrin-Assenmacher-Wesche and participants of the conference on "Forecasting and Monetary Policy" in Berlin, 2009. Corresponding author: malte.knueppel@bundesbank.de

the likely margin of future error is provided by the past forecast errors" [emphasis added]. Interestingly, all the forecasting institutions mentioned above indeed base their assessment of forecast uncertainty on past forecast errors.[2] However, surprisingly, the estimation of forecast uncertainty based on past forecast errors has hardly been investigated in the literature yet. A notable exception can be found in Williams and Goodman (1971).

The calculation of forecast uncertainty from past forecast errors can be performed in an extremely simple way. First, one collects all forecast errors for each forecast horizon. Then one performs a suitable transformation on these errors, reflecting the measure of dispersion to be reported. Typically, this means either taking absolute values or squared values of the forecast errors. In this work, I will focus on squared errors. For each horizon, the sample mean of the squared errors is calculated, i.e. an ordinary least squares (OLS) regression of the squared errors on a constant is performed. Although the estimation procedure consists simply of the calculation of horizon-specific sample means, I will refer to it as OLS estimation in order to contrast it with GLS and SUR estimation later on. OLS estimation yields consistent estimates of forecast uncertainty. It is apparently used by all the institutions mentioned above.[3] Yet, since the forecast errors are correlated across horizons, this procedure is not efficient.

This inefficiency is particularly pronounced for larger forecast horizons in small samples for two reasons. Firstly, the autocorrelation of forecast errors typically increases with the forecast horizon, so that estimates for large horizons tend to be rather imprecise. Secondly, the number of available forecast errors often decreases with the horizon. This is due to the fact that for the most recent forecasts, only the forecast errors for small horizons can be calculated, because only for these horizons realizations are available. If the frequency of forecasts publications equals

[2] The European Central Bank in its Monthly Bulletins as of September 2008 calculates forecast uncertainty based on a Bayesian VAR. However, this uncertainty is conditional on future paths for several exogenous variables like oil prices and exchange rates. Hence, in contrast to earlier publications and to all other institutions mentioned, the reported uncertainty is not a measure of unconditional forecast uncertainty. The Bank of Canada uses uncertainty estimated based on past forecast errors for smaller horizons and based on a model for larger horizons.

[3] Based on the estimated forecast uncertainty, in many cases prediction intervals covering a certain probability of the forecast density are calculated. These prediction intervals of course require distributional assumptions for the forecast errors.

the frequency of the forecast variables, the number of forecast errors decreases one-to-one with the forecast horizon. I will refer to such samples of forecast errors as samples of *recent* forecast errors.

Samples of recent forecast errors are frequently used in practice to estimate forecast uncertainty. These samples are present if a forecasting institution uses all forecast errors from the introduction of a new forecasting regime to the present. For example, since February 1998, the BoE has published quarterly forecasts based on market interest rates instead of constant interest rates. Suppose you want to use past forecast errors to assess the future forecast uncertainty for these forecasts, and that the last available realization comes from the fourth quarter of 2008 (henceforth 2008q4). If the forecasts contain a nowcast for the current quarter, then one can calculate 44 forecast errors for the nowcast (based on the forecasts from 1998q1 to 2008q4), 43 forecast errors for the 1-quarter-ahead forecast (based on the forecasts from 1998q1 to 2008q3), 42 forecast errors for the 2-quarter-ahead forecast (based on the forecasts from 1998q1 to 2008q2) etc. Thus, using all available forecast errors since February 1998 for an assessment of forecast uncertainty would mean using a sample with recent forecast errors.

In this work, I derive the covariance matrix of squared multi-step-ahead forecast errors under the assumption of optimal forecasts. In addition to the efficient estimator, i.e. the generalized least squares (GLS) estimator of forecast uncertainty, I also consider the estimator based on seemingly unrelated regressions (SUR estimator).[4]

The small-sample gains in forecast efficiency of these estimators are investigated for samples of recent forecast errors. It turns out that they have two important and at least partly surprising properties:

- The projection matrix of the GLS estimator does not depend on the distribution of the error terms of the data-generating process (DGP).
- The projection matrix of the SUR estimator does not depend on the DGP at all. The covariance matrix for the SUR estimator therefore does not need to

[4]The literature on GLS and SUR estimation with unequal number of observations is scarce. Concerning SUR estimation, this case is studied by Schmidt (1977) and Im (1994).

be estimated, but simply requires a certain structure. This is an intriguing property in small samples.

In practice, most forecasts are probably non-optimal. Therefore, the performance of the GLS and the SUR estimator is studied for such forecasts as well. Since the GLS estimator does not work well, I use a shrinkage estimator (henceforth SGLS estimator) instead, which shrinks the GLS estimates towards the OLS estimates. Monte Carlo studies show that only if the forecasting model is severely misspecified, the OLS estimator can sometimes be more efficient than the SUR and the SGLS estimator. However, in most cases studied, the SUR and the SGLS estimator yield better results, often even in cases of severe misspecification. The efficiency gains of the SUR estimator are typically larger than those of the SGLS estimator, so that the SUR estimator seems preferable.

Finally I apply the SUR estimator to the BoE's inflation forecasts for the consumer price index. The SUR estimator indicates that forecast uncertainty for the largest forecast horizons is likely to be strongly overestimated by OLS. Moreover, the claim of Clements (2004), Wallis (2004) and others that the BoE's fan charts fan out too quickly is investigated. It is found that this result is probably not related to an inefficient estimation of forecast uncertainty.

2 Optimal Forecasts

Every stationary DGP has a Wold-representation given by

$$y_t = \mu + \sum_{i=0}^{\infty} b_i \varepsilon_{t-i} \tag{1}$$

with $E[\varepsilon_t] = 0$, $E[\varepsilon_t^2] = \sigma^2$ and $b_0 = 1$. Here it is assumed that the fourth moment of ε_t exists, so that the kurtosis

$$\alpha = E\left[\varepsilon_t^4\right] / \sigma^4$$

is finite.

The optimal h-step-ahead forecast is

$$y_{t+h,t} = \mu + \sum_{i=0}^{\infty} b_{h+i}\varepsilon_{t-i}.$$

Hence, the h-step-ahead forecast error equals

$$e_{t+h,t} := y_t - y_{t+h,t} = \sum_{i=0}^{h-1} b_i \varepsilon_{t+h-i}. \tag{2}$$

Thus, $e_{t+h,t}$ is the error of the forecast made in period t for period $t+h$, and has a moving-average representation of order $h-1$ (henceforth MA($h-1$)-process).[5]

The variance of the h-step-ahead forecast error is given by

$$E\left(e_{t+h,t}^2\right) = \sigma_h^2 = \sigma^2 \sum_{i=0}^{h-1} b_i^2.$$

The variances for all forecast horizons are collected in the vector $\boldsymbol{\sigma}^2$, so that

$$\boldsymbol{\sigma}^2 = \left(\sigma_1^2, \sigma_2^2, \ldots, \sigma_H^2\right)',$$

where H denotes the largest forecast horizon. The estimates of forecast uncertainty will be denoted as

$$\hat{\boldsymbol{\sigma}}_m^2 = \left(\hat{\sigma}_{m,1}^2, \hat{\sigma}_{m,2}^2, \ldots, \hat{\sigma}_{m,H}^2\right)'$$

where m will refer to the estimation method used.

2.1 The Covariances of Squared Forecast Errors

From (2) it is obvious that the forecast errors e_{t_2,t_1} and e_{t_4,t_3} can be correlated, and the same holds for the squared forecast errors. In order to determine the covariance of two forecast errors $E\left[\left(e_{t_2,t_1}^2 - E\left(e_{t_2,t_1}^2\right)\right)\left(e_{t_4,t_3}^2 - E\left(e_{t_4,t_3}^2\right)\right)\right]$ with

[5] In case of optimal forecasts, stationarity of y_t is most likely unnecessary to obtain the expression for $e_{t+h,t}$ given above. However, the derivations of this expressions found in the literature as e.g. Patton and Timmermann (2007) and Diebold (1998, p. 341) always start from a stationary process for y_t.

$t_2 > t_1$, $t_4 > t_3$, define

$$\begin{aligned} m &= t_3 - t_1 \\ p &= t_2 - t_1 \\ q &= t_4 - t_3 \\ r &= \max(1, m+1) \\ s &= \min(p, m+q). \end{aligned} \tag{3}$$

Then the covariance can be expressed as

$$\begin{aligned} & E\left[\left(e^2_{t_2,t_1} - E\left(e^2_{t_2,t_1}\right)\right)\left(e^2_{t_4,t_3} - E\left(e^2_{t_4,t_3}\right)\right)\right] \\ &= \begin{cases} 0 & \text{if } s < r \\ \sigma^4(\alpha-1)\sum\limits_{i=0}^{s-r} b_i^2 b_{k-s+i}^2 + 2\sigma^4 \sum\limits_{i=0}^{s-r}\sum\limits_{j=0,j\neq i}^{s-r} b_i b_{k-s+i} b_j b_{k-s+j} & \text{if } s \geq r \end{cases} \\ \text{with } k &= \max(p, m+q). \end{aligned} \tag{4}$$

Here I use the convention $\sum_{i=0}^{0}\sum_{j=0,j\neq i}^{0} x_{ij} = 0$. The derivation of equation (4) is shown in Appendix A.1.

If the kurtosis equals $\alpha = 3$ (e.g. in case of a normal distribution)[6], expression (4) simplifies to

$$\begin{aligned} & E\left[\left(e^2_{t_2,t_1} - E\left(e^2_{t_2,t_1}\right)\right)\left(e^2_{t_4,t_3} - E\left(e^2_{t_4,t_3}\right)\right)\right] \\ &= \begin{cases} 0 & \text{if } s < r \\ 2\sigma^4 \sum\limits_{i=0}^{s-r}\sum\limits_{j=0}^{s-r} b_i b_{k-s+i} b_j b_{k-s+j} & \text{if } s \geq r \end{cases}. \end{aligned} \tag{5}$$

If, in addition, the true data-generating process is a first-order autoregressive process (henceforth AR(1)-process)

$$y_t = \rho y_{t-1} + \varepsilon_t,$$

[6] Note that no distribution with $\alpha < 1$ appears to be known. The smallest possible kurtosis seems to be $\alpha = 1$ for the discrete uniform distribution with 2 possible values. So the first term of the covariance of squared foreacst errors can apparently not be negative. Most distributions with infinite support have $\alpha \geq 3$. However, the covariance can of course be negative if the second term is negative.

like in many of the following examples, so that $b_i = \rho^i$, (5) simplifies to

$$E\left[\left(e_{t_2,t_1}^2 - E\left(e_{t_2,t_1}^2\right)\right)\left(e_{t_4,t_3}^2 - E\left(e_{t_4,t_3}^2\right)\right)\right]$$
$$= \begin{cases} 0 & \text{if } s < r \\ 2\sigma^4 \frac{\rho^{2(k-s)}}{(\rho^2-1)^2}\left(\rho^{2(s-r+1)} - 1\right)^2 & \text{if } s \geq r \text{ and } \rho \notin \{-1, 1\} \\ 2\sigma^4 (s-r+1)^2 & \text{if } s \geq r \text{ and } \rho \in \{-1, 1\} \end{cases}.$$

2.2 The Data Structure of Recent Forecast Errors

Suppose that the first forecast was made in period $T_1 - 1$, and that the last available realization comes from period T_2. Defining $N := T_2 - T_1 + 1$, the sample of all foreacst errors then contains N 1-step ahead forecast errors, $N - 1$ 2-step-ahead forecast errors etc. Obviously, it is required that $N \geq H$, so that there is at least one H-step-ahead forecast error.

Let the vector of squared 1-step-ahead forecast errors be

$$\mathbf{e}_1^2 = \left(e_{T_1,T_1-1}^2, e_{T_1+1,T_1}^2, \ldots, e_{T_2,T_2-1}^2\right),$$

the vector of squared 2-step-ahead forecast errors be

$$\mathbf{e}_2^2 = \left(e_{T_1+1,T_1-1}^2, e_{T_1+2,T_1}^2, \ldots, e_{T_2,T_2-2}^2\right),$$

etc. Then define the column vector with all squared forecast errors as

$$\mathbf{e}^2 = \left(\mathbf{e}_1^2, \mathbf{e}_2^2, \ldots, \mathbf{e}_H^2\right)'. \tag{6}$$

This vector has $\frac{1}{2}H(2N - H + 1)$ elements.

As an example, suppose that the forecasts started to be made in period 0 (i.e. $T_1 = 1$), and that the largest forecast horizon is 2 ($H = 2$), i.e. that 1- and 2-step ahead forecasts were produced. Further, suppose that the last realization observed is y_3 (i.e. $T_2 = 3$). This gives three 1-step-ahead forecast errors and two 2-step-ahead forecast errors. The available squared forecast errors are presented in Table 1.

		forecast horizon	
		1	2
forecast	0	$e^2_{1,0}$	$e^2_{2,0}$
from	1	$e^2_{2,1}$	$e^2_{3,1}$
period	2	$e^2_{3,2}$	

Table 1: An example data set

The (5×1) vector $\mathbf{e}^2$ is then given by

$$\mathbf{e}^2 = \left(\mathbf{e}_1^2, \mathbf{e}_2^2\right)' = \left(\left(e^2_{1,0}, e^2_{2,1}, e^2_{3,2}\right), \left(e^2_{2,0}, e^2_{3,1}\right)\right)'.$$

3 Efficient Estimation of Forecast Uncertainty

In order to estimate a model with correlated error terms efficiently, i.e. by GLS, one needs to know the covariance matrix of the error terms. Since the regressors are constants, the covariance matrix of the error terms equals the covariance matrix of the dependent variables which is given by

$$\mathbf{\Omega} = E\left[\left(\mathbf{e}^2 - E\left[\mathbf{e}^2\right]\right)\left(\mathbf{e}^2 - E\left[\mathbf{e}^2\right]\right)'\right].$$

Consider the covariance matrix of the vector $\mathbf{e}^2$ defined in the example of the previous Section, i.e. with $N = 3$ and $H = 2$. For the sake of simplification, assume $\alpha = 3$. Then using (5) yields the desired values and gives the matrix

$$\mathbf{\Omega} = 2\sigma^4 \begin{bmatrix} 1 & 0 & 0 & b_1^2 & 0 \\ 0 & 1 & 0 & 1 & b_1^2 \\ 0 & 0 & 1 & 0 & 1 \\ b_1^2 & 1 & 0 & (1+b_1^2)^2 & b_1^2 \\ 0 & b_1^2 & 1 & b_1^2 & (1+b_1^2)^2 \end{bmatrix}. \tag{7}$$

In order to estimate forecast uncertainty, define the regressor matrix

$$\mathbf{X}=\begin{bmatrix} \mathbf{1}_N & \mathbf{0}_N & \mathbf{0}_N & \dots & \mathbf{0}_N \\ \mathbf{0}_{N-1} & \mathbf{1}_{N-1} & \mathbf{0}_{N-1} & \dots & \mathbf{0}_{N-1} \\ \mathbf{0}_{N-2} & \mathbf{0}_{N-2} & \mathbf{1}_{N-2} & \dots & \mathbf{0}_{N-2} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ \mathbf{0}_{N-H+1} & \mathbf{0}_{N-H+1} & \mathbf{0}_{N-H+1} & \dots & \mathbf{1}_{N-H+1} \end{bmatrix}$$

where $\mathbf{1}_n$ denotes an $(n \times 1)$ vector of ones and $\mathbf{0}_n$ denotes an $(n \times 1)$ vector of zeros. For the example given above, the (5×2) regressor matrix is given by

$$\mathbf{X}=\begin{bmatrix} 1 & 1 & 1 & 0 & 0 \\ 0 & 0 & 0 & 1 & 1 \end{bmatrix}'.$$

Having defined these matrices, the model to be estimated and the properties of its error terms can be defined as

$$\begin{aligned} \mathbf{e}^2 &= \mathbf{X}\boldsymbol{\sigma}^2+\mathbf{u} \\ E\left[\mathbf{u}\mathbf{u}'\right] &= \boldsymbol{\Omega}. \end{aligned}$$

The GLS estimator is given by

$$\hat{\boldsymbol{\sigma}}^2_{GLS}=\left(\mathbf{X}'\boldsymbol{\Omega}^{-1}\mathbf{X}\right)^{-1}\mathbf{X}'\boldsymbol{\Omega}^{-1}\mathbf{e}^2 \tag{8}$$

whereas the OLS estimator which yields the sample means equals

$$\hat{\boldsymbol{\sigma}}^2_{OLS}=\left(\mathbf{X}'\mathbf{X}\right)^{-1}\mathbf{X}'\mathbf{e}^2.$$

The covariance matrices of these estimators are

$$Cov\left(\hat{\boldsymbol{\sigma}}^2_{GLS}\right)=\left(\mathbf{X}'\boldsymbol{\Omega}^{-1}\mathbf{X}\right)^{-1}$$

and

$$Cov\left(\hat{\boldsymbol{\sigma}}^2_{OLS}\right)=\left(\mathbf{X}'\mathbf{X}\right)^{-1}\mathbf{X}'\boldsymbol{\Omega}\mathbf{X}\left(\mathbf{X}'\mathbf{X}\right)^{-1}.$$

The measure of efficiency gains of the GLS estimator used in this work is defined by

$$\boldsymbol{\varphi}_{GLS} = 100 \ln \left(\sqrt{\frac{diag\left(Cov\left(\hat{\boldsymbol{\sigma}}^2_{OLS}\right)\right)}{diag\left(Cov\left(\hat{\boldsymbol{\sigma}}^2_{GLS}\right)\right)}} \right)$$

where the fraction bar denotes elementwise division, $\sqrt{\bullet}$ denotes the elementwise square root, *diag* extracts the diagonal of a matrix, and $\boldsymbol{\varphi}_{GLS}$ is the $(H \times 1)$ vector $\boldsymbol{\varphi}_{GLS} = \left(\varphi_{GLS,1}, \varphi_{GLS,2}, \ldots, \varphi_{GLS,H}\right)'$. So values larger than 0 indicate efficiency gains. For example, a value of $\varphi_{GLS,3} = 20$ means that the standard deviation of the GLS estimator for σ_3^2 is 20% lower than that of the OLS estimator.

For the example given above, the efficiency gains only depend on b_1. They are displayed in Figure 1. Efficiency gains can be obtained for $h = 2$, but not for $h = 1$. A similar phenomenon is found by Im (1994) for the SUR estimator with unequal numbers of observations and identical regressors. He shows that efficiency gains can only be obtained for the variable with a smaller number of observations.

Note that the efficiency gains for $h = 2$ are not monotonous with respect to $|b_1|$, attaining the lowest value at $|b_1| = 1$, i.e. at the point where the MA(1)-process of the 2-step-ahead forecast errors switches from invertibility to non-invertibility.

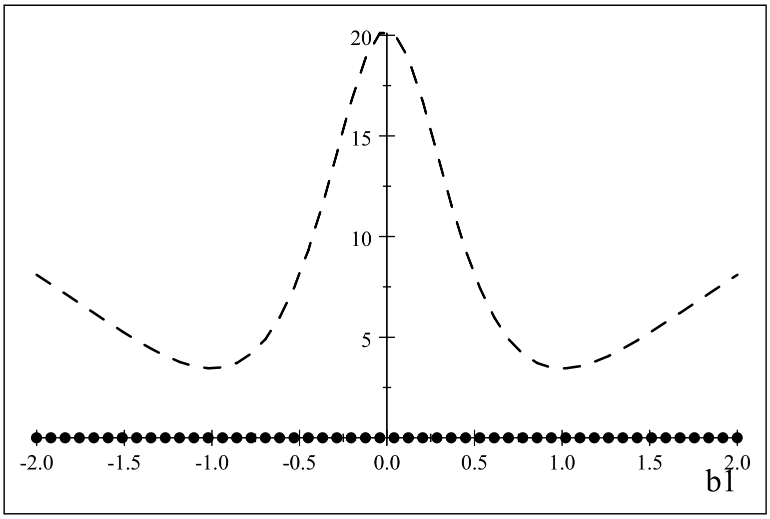

Figure 1: Efficiency gains $\varphi_{GLS,1}$ (circles) and $\varphi_{GLS,2}$ (dashed line) as functions of b_1 with $N = 3$, $H = 2$ and $\alpha = 3$

As Grenander and Rosenblatt (1957) show, OLS estimation is asymptotically as efficient as GLS estimation if the regressors consist exclusively of a constant. So, in the case presented here GLS estimation is interesting in small samples only. An important question is what a small sample means in this context. Since the answer depends on the DGP of the forecast errors and the number of forecast horizons, the question cannot be answered in general. However, an example might yield some insights.

Consider the macroeconomic forecasts of the BoE. Since August 2004, these are made for the current quarter and for the next 12 quarters, so $H = 13$. One of the variables to be forecast is real GDP growth in the UK. Estimating an AR(1)-process with a constant for quarterly real GDP growth from 1993q1 to 2008q3 yields $\rho = 0.42$ for the growth rate with respect to the previous quarter and $\rho = 0.88$ for the growth rate with respect to the previous year's quarter. Assuming that these are the true DGPs, the b_i's of (1) are simply given by $b_i = \rho^i$.

Considering $\alpha = 3$, the vectors $\boldsymbol{\varphi}_{GLS}$ displayed in Figures 2 and 3 are obtained. Obviously, the efficiency gains decrease with the number of available forecast errors, and they increase with the forecast horizon. As mentioned above, the latter has two reasons. Firstly, there are less forecast errors observed for larger horizons. Secondly, the autocorrelation of forecast errors increases with the forecast horizon. Both reasons lead to higher forecast uncertainty for larger horizons which can be reduced by GLS estimation.

The efficiency gains here also increase with decreasing persistence of the DGP. If the persistence is low ($\rho = 0.42$) and the sample is fairly small ($N = 20$, i.e. 5 years of data) the efficiency gains reach more than 40% for the 13-step-ahead forecast. Even if 28 13-step-ahead forecast errors are available (i.e. if $N = 40$), the efficiency gains for this horizon are still larger than 15%. However, if the sample exceeds 60 observations, the efficiency gains hardly exceed 10% even for the largest horizon.[7] So GLS estimation might be considered useful in situations where quarterly forecasts are made for up to 3 years ahead and the current forecasting regime has not been in place for more than 15 years.

[7] This result remains valid also if ρ is set to values lower than 0.42.

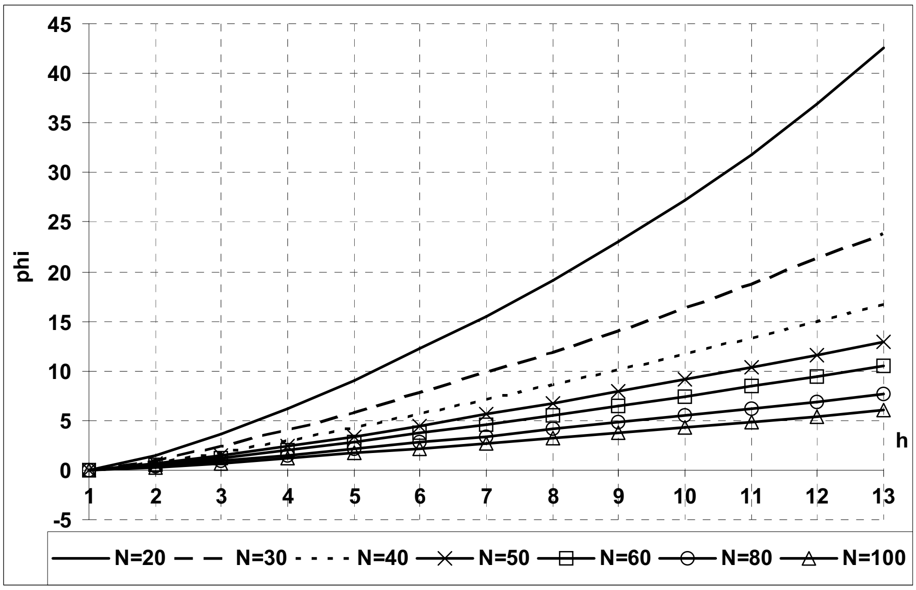

Figure 2: Efficiency gains of GLS versus OLS with $H = 13$, DGP is an AR(1)-process with $\rho = 0.42$.

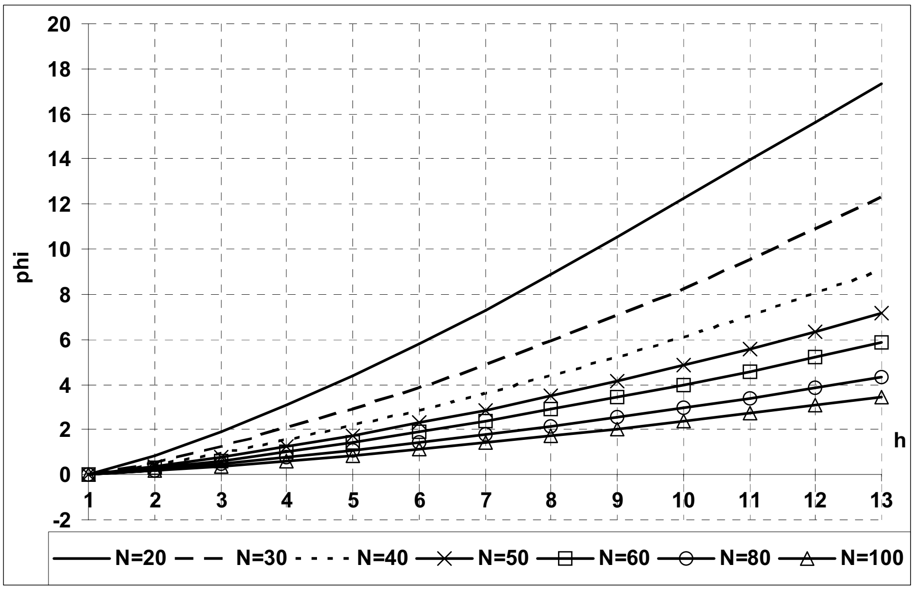

Figure 3: Efficiency gains of GLS versus OLS with $H = 13$, DGP is an AR(1)-process with $\rho = 0.88$.

4 SUR Estimation of Forecast Uncertainty

If the DGP is stationary, the covariance between the squared forecast errors $e^2_{t_1+n_1,t_1}$ and $e^2_{t_1+n_2,t_1}$ becomes very small when $|n_2 - n_1|$ becomes large. Actually, if all b_i's are smaller than 1 in absolute value, the largest possible covariance for two squared forecast errors $e^2_{t_1,t_1-n_1}$ and $e^2_{t_2,t_2-n_2}$ for given forecast horizons n_1 and n_2 is obtained when $t_1 = t_2$. Therefore, although some b_i's can be larger than 1 in absolute value if the DGP is stationary, a seemingly unrelated regression (SUR) estimation might be a promising approach for the estimation of forecast uncertainty.

The calculation of the individual covariances and the construction of the covariance matrices are relatively easy in this case. Define p and q as in (3), i.e. $p = t_2 - t_1$ and $q = t_4 - t_3$. Assuming that the SUR formulation is a good approximation, for the individual covariances one obtains

$$\begin{aligned} & E\left[\left(e^2_{t_2,t_1} - E\left(e^2_{t_2,t_1}\right)\right)\left(e^2_{t_4,t_3} - E\left(e^2_{t_4,t_3}\right)\right)\right] \\ \approx & \begin{cases} 0 & \text{if } t_2 \neq t_4 \\ \sigma^4(\alpha-1)\sum\limits_{i=0}^{s^*-1} b_i^4 + 2\sigma^4 \sum\limits_{i=0}^{s^*-1}\sum\limits_{j=0,j\neq i}^{s^*-1} b_i^2 b_j^2 & \text{if } t_2 = t_4 \end{cases} \\ \text{with } s^* = & \min(p,q). \end{aligned}$$

If the kurtosis equals $\alpha = 3$, this expression simplifies to

$$\begin{aligned} & E\left[\left(e^2_{t_2,t_1} - E\left(e^2_{t_2,t_1}\right)\right)\left(e^2_{t_4,t_3} - E\left(e^2_{t_4,t_3}\right)\right)\right] \\ \approx & \begin{cases} 0 & \text{if } t_2 \neq t_4 \\ 2\sigma^4 \sum\limits_{i=0}^{s^*-1}\sum\limits_{j=0}^{s^*-1} b_i^2 b_j^2 & \text{if } t_2 = t_4 \end{cases}. \end{aligned}$$

If, in addition, the true data-generating process is an AR(1)-process, this simplifies to

$$E\left[\left(e_{t_2,t_1}^2 - E\left(e_{t_2,t_1}^2\right)\right)\left(e_{t_4,t_3}^2 - E\left(e_{t_4,t_3}^2\right)\right)\right]$$
$$\approx \left\{ \begin{array}{cc} 0 & \text{if } t_2 \neq t_4 \\ 2\sigma^4 \frac{\left(\rho^{2(s^*+1)}-1\right)^2}{\left(\rho^2-1\right)^2} & \text{if } t_2 = t_4 \text{ and } \rho \notin \{-1,1\} \\ 2\sigma^4\left(s^*+1\right)^2 & \text{if } t_2 = t_4 \text{ and } \rho \in \{-1,1\} \end{array} \right. .$$

For the construction of the covariance matrix, it is helpful to define

$$g_{s^*} = \sigma^4\left(\alpha - 1\right)\sum_{i=0}^{s^*-1} b_i^4 + 2\sigma^4 \sum_{i=0}^{s^*-1}\sum_{j=0,j\neq i}^{s^*-1} b_i^2 b_j^2.$$

and

$$\mathbf{G}_{i,j} = \left[\begin{array}{cc} \mathbf{0}_{N-i,j} & g_{i-j+1}\mathbf{I}_{N-i} \end{array}\right]$$

where $\mathbf{0}_{n,m}$ denotes an $(n \times m)$ matrix of zeros, $\mathbf{I}_n$ denotes the identity matrix of size n, and with the convention that $\mathbf{G}_{i,0} = g_{i+1}\mathbf{I}_{N-i}$.

The covariance matrix for the SUR estimator is given by

$$\mathbf{\Omega}_{SUR} = \left[\begin{array}{ccccc} \mathbf{G}_{0,0} & \mathbf{G}_{1,1}' & \mathbf{G}_{2,2}' & \cdots & \mathbf{G}_{H-1,H-1}' \\ \mathbf{G}_{1,1} & \mathbf{G}_{1,0} & \mathbf{G}_{2,1}' & \cdots & \mathbf{G}_{H-1,H-2}' \\ \mathbf{G}_{2,2} & \mathbf{G}_{2,1} & \mathbf{G}_{2,0} & \cdots & \mathbf{G}_{H-1,H-3}' \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ \mathbf{G}_{H-1,H-1} & \mathbf{G}_{H-1,H-2} & \mathbf{G}_{H-1,H-3} & \cdots & \mathbf{G}_{H-1,0} \end{array}\right]$$

Thus, for the example given above with $N = 3$, $H = 2$ and $\alpha = 3$, the matrix $\mathbf{\Omega}_{SUR}$ equals

$$\mathbf{\Omega}_{SUR} = 2\sigma^4 \left[\begin{array}{ccccc} 1 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 1 & 0 \\ 0 & 0 & 1 & 0 & 1 \\ 0 & 1 & 0 & \left(1+b_1^2\right)^2 & 0 \\ 0 & 0 & 1 & 0 & \left(1+b_1^2\right)^2 \end{array}\right]. \tag{9}$$

ρ	N	$\boldsymbol{\varphi}'_{SUR}$								
		$h=1$	$h=2$	$h=3$	$h=4$	$h=5$	$h=6$	$h=7$	$h=8$	$h=9$
0.5	20	0.0	1.2	3.0	5.3	8.0	11.0	14.3	17.8	21.7
1.0	20	0.0	0.4	0.8	1.3	1.9	2.4	2.8	3.2	3.4
1.5	20	0.0	0.1	0.2	0.3	0.3	0.3	0.2	0.1	0.1
2.0	20	0.0	0.0	0.1	0.1	0.0	0.0	0.0	−0.0	−0.0

Table 2: Efficiency gains of the SUR estimator, varying ρ

The SUR estimator of forecast uncertainty is given by

$$\hat{\boldsymbol{\sigma}}^2_{SUR} = \left(\mathbf{X}'\boldsymbol{\Omega}^{-1}_{SUR}\mathbf{X}\right)^{-1}\mathbf{X}'\boldsymbol{\Omega}^{-1}_{SUR}\mathbf{e}^2. \tag{10}$$

The covariance matrix of the estimator $\hat{\boldsymbol{\sigma}}^2_{SUR}$ equals

$$Cov\left(\hat{\boldsymbol{\sigma}}^2\right) = \left(\mathbf{X}'\boldsymbol{\Omega}^{-1}_{SUR}\mathbf{X}\right)^{-1}\mathbf{X}'\boldsymbol{\Omega}^{-1}_{SUR}\boldsymbol{\Omega}\boldsymbol{\Omega}^{-1}_{\mathbf{SUR}}\mathbf{X}\left(\mathbf{X}'\boldsymbol{\Omega}^{-1}_{SUR}\mathbf{X}\right)^{-1}.$$

I define the efficiency gains of the SUR estimator as

$$\boldsymbol{\varphi}_{SUR} = 100\ln\left(\sqrt{\frac{diag\left(Cov\left(\hat{\boldsymbol{\sigma}}^2_{OLS}\right)\right)}{diag\left(Cov\left(\hat{\boldsymbol{\sigma}}^2_{SUR}\right)\right)}}\right)$$

where $\boldsymbol{\varphi}_{SUR}$ is an $(H \times 1)$ vector. It is clear that, in contrast to $\boldsymbol{\varphi}_{GLS}$, the elements of $\boldsymbol{\varphi}_{SUR}$ do not have to be greater than or equal to zero. The signs of the elements of $\boldsymbol{\varphi}_{SUR}$ depend on $\boldsymbol{\Omega}_{SUR}$ and $\boldsymbol{\Omega}$. In the light of the considerations at the beginning of this section, one would suppose that the OLS estimator could have a smaller variance than the SUR estimator if the b_i's are large in absolute value.

To investigate this possibility, I set $\alpha = 3$ and $\rho \in \{0.5, 1.0, 1.5, 2.0\}$ with $b_i = \rho^i$. The latter two of these values imply a strongly exploding forecast uncertainty. Moreover, $H = 9$ and $N = 20$. The results for $\boldsymbol{\varphi}_{SUR}$ are displayed in Table 2. Surprisingly, they show that even if forecast uncertainty explodes, the SUR estimator yields mostly smaller variances than the OLS estimator. Actually, only for $\rho = 2$ and $h \in \{8, 9\}$, the OLS estimator has a marginally smaller variance.

In order to investigate this further, N is varied in the following calculations. ρ is set to 2, all other parameters remain the same. The efficiency gains and

ρ	N	φ'_{SUR}								
		$h=1$	$h=2$	$h=3$	$h=4$	$h=5$	$h=6$	$h=7$	$h=8$	$h=9$
2.0	12	0.0	-0.0	-0.0	-0.1	-0.2	-0.3	-0.5	-0.9	-1.4
2.0	15	0.0	-0.0	-0.0	-0.0	-0.0	-0.1	-0.1	-0.2	-0.4
2.0	30	0.0	0.0	0.1	0.1	0.1	0.1	0.1	0.1	0.1

Table 3: Efficiency gains of the SUR estimator, varying N

losses displayed in Table 3 emerge. Apparently, the SUR estimator can have a marginally larger variance than the OLS estimator if the sample is very short and if the process of the forecast errors is strongly explosive, which is very unlikely to be a relevant case in practice.

It is also interesting to study the differences between the efficiency gains of the SUR and the GLS estimator. These are displayed for several values of ρ in Figure 4. For values of ρ smaller than or equal to 0.5, the SUR estimator is practically as efficient as the GLS estimator. For values of ρ around 0.75, differences become noticeable, but remain small. Even for $\rho = 0.9$, the efficiency gains of the SUR estimator equal more than half of those of the GLS estimator for all horizons. If $\rho = 1.0$, the efficiency gains of the SUR estimator reduce to about 40% of those of the GLS estimator. So, unless the DGP is very persistent, the SUR estimator performs almost as well as the GLS estimator.

5 Properties of the GLS and SUR Estimator

The GLS and the SUR estimator have interesting properties, which have partly become obvious in the preceding Section already. A striking feature from Figures 2, 3 and 4 is given by the equality of the variances of the OLS and the GLS estimator for the smallest forecast horizon. For the following investigation of issues like these, it is helpful to define the matrices

$$\boldsymbol{A} = \left(\mathbf{X}'\boldsymbol{\Omega}^{-1}\mathbf{X}\right)^{-1}\mathbf{X}'\boldsymbol{\Omega}^{-1}$$

and

$$\boldsymbol{A}_{SUR} = \left(\mathbf{X}'\boldsymbol{\Omega}_{SUR}^{-1}\mathbf{X}\right)^{-1}\mathbf{X}'\boldsymbol{\Omega}_{SUR}^{-1}.$$

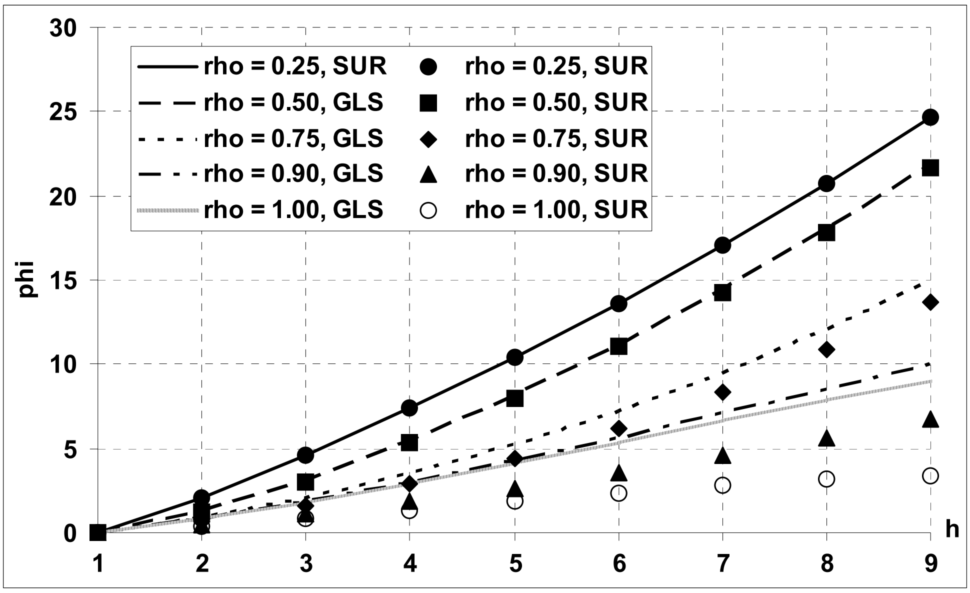

Figure 4: Efficiency gains of GLS and SUR versus OLS with $N = 20$ and $N = 9$, DGP is an AR(1)-process

These matrices are used for the calculations of the respective estimates in (8) and (10). They are multiplied with the vector of squared forecast errors $\mathbf{e}^2$. So the GLS projection matrix is given by $\mathbf{X}'\boldsymbol{A}$, and the SUR projection matrix by $\mathbf{X}'\boldsymbol{A}_{SUR}$.

Taking the example from above with $N = 3$, $H = 2$ and $\alpha = 3$, $\boldsymbol{A}$ equals

$$\boldsymbol{A} = \begin{bmatrix} \frac{1}{3} & \frac{1}{3} & \frac{1}{3} & 0 & 0 \\ \frac{1}{6}\left(2 - b_1^2\right) & -\frac{1}{6}\left(b_1^2 + 1\right) & \frac{1}{6}\left(2b_1^2 - 1\right) & \frac{1}{2} & \frac{1}{2} \end{bmatrix}. \quad (11)$$

For $\boldsymbol{A}_{SUR}$ one finds

$$\boldsymbol{A}_{SUR} = \begin{bmatrix} \frac{1}{3} & \frac{1}{3} & \frac{1}{3} & 0 & 0 \\ \frac{1}{3} & -\frac{1}{6} & -\frac{1}{6} & \frac{1}{2} & \frac{1}{2} \end{bmatrix}. \quad (12)$$

5.1 The Parameter-(In)dependence of the GLS and SUR Estimator

Up to now, only the special case where $\alpha = 3$ was considered. If this restriction is lifted, the covariance matrix of the GLS estimator with $N = 3$ and $H = 2$ becomes

$$\boldsymbol{\Omega} = \sigma^4 \begin{bmatrix} \alpha-1 & 0 & 0 & (\alpha-1)\,b_1^2 & 0 \\ 0 & \alpha-1 & 0 & \alpha-1 & (\alpha-1)\,b_1^2 \\ 0 & 0 & \alpha-1 & 0 & \alpha-1 \\ (\alpha-1)\,b_1^2 & \alpha-1 & 0 & (\alpha-1)\,(1+b_1^4)+4b_1^2 & (\alpha-1)\,b_1^2 \\ 0 & (\alpha-1)\,b_1^2 & \alpha-1 & (\alpha-1)\,b_1^2 & (\alpha-1)\,(1+b_1^4)+4b_1^2 \end{bmatrix}. \tag{13}$$

Note that the term $\alpha - 1$ cannot be factored out. However, calculating $\boldsymbol{A}$, surprisingly, yields the same result as with $\alpha = 3$. That is, the matrix $\boldsymbol{A}$ is again given by (11).

The covariance matrix of the SUR estimator with $N = 3$ and $H = 2$ and α unrestricted becomes

$$\boldsymbol{\Omega}_{SUR} = \sigma^4 \begin{bmatrix} \alpha-1 & 0 & 0 & 0 & 0 \\ 0 & \alpha-1 & 0 & \alpha-1 & 0 \\ 0 & 0 & \alpha-1 & 0 & \alpha-1 \\ 0 & \alpha-1 & 0 & (\alpha-1)\,(1+b_1^4)+4b_1^2 & 0 \\ 0 & 0 & \alpha-1 & 0 & (\alpha-1)\,(1+b_1^4)+4b_1^2 \end{bmatrix}.$$

Again, the term $\alpha - 1$ cannot be factored out. But again, calculating $\boldsymbol{A}_{SUR}$ gives the same result as with $\alpha = 3$, so that the matrix $\boldsymbol{A}_{SUR}$ is equal to (12). Surprisingly, in contrast to $\boldsymbol{A}$, $\boldsymbol{A}_{SUR}$ does not depend on b_1 either.

These results obtained for $N = 3$ and $H = 2$ might to be valid for other values of these parameters as well. That is, it is possible that $\mathbf{A}$ does not depend on α, and that $\boldsymbol{A}_{SUR}$ neither depends on α nor on $\mathbf{b}$, where $\mathbf{b} = (b_1, b_2, \ldots b_{H-1})$. This possibility can be studied analytically by calculating the derivatives $\partial\mathbf{A}/\partial\alpha$, $\partial\mathbf{A}_{SUR}/\partial\alpha$ and $\partial\mathbf{A}_{SUR}/\partial b_i$ with $i = 1, 2, \ldots, H-1$ and checking whether they equal zero.

Using the Symbolic Math Toolbox of MATLAB, I find that all the abovementioned derivatives indeed equal zero for $N = 2, 3, 4, 5$ and $2 \leq H \leq N$. For larger horizons, analytical investigations turn out to be impossible due to computational reasons. However, it is possible to increase N if H is small. If $H = 2$ or $H = 3$, N can be increased to 10, and the derivatives continue to equal zero. Yet, it could of course be possible that this is due to the still rather small values of N and H which can be studied in this manner.

Therefore, I also conduct a simulation study. In 1000 simulations, N is drawn from a discrete uniform distribution over the interval $[2, 60]$, H is drawn from a discrete uniform distribution over the interval $[2, \min(N, 15)]$, α equals $\chi + 1$ where χ is drawn from the χ_2^2 distribution (so that $E[\alpha] = 3$), and each element of $\mathbf{b}$ is drawn from a standard normal distribution. In these simulations, I find that the difference between a matrix $\boldsymbol{A}$ generated in the way described and a second matrix $\bar{\boldsymbol{A}}$ with the same N, H, and $\mathbf{b}$, but with $\alpha = \bar{\alpha}$ where $\bar{\alpha}$ is an arbitrary value (I use $\alpha = 3$) always equals zero for every element of these matrices. Along the same lines, the differences between the elements of a matrix $\boldsymbol{A}_{SUR}$ generated in the way described and the elements of a matrix $\bar{\boldsymbol{A}}_{SUR}$ with the same N and H, but with $\alpha = \bar{\alpha}$ and $\mathbf{b} = \bar{\mathbf{b}}$, where $\bar{\alpha}$ is an arbitrary value and $\bar{\mathbf{b}}$ is an arbitrary $((H-1) \times 1)$ vector (I use $\alpha = 3$ and $\mathbf{b} = \mathbf{1}'_{H-1}$) are always zero. So also for large values of N and H, apparently $\boldsymbol{A}$ does not depend on α and $\boldsymbol{A}_{SUR}$ neither depends on α nor on $\mathbf{b}$.[8]

So the GLS projection matrix $\mathbf{X}'\boldsymbol{A}$ does not depend on the distribution of the shocks, because α is the only distribution parameter that could have appeared in the projection matrix. The SUR projection matrix $\mathbf{X}'\boldsymbol{A}_{SUR}$ does not depend on the DGP at all. This means that the GLS estimator depends on the distribution of the shocks ε_t only through $\mathbf{e}^2$, but not through $\boldsymbol{A}$. Along the same lines, the SUR estimator depends on the DGP only through $\mathbf{e}^2$, but not through $\boldsymbol{A}_{SUR}$.

For the case of $H = 2$ and arbitrary values of all other parameters, the parame-

[8] It should be noted that both results depend on the data structure investigated here, where the number of available forecast errors decreases one-to-one with the forecast horizon. Simulations show that, in other data structures, $\mathbf{A}$ and $\mathbf{A}_{SUR}$ typically depend on α and $\mathbf{b}$.

An exception is a balanced sample of forecast errors, where all forecast errors come from the same time span. In this case, it is well known that OLS and SUR estimation yield the same results if the regressors are constants. Thus, in this case $\mathbf{A}_{SUR}$ does not depend on α and $\mathbf{b}$ either, but in contrast to the data structure studied here, no efficiency gains can be achieved with the SUR estimator.

ter independence of $\boldsymbol{A}_{SUR}$ can actually be proven analytically. As demonstrated in Appendix A.2, in this case $\boldsymbol{A}_{SUR}$ equals

$$\boldsymbol{A}_{SUR} = \begin{bmatrix} \frac{1}{N} & \frac{1}{N}\mathbf{1}'_{N-1} & \mathbf{0}'_{N-1} \\ \frac{1}{N} & -\frac{1}{N(N-1)}\mathbf{1}'_{N-1} & \frac{1}{N-1}\mathbf{1}'_{N-1} \end{bmatrix}.$$

The consequences of these findings are very useful in practice, because the covariance matrices $\boldsymbol{\Omega}$ and $\boldsymbol{\Omega}_{SUR}$ are usually unknown, and the samples under study can be very small.[9] The parameter independence of $\boldsymbol{A}_{SUR}$ implies that, although $\boldsymbol{\Omega}_{SUR}$ is unknown, and although the sample might be very small, the estimation uncertainty for $\boldsymbol{\Omega}_{SUR}$ does not matter. Only the known parameters H and N matter for the construction of $\boldsymbol{\Omega}_{SUR}$. For α and $\mathbf{b}$, arbitrary values can be assumed. Thus, the SUR estimator does not require the estimation of a covariance matrix. For the GLS estimator, an arbitrary value of α can be assumed, but $\mathbf{b}$ has to be estimated.

5.2 The Recursiveness of the GLS and SUR Estimator

Suppose that, as in the example above $N = 3$ and $\alpha = 3$. However, assume that $H = 3$, i.e. that the 3-step-ahead forecast error $e_{3,0}$ is available, so that the vector $\mathbf{e}^2$ becomes

$$\mathbf{e}^2 = \left(\mathbf{e}_1^2, \mathbf{e}_2^2, \mathbf{e}_3^2\right)' = \left(\left(e_{1,0}^2, e_{2,1}^2, e_{3,2}^2\right), \left(e_{2,0}^2, e_{3,1}^2\right), \left(e_{3,0}^2\right)\right)'.$$

Letting $\boldsymbol{\Omega}_{N=3,H=2}$ denote the covariance matrix (7) of the previous example with $N = 3$, $H = 2$ (and $\alpha = 3$), the covariance matrix for $H = 3$ is given by

$$\boldsymbol{\Omega} = 2\sigma^4 \begin{bmatrix} & & & & & b_2^2 \\ & & & & & b_1^2 \\ & & \frac{1}{2\sigma^4}\boldsymbol{\Omega}_{N=3,H=2} & & & 1 \\ & & & & & (1+b_2)^2 b_1^2 \\ & & & & & (1+b_1^2)^2 \\ b_2^2 & b_1^2 & 1 & (1+b_2)^2 b_1^2 & (1+b_1^2)^2 & (1+b_1^2+b_2^2)^2 \end{bmatrix}.$$

[9] Moreover, the forecasts are not exactly optimal. Otherwise, $\mathbf{b}$ could be precisely determined based on the forecast errors.

Along the same lines, let $\mathbf{X}_{N=3,H=2}$ denote the regressor matrix of the previous example with $N = 3$, $H = 2$. Then the regressor matrix for $H = 3$ equals

$$\mathbf{X} = \begin{bmatrix} \mathbf{X}_{N=3,H=2} & \mathbf{0}_5 \\ \mathbf{0}'_2 & 1 \end{bmatrix}. \tag{14}$$

Using these matrices to find $\boldsymbol{A}$ yields

$$\boldsymbol{A} = \begin{bmatrix} \frac{1}{3} & \frac{1}{3} & \frac{1}{3} & 0 & 0 & 0 \\ \frac{1}{6}\left(2 - b_1^2\right) & -\frac{1}{6}\left(b_1^2 + 1\right) & \frac{1}{6}\left(2b_1^2 - 1\right) & \frac{1}{2} & \frac{1}{2} & 0 \\ a_1 & a_2 & a_3 & \frac{1}{2}\left(1 - b_2\right) & \frac{1}{2}\left(b_2 - 1\right) & 1 \end{bmatrix} \tag{15}$$

with a_1, a_2 and a_3 being polynomials in b_1 and b_2 which are not written out to save space.

Comparing (11) to (15) shows that the estimates for $h = 1$ and $h = 2$, $\hat{\sigma}^2_{GLS,1}$ and $\hat{\sigma}^2_{GLS,2}$ are identical. So the information contained in $e^2_{3,0}$ is only used for the estimate for $h = 3$, $\hat{\sigma}^2_{GLS,3}$, but not for the smaller horizons. Along the same lines, the squared 2-step-ahead forecast errors $e^2_{2,0}$ and $e^2_{3,1}$ are used for the estimates $\hat{\sigma}^2_{GLS,2}$ and $\hat{\sigma}^2_{GLS,3}$, but not for the smaller horizon $\hat{\sigma}^2_{GLS,1}$. Moreover, it is interesting to see that $\hat{\sigma}^2_{GLS,1} = \hat{\sigma}^2_{OLS,1}$ and that the estimates $\hat{\sigma}^2_{GLS,2}$ and $\hat{\sigma}^2_{GLS,3}$ equal the sum of the OLS estimates $\hat{\sigma}^2_{OLS,2}$ and $\hat{\sigma}^2_{OLS,3}$, respectively, and of a weighted sum of the squared forecast errors from smaller horizons whose expectation must equal zero.[10]

For the SUR estimator, the same can be observed. Letting $\mathbf{\Omega}_{SUR,N=3,H=2}$ denote the covariance matrix (9) of the previous example with $N = 3$, $H = 2$ (and

[10] This zero-expectation property follows from the fact that the GLS estimates and the OLS estimates are consistent.

$\alpha = 3$), the covariance matrix for $H = 3$ is given by

$$\boldsymbol{\Omega}_{SUR}{=}2\sigma^4 \begin{bmatrix} & & & & & 0 \\ & & & & & 0 \\ & \frac{1}{2\sigma^4}\boldsymbol{\Omega}_{SUR,N=3,H=2} & & & & 1 \\ & & & & & 0 \\ & & & & & (1+b_1^2)^2 \\ 0 & 0 & 1 & 0 & (1+b_1^2)^2 & (1+b_1^2+b_2^2)^2 \end{bmatrix}.$$

Using $\boldsymbol{\Omega}_{SUR}$ and the regressor matrix (14) to find $\boldsymbol{A}_{SUR}$ gives

$$\boldsymbol{A}_{SUR} = \begin{bmatrix} \frac{1}{3} & \frac{1}{3} & \frac{1}{3} & 0 & 0 & 0 \\ \frac{1}{3} & -\frac{1}{6} & -\frac{1}{6} & \frac{1}{2} & \frac{1}{2} & 0 \\ \frac{1}{3} & -\frac{1}{6} & -\frac{1}{6} & \frac{1}{2} & -\frac{1}{2} & 1 \end{bmatrix}.$$

So also with the SUR estimator, the 3-step-ahead forecast errors are only used for the estimate $\hat{\sigma}^2_{SUR,3}$, and the 2-step-ahead forecasts are only used for the estimates $\hat{\sigma}^2_{SUR,2}$ and $\hat{\sigma}^2_{SUR,3}$. Also $\hat{\sigma}^2_{SUR,1} = \hat{\sigma}^2_{OLS,1}$, and the estimates $\hat{\sigma}^2_{SUR,2}$ and $\hat{\sigma}^2_{SUR,3}$ equal the sum of the OLS estimates $\hat{\sigma}^2_{OLS,2}$ and $\hat{\sigma}^2_{OLS,3}$, respectively, and of a weighted sum of the squared forecast errors from smaller horizons whose expectation equals zero.

To further investigate the hypotheses that the matrices $\boldsymbol{A}$ and $\boldsymbol{A}_{SUR}$ have the recursiveness properties described above, that $\hat{\sigma}^2_{GLS,1} = \hat{\sigma}^2_{SUR,1} = \hat{\sigma}^2_{OLS,1}$, and that the estimates $\hat{\sigma}^2_{GLS,\tilde{h}}$ and $\hat{\sigma}^2_{SUR,\tilde{h}}$ are sums containing the OLS estimate $\hat{\sigma}^2_{OLS,\tilde{h}}$, I use the same simulation design as described above. That is, I use random draws of $N, H, \mathbf{b}$ and α to calculate matrices $\boldsymbol{A}$ and $\boldsymbol{A}_{SUR}$ to check whether the hypotheses are rejected.

I find that for the GLS estimator, the matrix $\boldsymbol{A}$ always has a structure which can indeed be written as

$$
\boldsymbol{A} = \begin{bmatrix}
\frac{1}{N}\mathbf{1}'_N & \mathbf{0}'_{N-1} & \mathbf{0}'_{N-2} & \cdots & \mathbf{0}'_{N-H+3} & \mathbf{0}'_{N-H+2} & \mathbf{0}'_{N-H+1} \\
\mathbf{a}_{1,1} & \frac{1}{N-1}\mathbf{1}'_{N-1} & \mathbf{0}'_{N-2} & \cdots & \mathbf{0}'_{N-H+3} & \mathbf{0}'_{N-H+2} & \mathbf{0}'_{N-H+1} \\
\mathbf{a}_{2,1} & \mathbf{a}_{2,2} & \frac{1}{N-2}\mathbf{1}'_{N-2} & \cdots & \mathbf{0}'_{N-H+3} & \mathbf{0}'_{N-H+2} & \mathbf{0}'_{N-H+1} \\
\mathbf{a}_{3,1} & \mathbf{a}_{3,2} & \mathbf{a}_{3,3} & \cdots & \mathbf{0}'_{N-H+3} & \mathbf{0}'_{N-H+2} & \mathbf{0}'_{N-H+1} \\
\vdots & \vdots & \vdots & \ddots & \vdots & \vdots & \vdots \\
\mathbf{a}_{H-4,1} & \mathbf{a}_{H-4,2} & \mathbf{a}_{H-4,3} & \cdots & \mathbf{0}'_{N-H+3} & \mathbf{0}'_{N-H+2} & \mathbf{0}'_{N-H+1} \\
\mathbf{a}_{H-3,1} & \mathbf{a}_{H-3,2} & \mathbf{a}_{H-3,3} & \cdots & \frac{1}{N-H+3}\mathbf{1}'_{N-H+3} & \mathbf{0}'_{N-H+2} & \mathbf{0}'_{N-H+1} \\
\mathbf{a}_{H-2,1} & \mathbf{a}_{H-2,2} & \mathbf{a}_{H-2,3} & \cdots & \mathbf{a}_{H-2,H-2} & \frac{1}{N-H+2}\mathbf{1}'_{N-H+2} & \mathbf{0}'_{N-H+1} \\
\mathbf{a}_{H-1,1} & \mathbf{a}_{H-1,2} & \mathbf{a}_{H-1,3} & \cdots & \mathbf{a}_{H-1,H-2} & \mathbf{a}_{H-1,H-1} & \frac{1}{N-H+1}\mathbf{1}'_{N-H+1}
\end{bmatrix} \tag{16}
$$

where $\mathbf{a}_{i,j}$ is a row vector whose elements are polynomials in the elements of $\mathbf{b}$. The same structure is found for $\boldsymbol{A}_{SUR}$, but, as discovered above, the row vector $\mathbf{a}_{i,j}$ does not depend on $\mathbf{b}$.

This structure of $\boldsymbol{A}$ and $\boldsymbol{A}_{SUR}$ indeed implies that the inclusion of forecast errors of horizons larger than a certain $\tilde{h}$ do not affect the estimates for the horizons $h \leq \tilde{h}$. Thus, the GLS and the SUR estimator have a recursive structure, where for the estimation of forecast uncertainty for the horizon $\tilde{h}$, only forecast errors for the horizons $h = 1, 2, \ldots, \tilde{h}$ are employed. This means, for example, that the results for $h = 1, 2, \ldots, \tilde{h}$ presented in Figures 2, 3 and 4 would not change if the forecast errors for the horizons $\tilde{h}, \tilde{h}+1, \ldots, 9$ were excluded from the estimation. Of course, from (16) it also follows that $\hat{\sigma}^2_{GLS,1} = \hat{\sigma}^2_{SUR,1} = \hat{\sigma}^2_{OLS,1}$ and that $\hat{\sigma}^2_{GLS,\tilde{h}}$ and $\hat{\sigma}^2_{SUR,\tilde{h}}$ are sums containing the OLS estimate $\hat{\sigma}^2_{OLS,\tilde{h}}$.

Stated formally, the structure of $\boldsymbol{A}$ implies that the GLS estimate of $\sigma^2_{\tilde{h}}$ can be written as

$$\hat{\sigma}^2_{GLS,\tilde{h}} = \mathbf{a}_{\tilde{h}-1,1}\mathbf{e}^2_1 + \mathbf{a}^2_{\tilde{h}-1,2}\mathbf{e}^2_2 + \ldots + \mathbf{a}_{\tilde{h}-1,\tilde{h}-1}\mathbf{e}^2_{\tilde{h}-1} + \hat{\sigma}^2_{OLS,\tilde{h}} \tag{17}$$

where the elements $\mathbf{a}_{\tilde{h}-1,j}$ have the interesting property

$$\mathbf{a}_{\tilde{h}-1,j}\mathbf{1}_{N-j+1} = 0, \tag{18}$$

so that $\hat{\sigma}^2_{GLS,\tilde{h}}$ is the sum of the OLS estimate $\hat{\sigma}^2_{OLS,\tilde{h}}$ and $\tilde{h}-1$ summation terms which *each* have an expectation of zero.[11] The same holds for $\hat{\sigma}^2_{SUR,\tilde{h}}$.

If the SUR estimator is employed, actually the matrix $\boldsymbol{A}_{SUR}$ might be constructed without a prior determination of $\boldsymbol{\Omega}_{SUR}$. While the structure displayed in (16) is common to $\boldsymbol{A}$ and $\boldsymbol{A}_{SUR}$, the matrix $\boldsymbol{A}_{SUR}$ can be precisely determined using the equations

$$\begin{aligned} \mathbf{a}_{i,j} &= \begin{bmatrix} \frac{1}{N-j+1} & \mathbf{c}_j \end{bmatrix} \\ \mathbf{c}_j &= -\frac{1}{(N-j+1)(N-j)}\mathbf{1}'_{N-j}. \end{aligned} \tag{19}$$

Note that, as claimed in (18),

$$\frac{1}{N-j+1} + \mathbf{c}_j \mathbf{1}_{N-j} = 0$$

So given N and H, the SUR estimates can be obtained simply by setting up $\boldsymbol{A}_{SUR}$ as in (16) with $\mathbf{a}_{i,j}$ determined by (19). The product $\boldsymbol{A}_{SUR}\mathbf{e}^2$ yields the SUR estimates.

6 Problems in Practice

6.1 Unknown Covariance Matrix

In general, the covariance matrix of the squared forecast errors is unknown in practice. This is irrelevant for the SUR estimator, because its covariance matrix depends on the known parameters N and H only. However, GLS estimation requires the estimation of the covariance matrix $\boldsymbol{\Omega}$. If the forecasts were optimal, $\mathbf{b}$ could be calculated directly from the forecast errors and α could be set arbitrarily. However, forecasts are typically not exactly optimal.

Thus, the estimator to be employed in this case is a feasible GLS (henceforth FGLS) estimator. The covariance matrix $\boldsymbol{\Omega}$ here is estimated based on the em-

[11] That is, not only $E\left[\mathbf{a}_{\tilde{h}-1,1}\mathbf{e}_1^2 + \mathbf{a}^2_{\tilde{h}-1,2}\mathbf{e}_2^2 + \ldots + \mathbf{a}_{\tilde{h}-1,\tilde{h}-1}\mathbf{e}^2_{\tilde{h}-1}\right] = 0$ holds, but also $E\left[\mathbf{a}_{\tilde{h}-1,j}\mathbf{e}_j^2\right] = 0$ for $j = 1, 2, \ldots, \tilde{h}-1$.
It might also be interesting to note that (17) corresponds to the result of Grenander and Rosenblatt (1957) in that the GLS estimate asymptotically equals the OLS estimate.

pirical covariances of the squared forecast errors as described in Appendix A.3.[12] This estimated matrix $\hat{\mathbf{\Omega}}$ is then used to obtain the FGLS estimates according to

$$\hat{\boldsymbol{\sigma}}^2_{FGLS} = \left(\mathbf{X}'\hat{\mathbf{\Omega}}^{-1}\mathbf{X}\right)^{-1}\mathbf{X}'\hat{\mathbf{\Omega}}^{-1}\mathbf{e}^2.$$

However, this procedure sometimes leads to huge outliers where the FGLS estimates differ enormously from the OLS estimates and from the true values. The FGLS estimates can even become negative. Therefore, a very simple method of shrinking is used, giving rise to the shrunk (feasible) GLS (henceforth SGLS) estimator

$$\hat{\boldsymbol{\sigma}}^2_{SGLS} = \mathbf{w} \odot \hat{\boldsymbol{\sigma}}^2_{OLS} + (\mathbf{1}_H - \mathbf{w}) \odot \hat{\boldsymbol{\sigma}}^2_{FGLS}$$

with

$$\mathbf{w} = \frac{\left(\hat{\boldsymbol{\sigma}}^2_{OLS} - \hat{\boldsymbol{\sigma}}^2_{FGLS}\right)^2}{\mathbf{1}_H - \left(\hat{\boldsymbol{\sigma}}^2_{OLS} - \hat{\boldsymbol{\sigma}}^2_{FGLS}\right)^2}.$$

The fraction bar denotes elementwise division, raising a vector to the second power here means raising each element of that vector to the second power, and $\odot$ denotes Schur multiplication.

If the difference between the OLS and the FGLS estimator is very large, $\mathbf{w}$ is close to $\mathbf{1}_H$. If the difference is very small, $\mathbf{w}$ is close to $\mathbf{0}_H$. Therefore, with this approach, the larger the difference between the OLS and the FGLS estimator is, the stronger the estimator $\hat{\boldsymbol{\sigma}}^2_{SGLS}$ is shrunk towards the OLS estimator.[13] Note that $\hat{\boldsymbol{\sigma}}^2_{SGLS}$ is an unbiased estimator, because it is a weighted average of the two unbiased estimators $\hat{\boldsymbol{\sigma}}^2_{OLS}$ and $\hat{\boldsymbol{\sigma}}^2_{FGLS}$.

In order to compare the efficiency gains of $\hat{\boldsymbol{\sigma}}^2_{FGLS}$, $\hat{\boldsymbol{\sigma}}^2_{SGLS}$ and $\hat{\boldsymbol{\sigma}}^2_{FGLS}$, I consider the case of optimal forecasts and an AR(1)-process with $\rho = 0.5$ and normally distributed error terms. The results in Table 4 show that, even in case of optimal forecasts, the FGLS estimator only works well in situations where efficiency gains are very small anyway, i.e. in relatively large samples. In the case of small samples, i.e. in case of a small N given a certain value of H, the FGLS estimator leads to

[12] If the forecasts were close to optimal, one could estimate $\mathbf{b}$ reasonably well from the forecast errors and construct the covariance matrix based on this estimate. However, in practice most forecasts are unlikely to be close to optimal, so that the estimation of $\mathbf{b}$ can be problematic.

[13] Another possibility would be to shrink the covariance matrix towards some target matrix. However, this approach will not be pursued here.

N	3	10	10	20
H	2	2	3	3
$\varphi_{GLS,2}$	5.3	1.7	1.8	0.9
$\varphi_{SGLS,2}$	0.0	1.8	1.7	1.1
$\varphi_{FGLS,2}$	−20.7	2.0	−2.6	1.0
$\varphi_{GLS,3}$			4.2	2.5
$\varphi_{SGLS,3}$			1.5	2.0
$\varphi_{FGLS,3}$			−15.1	−6.7

Table 4: Efficiency gains for optimal forecasts with $\alpha = 3$ and $b_1 = 0.5$. Efficiency gains are determined by 10000 simulations.

efficiency losses. For example, if $N = 3$ and $H = 2$, the FGLS estimator of the 3-step-ahead forecast uncertainty is much less efficient than the OLS estimator. This is due to the fact that the estimation uncertainty for the covariance matrix is quite large. Therefore, the FGLS estimator will not be considered in what follows. It will only be used to determine the SGLS estimator. The SGLS estimator is in general more efficient than the OLS estimator.

6.2 Non-Optimal Forecasts

Economic forecasts are often found to be non-optimal[14], which indicates that many forecasting models are probably more than just marginally misspecified. Actually, this is one of the reasons why most institutions measure forecast uncertainty using past forecast errors. If the forecasts were optimal, the forecast uncertainty could be calculated employing the forecasting model. Therefore, the results found above give only limited guidance to practitioners.

If the DGP is given by (1), and the forecasts are non-optimal, the forecast errors have the representation

$$e_{t+h,t} = a_h + \sum_{i=0}^{\infty} d_{h,i}\varepsilon_{t+h-i}$$

with $d_{h,0} = 1$ for all h. Hence, all forecast errors can be correlated, the forecast uncertainty can decrease with h etc. However, as in the case of optimal forecasts,

[14] see, e.g. Brown and Maital (1981) or Zarnowitz (1985).

the inequality

$$E\left[e_{t,t-h}e_{t,t-j}\right] \geq \sigma^2$$

with $j \geq 1$ continues to hold. That is, the forecast errors of forecasts for a certain period are always strongly correlated due to the shock in that period. This strong correlation does not depend on forecast optimality. Since the SUR estimator derived for optimal forecasts is based on the covariances of $e_{t,t-h}^2$ and $e_{t,t-j}^2$, it might be the case that this estimator produces reasonable results even in the case of non-optimal forecasts.

In addition to the SUR estimator, the SGLS estimator will be considered. For the estimation of the required covariance matrix of the squared forecast errors, I also use the restrictions derived for optimal forecasts.[15]

In the following, I investigate three important problems which can occur in the forecasting process and lead to non-optimal forecasts: Bias, dynamic misspecification, and structural breaks.

6.2.1 Bias

Suppose that the DGP is given by

$$y_t = \mu + \rho y_{t-1} + \varepsilon_t.$$

The forecasting model does not contain a constant but uses the correct autoregressive parameter, so that the forecasts are determined by

$$\hat{y}_{t+h,t} = \rho^h y_t.$$

Here and in the following, the error terms ε_t are identically, independently and normally distributed. If $\mu \neq 0$, the forecasting model is misspecified, the forecasts are biased, and the elements of the covariance matrix of the squared forecast errors cannot be determined by (4). Nevertheless, I will employ the covariance matrices

[15] In principle, the covariance matrix of the squared forecast errors could be determined without restrictions. However, the fact that the samples under study can be very small favours their use. Therefore, and in order to facilitate comparisons with the SUR estimator, I use the restrictions derived for optimal forecasts.

derived for optimal forecasts for the SUR and the SGLS estimator.

In addition to the SUR and SGLS estimator, here I will also consider the determination of forecast uncertainty based on the forecasting model. The model-based uncertainty is simply given by

$$\boldsymbol{\sigma}_{MB}^2 = \sigma^2 \left(\sum_{i=0}^{0} \rho^{2i}, \sum_{i=0}^{1} \rho^{2i}, \ldots, \sum_{i=0}^{H-1} \rho^{2i} \right)'.$$

I abstract from estimation uncertainty, so that the parameters σ and ρ are known. Thus, the efficiency gains for the model-based uncertainty are determined by

$$\boldsymbol{\varphi}_{MB} = 100 \ln \left(\sqrt{diag\left(Cov\left(\hat{\boldsymbol{\sigma}}_{OLS}^2\right)\right)} \div \sqrt{\left(\boldsymbol{\sigma}_{MB}^2 - \boldsymbol{\sigma}^2\right)^2} \right).$$

Since there is no estimation uncertainty with respect to $\boldsymbol{\sigma}_{MB}^2$, the only reason why $\boldsymbol{\sigma}_{MB}^2$ differs from $\boldsymbol{\sigma}^2$ is the bias of $\boldsymbol{\sigma}_{MB}^2$ due to neglecting μ in the forecasting model.

Simulation results are reported in Table 5 for $\rho = 0.5$, $N = 20$ and $H = 9$. The biasedness of the forecasts leads only to minor reductions of the efficiency gains of the SUR estimator. This holds even for extremely large biases like $\mu = 10$. The presence of bias can apparently not cause efficiency losses of the SUR estimator as long as the dynamic specification of the forecasting model is correct. The SGLS estimator also leads to efficiency gains for all values of μ considered, but the gains are markedly smaller than those of the SUR estimator.

The efficiency gains obtained by using model-based uncertainty strongly depend on the value of μ. For $\mu = 0.5$, the model-based calculation of uncertainty is more efficient than OLS estimation. For $\mu = 1$ and $\mu = 10$, i.e. in case of strong misspecification of the model, the OLS estimator is more efficient. It should be noted that $\boldsymbol{\sigma}_{MB}^2$ is always biased downwards, i.e. $\boldsymbol{\sigma}_{MB}^2$ always understates the true forecast uncertainty if $\mu \neq 0$.

	$\mu = 0$			$\mu = 0.5$			$\mu = 1$			$\mu = 100$		
	$SGLS$	SUR	MB	$SGLS$	SUR	MB	$SGLS$	SUR	MB	$SGLS$	SUR	MB
$h = 1$	0.0	0.0	∞	0.0	0.0	41.3	0.0	0.0	-60.6	0.0	0.0	-309.7
$h = 2$	1.5	1.3	∞	1.6	1.2	18.4	1.2	1.0	-70.0	0.4	1.0	-308.4
$h = 3$	3.2	3.2	∞	3.3	2.8	13.5	2.6	2.4	-71.1	0.4	2.4	-306.8
$h = 4$	3.7	5.2	∞	4.3	4.7	12.8	3.5	4.3	-70.2	0.3	4.1	-304.9
$h = 5$	3.8	7.9	∞	5.0	6.9	13.7	3.9	6.7	-68.3	0.2	6.4	-302.5
$h = 6$	4.2	10.6	∞	5.6	9.3	15.4	4.1	9.3	-65.9	0.2	8.9	-299.9
$h = 7$	4.5	13.4	∞	5.8	12.0	17.7	4.0	12.0	-63.3	0.2	11.8	-297.0
$h = 8$	5.1	17.0	∞	5.8	15.3	21.0	3.9	15.0	-60.3	0.1	15.4	-293.4
$h = 9$	5.1	20.9	∞	5.9	18.8	24.4	3.7	18.7	-56.6	0.1	18.9	-289.9

Table 5: Efficiency gains of the GLS estimator, the SUR estimator, and the model-based determination of forecast uncertainty with $N = 20$ and $H = 9$ if the true DGP contains a constant μ while the forecasting model does not. $\rho = 0.5$ in the DGP and the forecasting model. Results are based on 10000 simulations.

6.2.2 AR(1)-Processes

Suppose that the DGP is given by

$$y_t = \rho y_{t-1} + \varepsilon_t,$$

and the forecasts are made according to

$$\hat{y}_{t+h,t} = \hat{\rho}^h y_t.$$

If $\hat{\rho} \neq \rho$, the forecasting model's dynamics are misspecified. Simulations with under- and overestimations of ρ are reported in Table 6 for $N = 20$ and $H = 9$.

The results in Table 6 indicate that efficiency losses are unlikely if ρ is underestimated. The efficiency gains of the SUR estimator for $\rho = 1.0$ are larger than those of the SGLS estimator, but still very small. This, however, is unrelated to the dynamic misspecification as e.g. Table 2 shows. Interestingly, the efficiency gains of the SUR estimator increase when the misspecification becomes stronger.

If ρ is overestimated, efficiency losses can occur with the SGLS and with the SUR estimator. With the SUR estimator, losses are restricted to the case of strong overestimation of ρ. If ρ equals 0.9 and the forecasting model uses $\hat{\rho} = 1.0$, the SUR estimator yields small efficiency gains for all horizons (except $h = 1$, of course), whereas with the SGLS estimator, a slight efficiency loss occurs for $h = 2$.

	$\rho = 1.0$		$\rho = 1.0$		$\rho = 0.6$		$\rho = 0.9$	
	$\hat{\rho} = 0.6$		$\hat{\rho} = 0.9$		$\hat{\rho} = 1.0$		$\hat{\rho} = 1.0$	
	SGLS	*SUR*	*SGLS*	*SUR*	*SGLS*	*SUR*	*SGLS*	*SUR*
$h = 1$	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
$h = 2$	0.0	0.2	0.1	0.2	−0.6	0.4	−0.4	0.3
$h = 3$	0.0	0.5	0.2	0.4	−1.0	0.6	0.0	0.8
$h = 4$	0.0	0.9	0.1	0.6	−1.2	0.0	0.2	1.2
$h = 5$	0.0	1.3	0.0	0.8	−1.5	−0.9	0.0	1.6
$h = 6$	0.0	1.8	0.0	0.9	−1.4	−2.7	0.0	1.7
$h = 7$	0.0	2.3	0.0	1.1	−1.4	−5.0	0.0	1.5
$h = 8$	0.0	2.8	0.0	1.4	−1.2	−8.1	0.0	1.2
$h = 9$	0.0	3.3	0.0	1.7	−1.3	−12.3	0.0	0.6

Table 6: Efficiency gains of the SGLS and the SUR estimators with $N = 20$ and $H = 9$ if the autoregressive coefficient of the forecasting model is misspecified. Results are based on 10000 simulations.

6.2.3 AR(2)-Processes

Assume that the DGP is given by

$$y_t = \rho_1 y_{t-1} + \rho_2 y_{t-1} + \varepsilon_t,$$

with the eigenvalues of this process denoted by λ_1 and λ_2. The forecasting model is

$$y_t = \hat{\rho} y_{t-1} + u_t$$

with $E[u_t] = 0$. $\hat{\rho}$ is the asymptotic result of a regression of y_t on y_{t-1}. Thus, the misspecification arises from using a too small lag order. The least squares estimator equals the Yule-Walker estimator asymptotically, so that

$$\lim_{T \to \infty} \hat{\rho} = \frac{\rho_1}{1 - \rho_2}$$

holds, where T denotes the sample size.

Several values of ρ_1 and ρ_2 are considered. The results of the simulations are displayed in Table 7. Obviously, the SGLS and SUR estimator are more efficient than the OLS estimator in many cases. Only if the true DGP contains an eigenvalue which is close to unity and closer to unity than the eigenvalue of the

ρ_1	0.750		0.100		1.200		0.080		1.425		0.095		1.0945	
ρ_2	−0.500		0.800		−0.500		0.900		−0.500		0.900		−0.100	
$\lvert\lambda_1\rvert$	0.707		0.946		0.707		0.990		0.800		0.997		0.994	
$\lvert\lambda_2\rvert$	0.707		0.846		0.707		0.910		0.625		0.902		0.101	
$\hat{\rho}$	0.500		0.500		0.800		0.800		0.950		0.950		0.995	
	SGLS	SUR	SGLS	SUR	SGLS	SUR	SGLS	SUR	SGLS	SUR	SGLS	SUR	SGLS	SUR
$h = 1$	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
$h = 2$	1.6	1.2	0.9	0.8	1.1	0.9	0.5	−0.1	0.7	0.5	−7.4	−16.6	0.2	0.4
$h = 3$	3.9	3.5	1.7	1.8	1.9	2.1	1.0	1.2	0.8	1.2	1.5	1.6	0.9	1.1
$h = 4$	6.2	6.5	1.9	3.0	1.8	3.3	0.3	0.7	0.4	2.0	−0.1	−5.5	0.7	1.8
$h = 5$	7.7	10.3	2.0	4.5	1.6	4.9	0.4	1.3	0.2	2.7	1.1	2.3	0.3	2.4
$h = 6$	8.0	13.6	2.0	6.3	1.4	6.7	0.2	1.3	0.1	3.3	0.0	−1.8	0.2	3.0
$h = 7$	8.0	16.5	1.9	7.9	1.6	9.1	0.2	1.7	0.1	3.9	0.5	1.9	0.2	3.5
$h = 8$	8.1	19.6	1.8	9.8	1.5	12.1	0.1	1.9	0.1	4.4	0.2	−0.2	0.1	3.8
$h = 9$	7.5	23.3	1.6	11.7	1.5	16.0	0.2	2.4	0.0	4.8	0.1	1.5	0.0	4.0

Table 7: Efficiency gains of the GLS and the SUR estimators with $N = 20$ and $H = 9$ if the true DGP is an AR(2)-process while the forecasting model is an AR(1)-model. Results are based on 10000 simulations.

forecasting model, for some forecast horizons OLS is found to be preferable.

The most extreme case is observed for $\rho_1 = 0.095$ and $\rho_2 = 0.9$ resulting in an eigenvalue with an absolute value of 0.997 and $\hat{\rho} = 0.95$. In this case, for $h = 2$, the efficiency losses of SGLS and SUR estimation reach 7% and 17%, respectively. However, if at least one of the two conditions $\max(|\lambda_1|, |\lambda_2|) \approx 1$ and $\max(|\lambda_1|, |\lambda_2|) > \hat{\rho}$ is not fulfilled, SGLS and SUR appear to be more efficient than OLS. So the situations where OLS is preferable seem to be restricted to a rather limited part of the parameter space. The efficiency gains achieved with the SUR estimator are in general substantially larger than those with the SGLS estimator.

6.2.4 Structural Breaks

Another problem that can occur in practice is a change of the DGP due to a structural break, which usually leads to a change in the properties of the squared forecast errors. Suppose that, as in the examples above, we have $N = 3$ and $H = 2$, so that the vector of squared forecast errors is

$$\mathbf{e}^2 = \left(\mathbf{e}_1^2, \mathbf{e}_2^2\right)' = \left(\left(e_{1,0}^2, e_{2,1}^2, e_{3,2}^2\right), \left(e_{2,0}^2, e_{3,1}^2\right)\right)'.$$

Now consider the case where a structural break of the DGP occurs in period T_1+1. Since the first forecast is made in T_1-1, this implies that $e_{1,0}^2$ is not affected by the structural break, but all other elements of $\mathbf{e}^2$ are. When considering recent forecast errors, the forecaster is usually interested in the current forecast uncertainty, so this would be the uncertainty after the structural break.

The OLS and hence all other estimators yield biased results for $h = 1$, because they use a squared forecast error that occurred before the structural break ($e^2_{1,0}$) in addition to the squared forecast errors that occurred after the structural break ($e^2_{2,1}$,$e^2_{3,2}$). For $h = 2$, the OLS estimator only uses squared forecast errors that occurred after the structural break ($e^2_{2,0}$,$e^2_{3,1}$), whereas the SGLS and the SUR estimator use all elements of $\mathbf{e}^2$, including $e^2_{1,0}$. Hence the OLS estimate for $h = 2$ is unbiased, whereas the SGLS and the SUR estimate are biased. Therefore, in the case of a structural break in the DGP, the OLS estimator in general should be preferred. Yet, if the structural break only leads to small changes of forecast uncertainty, it could be that the gains from the smaller variance of the SGLS and the SUR estimator outweigh the loss due to the bias.

7 Applications to the Bank of England's Inflation Forecasts

7.1 The Uncertainty About the 2- to 3-Year-Ahead Forecasts

In the BoE Inflation Reports as of August 2004, inflation forecasts, conditioned on the interest rate path expected by market participants (henceforth forecasts based on market rates), are published for up to 13 quarters ahead. Since the smallest forecast horizon actually corresponds to a nowcast for the current quarter, the largest forecast horizon corresponds to a 3-year-ahead forecast. Before August 2004, the largest forecast horizon was 9, corresponding to 2-year-ahead forecasts. So if the forecast uncertainty concerning the forecast horizons 10 to 13 is to be estimated, only very few forecast errors are available. With the last observation coming from 2008q4, the number of forecast errors ranges from 6 (horizon 13) to 9 (horizon 10). For the nowcasts as of August 2004, i.e. the current quarter forecasts, we have 18 forecast errors, so that $N = 18$.

If only the forecast horizons 1 to 9 are to be investigated, the sample of forecasts based on market rates starts in 1998q1. In this case $N = 44$. In 2004q1, the BoE switched from targeting and forecasting the inflation of the all items retail prices index excluding mortgage interest payments (henceforth RPIX inflation) to the inflation of the consumer price index (henceforth CPI inflation). However, this change does not seem to have caused a structural break in forecast uncertainty.[16] Thus, the forecast uncertainty for horizons 1 to 9 can be estimated based on a relatively large sample.

I use the OLS and the SUR estimator to estimate the forecast uncertainty of the BoE. The SGLS estimator is not employed because, as found above, its efficiency gains generally are only small. In Table 8, estimation results for both samples, the shorter one with $N = 18$ and the larger one with $N = 44$ are presented. Instead of the estimated expected values of the squared forecast errors, their square roots are reported.[17] It is found that in the larger sample, the differences between the SUR and the OLS estimator are small. The SUR estimator yields slightly smaller estimates at longer horizons. In the shorter sample, however, the differences between both estimators are large at least for horizons greater than 7. For the 3-year-ahead forecast, the OLS estimator gives a result that is almost 50% larger than the one of the SUR estimator. For $10 \leq h \leq 13$, the differences always exceed 20%, and the roots of the estimated expected squared forecast errors range from 1.30 to 1.37 with the OLS estimator but only from 0.85 to 1.05 with the SUR estimator.

Looking at the larger sample, it seems that forecast uncertainty hardly increases with the forecast horizon if the forecast horizon exceeds $h = 5$. The roots of the estimated expected squared errors equal about 0.75 for $6 \leq h \leq 9$. Moreover, all the forecast values not reported here are usually quite close to the inflation target for larger horizons. Thus, it appears unlikely that the forecast uncertainty

[16] In each projection, the BoE reports a parameter for the uncertainty of each forecast horizon. These parameters can change from forecast to forecast, depending on the assessment of current forecast uncertainty by the BoE. From 2003q4 to 2004q1, no major change in reported forecast uncertainty took place. Looking at the squared forecast errors, there is no indication of a change in uncertainty as of 2004q1 either.

[17] The BoE reports a parameter for uncertainty that is related to the standard deviation of the forecast errors, so possible comparisons are easier when the square roots of the estimated expected values of the squared forecast errors are presented here. For the exact relation between the uncertainty parameter reported by the BoE and the standard deviation of the forecast errors, see Wallis (2004, p. 66).

	sample 2004q3 to 2008q4			sample 1998q1 to 2008q4		
	$\hat{\sigma}_{OLS,h}$	$\hat{\sigma}_{SUR,h}$	$100 \ln \left(\frac{\hat{\sigma}_{OLS,h}}{\hat{\sigma}_{SUR,h}} \right)$	$\hat{\sigma}_{OLS,h}$	$\hat{\sigma}_{SUR,h}$	$100 \ln \left(\frac{\hat{\sigma}_{OLS,h}}{\hat{\sigma}_{SUR,h}} \right)$
$h = 1$	**0.19**	**0.19**	**0.0**	**0.17**	**0.17**	**0.0**
$h = 2$	0.47	0.47	0.3	0.35	0.35	0.2
$h = 3$	0.66	0.65	1.5	0.46	0.46	0.6
$h = 4$	0.91	0.89	2.0	0.62	0.61	1.0
$h = 5$	**1.10**	**1.07**	**2.7**	**0.72**	**0.71**	**1.7**
$h = 6$	1.19	1.14	3.7	0.76	0.74	2.7
$h = 7$	1.19	1.10	7.2	0.76	0.73	3.9
$h = 8$	1.18	1.05	12.5	0.75	0.72	4.6
$h = 9$	**1.25**	**1.06**	**16.5**	**0.77**	**0.74**	**4.2**
$h = 10$	1.30	1.05	21.6			
$h = 11$	1.35	1.03	26.7			
$h = 12$	1.34	0.94	35.3			
$h = 13$	**1.37**	**0.85**	**47.7**			

Table 8: Square roots of OLS and SUR estimates of the expected values of the squared forecast errors of the BoE's inflation forecasts in different samples. The results for the nowcast as well as the 1-, 2- and 3-year-ahead forecasts are shown in bold.

for the horizons $10 \leq h \leq 13$ is much higher than the uncertainty for the horizons $6 \leq h \leq 9$. Therefore, the values obtained with the SUR estimator for $10 \leq h \leq 13$ in the smaller sample appear far more plausible than the values obtained with the OLS estimator.

Of course, it would be interesting to investigate the significance of the differences between both estimators. However, this would require a reliable measure of uncertainty about uncertainty for both estimators, which is not available yet.

7.2 The Width of the RPIX Inflation Fan Charts

Many empirical studies of the BoE's forecast errors as Clements (2004), Dowd (2007) and Wallis (2003, 2004) focus on the inflation forecasts starting in 1997q3, because the BoE then started to publish density forecasts. These studies conclude that the dispersion of these densities is too large except for short horizons. Wallis (2004) and Clements (2004) reach their conclusions by looking at the 1-year-ahead forecasts, so the effects of correlations of forecast errors across horizons are not taken into account.

The BoE decides on the dispersion of the fan charts based on past forecast errors. Thus, it could be possible that, due to a small sample problem, the true dispersion was overestimated by the BoE. Moreover, it could be that the studies cited rely on time spans where the true dispersion was underestimated by the methods used.

Here I try to investigate both hypotheses. I start by looking at the BoE's inflation forecast errors prior to the publication of density forecasts. The BoE became an inflation targeter at the end of 1992 and published its first Inflation Report in 1993q1. It seems probable that the forecasting regime changed with the announcement of an inflation target, so that it can be supposed that forecasts made before 1993 have produced errors different from those of forecasts made after that date, except maybe those for short horizons. However, according to Britton, Fisher, and Whitley (1998), the BoE used the forecast errors from the last ten years when it constructed the first fan charts. Yet, since the forecasts prior to 1993 are not publicly available[18], and for the reason mentioned above, I use a shorter sample of forecast errors starting in 1993. With the first forecast coming from 1993q1 and with 1997q2 as the last available realization before the publication of inflation forecasts, N equals 18. Since there are missing values for $h = 8$ and $h = 9$, H is set to 7.

The second hypothesis is investigated by studying the forecast uncertainty based on data from the same time span as the one studied by Wallis (2004), so that the last available realization comes from 2003q4. This implies that N equals 26. The data sets used by Clements (2004) and Dowd (2007) are similar, ending in 2003q1 and 2004q1, respectively.

Results obtained with the OLS and the SUR estimator are displayed in Table 9. Instead of the estimated expected values of the squared forecast errors, again their square roots are reported.

It turns out that the estimated forecast uncertainty in the first sample is much larger than in the second sample for all horizons. Moreover, the SUR and the OLS estimator yield similar results. They differ most strongly for $h = 5$ and $h = 6$, with the differences attaining 4% to 5%. For these horizons, the SUR estimator

[18] They are not available from the website of the BoE, in contrast to the forecasts since 1993.

	$h=1$	$h=2$	$h=3$	$h=4$	$h=5$	$h=6$	$h=7$	$h=8$	$h=9$
sample 1993q1 to 1997q2									
$\hat{\sigma}_{OLS,h}$	**0.25**	0.48	0.59	0.68	**0.73**	0.68	0.67		
$\hat{\sigma}_{SUR,h}$	**0.25**	0.48	0.59	0.66	**0.71**	0.65	0.67		
$100 \ln \left(\frac{\hat{\sigma}_{OLS,h}}{\hat{\sigma}_{SUR,h}} \right)$	**0.0**	0.7	−0.7	1.9	**3.7**	4.8	−0.3		
sample 1997q3 to 2003q4									
$\hat{\sigma}_{OLS,h}$	**0.16**	0.25	0.29	0.38	**0.39**	0.39	0.43	0.49	**0.52**
$\hat{\sigma}_{SUR,h}$	**0.16**	0.24	0.29	0.38	**0.42**	0.41	0.44	0.49	**0.52**
$100 \ln \left(\frac{\hat{\sigma}_{OLS,h}}{\hat{\sigma}_{SUR,h}} \right)$	**0.0**	0.5	−2.1	−1.1	**−5.3**	−4.4	−1.9	0.4	**0.2**
differences between samples, calculated as $100 \ln \left(\frac{\text{estimate of second sample}}{\text{estimate of first sample}} \right)$ based on									
$\hat{\sigma}_{OLS,h}$	**−43**	−67	−71	−58	**−62**	−55	−44		
$\hat{\sigma}_{SUR,h}$	**−43**	−67	−70	−55	**−53**	−46	−42		

Table 9: Square roots of OLS and SUR estimates of the expected values of the squared forecast errors of the BoE's inflation forecasts in different samples. The results for the nowcast as well as the 1- and 2-year-ahead forecasts are shown in bold.

yields lower values in the first sample and larger values in the second.

Based on the OLS estimator, one concludes that the standard deviation of the forecast errors in the second sample for $h = 5$ and $h = 6$ is 62% and 55% smaller, respectively, than in the first sample. Based on the SUR estimator, both numbers would decrease by 9 percentage points. In Clements (2004), and Wallis (2004) it is the 1-year-ahead forecasts, corresponding to $h = 5$, for which the fan charts are studied and considered as too wide. Looking at the numbers in Table 9, it appears very unlikely that this finding would have changed if the BoE and the researchers had taken the correlation of forecast errors among horizons into account. However, the significance levels at which the null hypothesis of correct width of the fan charts were rejected by the researchers would most certainly have increased.

8 Conclusion

In this paper, the joint estimation of forecast uncertainty for multi-step-ahead forecasts in samples of recent forecast errors is investigated. In order to make such an estimation possible, the formula for the covariances of squared forecast errors from optimal forecasts is derived. Using these covariances, GLS and SUR estimators of forecast uncertainty based on recent forecast errors are constructed.

The efficiency gains due to GLS and SUR estimation vanish asymptotically. In small samples, however, the efficiency gains over the OLS estimator can be large. They strongly depend on the persistence of the DGP and the number of forecast horizons H. If persistence is not too large, the GLS estimator and the SUR estimator yield similar results.

Several interesting properties of the GLS and the SUR estimator are observed. Although the covariance matrix of the squared forecast errors depends on the distribution of the shocks to the DGP, the GLS projection matrix does not depend on them. Moreover, the SUR projection matrix does not depend on the DGP at all.

This important result implies that, if the covariances of the forecast errors are unknown, the covariance matrix of the SUR estimator does not have to be estimated. One simply needs to impose a certain structure. For both estimators, the estimation of uncertainty for forecast horizon h does not use forecast errors of horizons larger than h. So only information from the 1- to h-step-ahead forecast errors are employed for the estimation. This implies that efficiency gains do not depend on the inclusion of errors from larger forecast horizons.

If the forecasts are non-optimal, the SUR estimator derived for optimal forecast errors is mostly found to be more efficient than the OLS estimator in several Monte Carlo studies. In general, it is also more efficient than the SGLS estimator defined above. Bias does not seem to affect the superiority of the SUR estimator. In the case of severe dynamic misspecification, however, the OLS estimator can sometimes be more efficient. If a structural break affecting forecast uncertainty occurs within the sample, the OLS estimator is in general more efficient than the SUR estimator unless the break is small.

An application to the BoE's inflation forecasts shows that the uncertainty for 2- to 3-year-ahead forecasts is likely to be overestimated by OLS. The SUR estimator gives more plausible results. Another application finds that the BoE might have slightly underestimated the true forecast uncertainty for medium horizons, including the 1-year-ahead forecast, prior to the publication of fan charts, possibly leading to fan charts for these horizons which were marginally too wide. In the sample where these fan charts were studied by several researchers as Clements (2004) and Wallis (2004), an individual estimation of uncertainty for each forecast horizon apparently leads to a small overestimation of forecast uncertainty for medium horizons. However, independently of the estimation method used, the differences in estimated forecast uncertainties between the sample prior to the publication of the fan charts and the sample used by researchers to evaluate the fan charts are large. Therefore, it is unlikely that the conclusion of the researchers that the fan charts are too wide for the 1-year-ahead forecast would have changed. Yet, the significance levels at which this conclusion was reached might be higher than claimed.

References

Britton, E., P. Fisher, and J. Whitley (1998): "The Inflation Report Projections: Understanding the Fan Chart," *Bank of England Quarterly Bulletin*, pp. 30–37.

Brown, B. W., and S. Maital (1981): "What Do Economists Know? An Empirical Study of Experts' Expectations," *Econometrica*, 49, 491–504.

Clements, M. P. (2004): "Evaluating the Bank of England Density Forecasts of Inflation," *The Economic Journal*, 114, 844–866.

Clements, M. P., and D. F. Hendry (1998): *Forecasting Economic Time Series*. Cambridge University Press, Cambridge.

Diebold, F. X. (1998): *Elements of Forecasting*. South-Western College Publishing, Cincinnati.

Dowd, K. (2007): "Too Good to Be True? The (In)Credibility of the UK Inflation Fan Charts," *Journal of Macroeconomics*, 29, 91–102.

Ericsson, N. R. (2002): "Predictable Uncertainty in Economic Forecasting," in *A Companion to Economic Forecasting*, ed. by D. F. H. Michael P. Clements, chap. 2, pp. 19–44. Blackwell.

Grenander, U., and M. Rosenblatt (1957): *Statistical Analysis of Stationary Time Series*. John Wiley and Sons, New York.

Im, E. I. (1994): "Unequal Numbers of Observations and Partial Efficiency Gain," *Economics Letters*, 46, 291–294.

Patton, A. J., and A. Timmermann (2007): "Properties of Optimal Forecasts under Asymmetric Loss and Nonlinearity," *Journal of Econometrics*, 140, 884–918.

Schmidt, P. (1977): "Estimation of Seemingly Unrelated Regressions with Unequal Numbers of Observations," *Journal of Econometrics*, 5(3), 365–377.

WALLIS, K. F. (1989): "Macroeconomic Forecasting: A Survey," *The Economic Journal*, 99, 28–61.

——— (2003): "Chi-Squared Tests of Interval and Density Forecasts, and the Bank of England's Fan Charts," *International Journal of Forecasting*, 19, 165–175.

——— (2004): "An Assessment of Bank of England and National Institute Inflation Forecast Uncertainties," *National Institute Economic Review*, 189, 64–71.

WILLIAMS, W. H., AND M. L. GOODMAN (1971): "A Simple Method for the Construction of Empirical Confidence Limits for Economic Forecasts," *Journal of the American Statistical Association*, 66(336), 752– 754.

ZARNOWITZ, V. (1985): "Rational Expectations and Macroeconomic Forecasts," *Journal of Business and Economic Statistics*, 3, 293–311.

A Appendix

A.1 Covariance of Squared Forecast Errors

In order to see how the covariance between the squared forecast errors $e^2_{t_2,t_1}$ and $e^2_{t_4,t_3}$ is determined, consider first an easy example with $t_3 > t_1, t_4 > t_2$. Of course, we also have $t_2 > t_1, t_4 > t_3$. Define m, n and q as

$$\begin{aligned} m &= t_3 - t_1 \\ n &= t_2 - t_1 \\ q &= t_4 - t_3. \end{aligned}$$

Now, one has to identify those shocks of $e^2_{t_2,t_1}$ and $e^2_{t_4,t_3}$ which overlap.

$$\begin{aligned} \omega &= E\left[\left(e^2_{t_2,t_1} - E\left(e^2_{t_2,t_1}\right)\right)\left(e^2_{t_4,t_3} - E\left(e^2_{t_4,t_3}\right)\right)\right] \\ &= E\left[\left(\left(\sum_{i=0}^{n-1} b_i \varepsilon_{t_2-i}\right)^2 - \sum_{i=0}^{n-1} b_i^2 \sigma^2\right)\left(\left(\sum_{i=0}^{q-1} b_i \varepsilon_{t_4-i}\right)^2 - \sum_{i=0}^{q-1} b_i^2 \sigma^2\right)\right] \end{aligned}$$

First, identify the shocks contained in $e^2_{t_2,t_1}$ not overlapping with those of $e^2_{t_4,t_3}$ and set the corresponding expectation term to 0.

$$\begin{aligned} \omega = & \; E\left[\left(\left(\sum_{i=0}^{n-m-1} b_i \varepsilon_{t_2-i}\right)^2 - \sum_{i=0}^{n-m-1} b_i^2 \sigma^2\right)\left(\left(\sum_{i=0}^{q-1} b_i \varepsilon_{t_4-i}\right)^2 - \sum_{i=0}^{q-1} b_i^2 \sigma^2\right)\right] \\ & \underbrace{+E\left[\left(\left(\sum_{i=n-m}^{n-1} b_i \varepsilon_{t_2-i}\right)^2 - \sum_{i=n-m}^{n-1} b_i^2 \sigma^2\right)\left(\left(\sum_{i=0}^{q-1} b_i \varepsilon_{t_4-i}\right)^2 - \sum_{i=0}^{q-1} b_i^2 \sigma^2\right)\right]}_{=0} \end{aligned}$$

Next, identify the shocks of $e^2_{t_4,t_3}$ not overlapping with those remaining of $e^2_{t_2,t_1}$ and set the corresponding expectation term to 0.

$$\begin{aligned}
\omega &= \underbrace{E\left[\left(\left(\sum_{i=0}^{n-m-1} b_i\varepsilon_{t_2-i}\right)^2 - \sum_{i=0}^{n-m-1} b_i^2\sigma^2\right)\left(\left(\sum_{i=0}^{q+m-n-1} b_i\varepsilon_{t_4-i}\right)^2 - \sum_{i=0}^{q+m-n-1} b_i^2\sigma^2\right)\right]}_{=0} \\
&\quad + E\left[\left(\left(\sum_{i=0}^{n-m-1} b_i\varepsilon_{t_2-i}\right)^2 - \sum_{i=0}^{n-m-1} b_i^2\sigma^2\right)\left(\left(\sum_{i=q+m-n}^{q-1} b_i\varepsilon_{t_4-i}\right)^2 - \sum_{i=q+m-n}^{q-1} b_i^2\sigma^2\right)\right] \\
&= E\left[\left(\left(\sum_{i=0}^{n-m-1} b_i\varepsilon_{t_2-i}\right)^2 - \sum_{i=0}^{n-m-1} b_i^2\sigma^2\right)\left(\left(\sum_{i=q+m-n}^{q-1} b_i\varepsilon_{t_4-i}\right)^2 - \sum_{i=q+m-n}^{q-1} b_i^2\sigma^2\right)\right]
\end{aligned}$$

Having obtained this expression for ω, the indices should now be rewritten such that it becomes clear which error terms are identical. To this end, note that

$$\begin{aligned}
t_4 - t_2 &= q + m - n \\
&\Rightarrow t_4 - (q+m-n) - i = t_2 - i
\end{aligned}$$

and in order to simplify notation, set

$$t := t_2.$$

This gives

$$\begin{aligned}
\omega = E&\left[\left(\left(\sum_{i=0}^{n-m-1} b_i\varepsilon_{t_2-i}\right)^2 - \sum_{i=0}^{n-m-1} b_i^2\sigma^2\right)\left(\left(\sum_{i=q+m-n}^{q-1} b_i\varepsilon_{t_4-i}\right)^2 - \sum_{i=q+m-n}^{q-1} b_i^2\sigma^2\right)\right] \\
= E&\left[\begin{array}{c}\left(\left(\sum_{i=0}^{n-m-1} b_i\varepsilon_{t_2-i}\right)^2 - \sum_{i=0}^{n-m-1} b_i^2\sigma^2\right) \\ \times\left(\left(\sum_{i=0}^{n-m-1} b_{q+m-n+i}\varepsilon_{t_4-(q+m-n)-i}\right)^2 - \sum_{i=0}^{n-m-1} b_{q+m-n+i}^2\sigma^2\right)\end{array}\right] \\
= E&\left[\left(\left(\sum_{i=0}^{n-m-1} b_i\varepsilon_{t-i}\right)^2 - \sum_{i=0}^{n-m-1} b_i^2\sigma^2\right)\left(\left(\sum_{i=0}^{n-m-1} b_{q+m-n+i}\varepsilon_{t-i}\right)^2 - \sum_{i=0}^{n-m-1} b_{q+m-n+i}^2\sigma^2\right)\right].
\end{aligned}$$

Now we drop the restrictions $t_3 > t_1$ and $t_4 > t_2$. Define

$$\begin{aligned} r &= \max(1, m+1) \\ s &= \min(n, m+q). \end{aligned}$$

Then it turns out that the covariance between the squared forecast errors $e^2_{t_2,t_1}$ and $e^2_{t_4,t_3}$ is given by

$$\begin{aligned} & E\left[\left(e^2_{t_2,t_1} - E\left(e^2_{t_2,t_1}\right)\right)\left(e^2_{t_4,t_3} - E\left(e^2_{t_4,t_3}\right)\right)\right] \\ = & \begin{cases} 0 & \text{if } s < r \\ E\left[\left(\left(\sum_{i=0}^{s-r} b_i \varepsilon_{t-i}\right)^2 - \sum_{i=0}^{s-r} b_i^2 \sigma^2\right)\left(\left(\sum_{i=0}^{s-r} b_{k-s+i} \varepsilon_{t-i}\right)^2 - \sum_{i=0}^{s-r} b_{k-s+i}^2 \sigma^2\right)\right] & \text{if } s \geq r \end{cases} \\ \text{with } k = & \max(p, m+q). \end{aligned}$$

so that for the example with restrictions given above, $s - r = n - m - 1$ and $k - s = q + m - n$.

In order to calculate the covariance, it is useful to write out the summations as

$$\begin{aligned} & E\left[\left(\left(\sum_{i=0}^{s-r} b_i \varepsilon_{t-i}\right)^2 - \sum_{i=0}^{s-r} b_i^2 \sigma^2\right)\left(\left(\sum_{i=0}^{s-r} b_{k-s+i} \varepsilon_{t-i}\right)^2 - \sum_{i=0}^{s-r} b_{k-s+i}^2 \sigma^2\right)\right] \\ = & E[\theta_1 \theta_2] \end{aligned}$$

with

$$\theta_1 = \begin{pmatrix} \left(b_0^2 \varepsilon_t^2 - b_0^2 \sigma^2\right) & +b_0 \varepsilon_t b_1 \varepsilon_{t-1} & \cdots & +b_0 \varepsilon_t b_{s-r} \varepsilon_{t-(s-r)} \\ +b_1 \varepsilon_{t-1} b_0 \varepsilon_t & +\left(b_1^2 \varepsilon_{t-1}^2 - b_1^2 \sigma^2\right) & \cdots & +b_1 \varepsilon_{t-1} b_{s-r} \varepsilon_{t-(s-r)} \\ \vdots & \vdots & \ddots & \vdots \\ +b_{s-r} \varepsilon_{t-(s-r)} b_0 \varepsilon_t & +b_{s-r} \varepsilon_{t-(s-r)} b_1 \varepsilon_{t-1} & \cdots & +\left(b_{s-r}^2 \varepsilon_{t-(s-r)}^2 - b_{s-r}^2 \sigma^2\right) \end{pmatrix}$$

$$\theta_2 = \begin{pmatrix} \left(b_{k-s}^2 \varepsilon_t^2 - b_{k-s}^2 \sigma^2\right) & +b_{k-s} \varepsilon_t b_{k-s+1} \varepsilon_{t-1} & \cdots & +b_{k-s} \varepsilon_t b_{k-r} \varepsilon_{t-(s-r)} \\ +b_{k-s+1} \varepsilon_{t-1} b_{k-s} \varepsilon_t & +\left(b_{k-s+1}^2 \varepsilon_{t-1}^2 - b_{k-s+1}^2 \sigma^2\right) & \cdots & +b_{k-s+1} \varepsilon_{t-1} b_{k-r} \varepsilon_{t-(s-r)} \\ \vdots & \vdots & \ddots & \vdots \\ +b_{k-r} \varepsilon_{t-(s-r)} b_{k-s} \varepsilon_t & +b_{k-r} \varepsilon_{t-(s-r)} b_{k-s+1} \varepsilon_{t-1} & \cdots & +\left(b_{k-r}^2 \varepsilon_{t-(s-r)}^2 - b_{k-r}^2 \sigma^2\right) \end{pmatrix}.$$

θ_1 and θ_2 each consist of $(s-r+1)^2$ summands. The expectation of the product of these summands only differs from zero if summand i of θ_1 is multiplied by summand i of θ_2 with $i = \left(1, 2, \ldots, (s-r+1)^2\right)$. If summand i of θ_1 is multiplied by summand j of θ_2 with $i \neq j$, this product contains a term $\varepsilon_\tau \varepsilon_{\tau+\tau_1} \varepsilon_{\tau+\tau_2} \varepsilon_{\tau+\tau_3}$ with $\tau_1 \neq 0, \tau_2 \neq 0, \tau_3 \neq 0$. Since the expectation of ε_τ is zero, and ε_τ is uncorrelated with the shocks $\varepsilon_{\tau+\tau_1}, \varepsilon_{\tau+\tau_2}$ and $\varepsilon_{\tau+\tau_3}$ the expectation of the product of these summands equals zero. It follows that $E[\theta_1\theta_2]$ simplifies to

$$E[\theta_1\theta_2] =$$

$$E\begin{bmatrix} b_0^2 b_{k-s}^2 \left(\varepsilon_t^2 - \sigma^2\right)^2 & +b_0 b_1 b_{k-s} b_{k-s+1} \varepsilon_t^2 \varepsilon_{t-1}^2 & \cdots & +b_0 b_{s-r} b_{k-s} b_{k-r} \varepsilon_t^2 \varepsilon_{t-(s-r)}^2 \\ +b_1 b_0 b_{k-s+1} b_{k-s} \varepsilon_{t-1}^2 \varepsilon_t^2 & +b_1^2 b_{k-s+1}^2 \left(\varepsilon_{t-1}^2 - \sigma^2\right)^2 & \cdots & +b_1 b_{s-r} b_{k-s+1} b_{k-r} \varepsilon_{t-1}^2 \varepsilon_{t-(s-r)}^2 \\ \vdots & \vdots & \ddots & \vdots \\ +b_{s-r} b_0 b_{k-r} b_{k-s} \varepsilon_{t-(s-r)}^2 \varepsilon_t^2 & +b_{s-r} b_1 b_{k-r} b_{k-s+1} \varepsilon_{t-(s-r)}^2 \varepsilon_{t-1}^2 & \cdots & +b_{s-r}^2 b_{k-r}^2 \left(\varepsilon_{t-(s-r)}^2 - \sigma^2\right)^2 \end{bmatrix}.$$

Noting that $E\left[\left(\varepsilon_t^2 - \sigma^2\right)^2\right]$ equals

$$E\left[\left(\varepsilon_t^2 - \sigma^2\right)^2\right] = E\left[\varepsilon_t^4\right] - 2E\left[\varepsilon_t^2\right]\sigma^2 + \sigma^4 = E\left[\varepsilon_t^4\right] - \sigma^4$$

yields the final result

$$\begin{aligned} & E\left[\left(e_{t_2,t_1}^2 - E\left(e_{t_2,t_1}^2\right)\right)\left(e_{t_4,t_3}^2 - E\left(e_{t_4,t_3}^2\right)\right)\right] \\ = & \left(E\left[\varepsilon_t^4\right] - \sigma^4\right) \sum_{i=0}^{s-r} b_i^2 b_{k-s+i}^2 + 2\sigma^4 \sum_{i=0}^{s-r} \sum_{j=0, j\neq i}^{s-r} b_i b_{k-s+i} b_j b_{k-s+j} \end{aligned}$$

where the first summation term collects all elements on the "main diagonal" of $E[\theta_1\theta_2]$, and the second term collects all "off-diagonal" elements.

A.2 Parameter Independence of the SUR Estimator

Suppose that $H = 2$. In this case, following the setup of Im (1994), the regressor matrix can be written as

$$\mathbf{X} = \begin{bmatrix} \mathbf{1}_N & \mathbf{0}_N \\ \mathbf{0}_{N-1} & \mathbf{1}_{N-1} \end{bmatrix} = \begin{bmatrix} 1 & 0 \\ \mathbf{1}_{N-1} & \mathbf{0}_{N-1} \\ \mathbf{0}_{N-1} & \mathbf{1}_{N-1} \end{bmatrix}$$

and the SUR covariance matrix equals

$$\begin{aligned} \mathbf{\Omega}_{SUR} &= \begin{bmatrix} \mathbf{G}_{0,0} & \mathbf{G}_{1,1}' \\ \mathbf{G}_{1,1} & \mathbf{G}_{1,0} \end{bmatrix} \\ &= \begin{bmatrix} g_1 \mathbf{I}_N & \begin{bmatrix} \mathbf{0}_{N-1} & g_1 \mathbf{I}_{N-1} \end{bmatrix}' \\ \begin{bmatrix} \mathbf{0}_{N-1} & g_1 \mathbf{I}_{N-1} \end{bmatrix} & g_2 \mathbf{I}_{N-1} \end{bmatrix} \\ &= \begin{bmatrix} g_1 & \mathbf{0}_{N-1}' & \mathbf{0}_{N-1}' \\ \mathbf{0}_{N-1} & g_1 \mathbf{I}_{N-1} & g_1 \mathbf{I}_{N-1} \\ \mathbf{0}_{N-1} & g_1 \mathbf{I}_{N-1} & g_2 \mathbf{I}_{N-1} \end{bmatrix}. \end{aligned}$$

Inverting the covariance matrix yields

$$\mathbf{\Omega}_{SUR}^{-1} = \begin{bmatrix} \frac{1}{g_1} & \mathbf{0}_{N-1}' & \mathbf{0}_{N-1}' \\ \mathbf{0}_{N-1} & -\frac{g_2}{g_1(g_1-g_2)} \mathbf{I}_{N-1} & \frac{1}{g_1-g_2} \mathbf{I}_{N-1} \\ \mathbf{0}_{N-1} & \frac{1}{g_1-g_2} \mathbf{I}_{N-1} & -\frac{1}{g_1-g_2} \mathbf{I}_{N-1} \end{bmatrix}.$$

The expression $\mathbf{X}'\mathbf{\Omega}_{SUR}^{-1}\mathbf{X}$ can be transformed as follows

$$\begin{aligned}
&\mathbf{X}'\mathbf{\Omega}_{SUR}^{-1}\mathbf{X} \\
= &\begin{bmatrix} 1 & \mathbf{1}'_{N-1} & \mathbf{0}'_{N-1} \\ 0 & \mathbf{0}'_{N-1} & \mathbf{1}'_{N-1} \end{bmatrix}
\begin{bmatrix} \frac{1}{g_1} & \mathbf{0}'_{N-1} & \mathbf{0}'_{N-1} \\ \mathbf{0}_{N-1} & -\frac{g_2}{g_1(g_1-g_2)}\mathbf{I}_{N-1} & \frac{1}{g_1-g_2}\mathbf{I}_{N-1} \\ \mathbf{0}_{N-1} & \frac{1}{g_1-g_2}\mathbf{I}_{N-1} & -\frac{1}{g_1-g_2}\mathbf{I}_{N-1} \end{bmatrix}
\begin{bmatrix} 1 & 0 \\ \mathbf{1}_{N-1} & \mathbf{0}_{N-1} \\ \mathbf{0}_{N-1} & \mathbf{1}_{N-1} \end{bmatrix} \\
= &\begin{bmatrix} \frac{1}{g_1} & -\frac{g_2}{g_1(g_1-g_2)}\mathbf{1}'_{N-1} & \frac{1}{g_1-g_2}\mathbf{1}'_{N-1} \\ 0 & \frac{1}{g_1-g_2}\mathbf{1}'_{N-1} & -\frac{1}{g_1-g_2}\mathbf{1}'_{N-1} \end{bmatrix}
\begin{bmatrix} 1 & 0 \\ \mathbf{1}_{N-1} & \mathbf{0}_{N-1} \\ \mathbf{0}_{N-1} & \mathbf{1}_{N-1} \end{bmatrix} \\
= &\begin{bmatrix} \frac{1}{g_1} - \frac{g_2}{g_1(g_1-g_2)}\mathbf{1}'_{N-1}\mathbf{1}_{N-1} & \frac{1}{g_1-g_2}\mathbf{1}'_{N-1}\mathbf{1}_{N-1} \\ \frac{1}{g_1-g_2}\mathbf{1}'_{N-1}\mathbf{1}_{N-1} & -\frac{1}{g_1-g_2}\mathbf{1}'_{N-1}\mathbf{1}_{N-1} \end{bmatrix} \\
= &\begin{bmatrix} \frac{1}{g_1}\left(1 - \frac{g_2}{g_1-g_2}(N-1)\right) & \frac{1}{g_1-g_2}(N-1) \\ \frac{1}{g_1-g_2}(N-1) & -\frac{1}{g_1-g_2}(N-1) \end{bmatrix}.
\end{aligned}$$

Since the determinant of $\mathbf{X}'\mathbf{\Omega}_{SUR}^{-1}\mathbf{X}$ is given by

$$\det\left(\mathbf{X}'\mathbf{\Omega}_{SUR}^{-1}\mathbf{X}\right) = -\frac{N}{g_1(g_1-g_2)}(N-1)$$

the inverse of $\mathbf{X}'\mathbf{\Omega}_{SUR}^{-1}\mathbf{X}$ equals

$$\left(\mathbf{X}'\mathbf{\Omega}_{SUR}^{-1}\mathbf{X}\right)^{-1} = \begin{bmatrix} \frac{1}{N}g_1 & \frac{1}{N}g_1 \\ \frac{1}{N}g_1 & -\frac{1}{N}\frac{g_1-Ng_2}{N-1} \end{bmatrix}.$$

When the inverse of $\mathbf{X}'\mathbf{\Omega}_{SUR}^{-1}\mathbf{X}$ is multiplied by $\mathbf{X}'\mathbf{\Omega}_{SUR}^{-1}$, we finally get

$$\begin{aligned}
&\left(\mathbf{X}'\mathbf{\Omega}_{SUR}^{-1}\mathbf{X}\right)^{-1}\mathbf{X}'\mathbf{\Omega}_{SUR}^{-1} \\
= &\begin{bmatrix} \frac{1}{N}g_1 & \frac{1}{N}g_1 \\ \frac{1}{N}g_1 & -\frac{1}{N}\frac{g_1-Ng_2}{N-1} \end{bmatrix}
\begin{bmatrix} \frac{1}{g_1} & -\frac{g_2}{g_1(g_1-g_2)}\mathbf{1}'_{N-1} & \frac{1}{g_1-g_2}\mathbf{1}'_{N-1} \\ 0 & \frac{1}{g_1-g_2}\mathbf{1}'_{N-1} & -\frac{1}{g_1-g_2}\mathbf{1}'_{N-1} \end{bmatrix} \\
= &\begin{bmatrix} \frac{1}{N}g_1\frac{1}{g_1} & \left(\frac{1}{N}g_1\left(-\frac{g_2}{g_1(g_1-g_2)}\right) + \frac{1}{N}g_1\frac{1}{g_1-g_2}\right)\mathbf{1}'_{N-1} & \left(\frac{1}{N}g_1\frac{1}{g_1-g_2} - \frac{1}{N}g_1\frac{1}{g_1-g_2}\right)\mathbf{1}'_{N-1} \\ \frac{1}{N}g_1\frac{1}{g_1} & \left(\frac{1}{N}g_1\left(-\frac{g_2}{g_1(g_1-g_2)}\right) - \frac{1}{N}\frac{g_1-Ng_2}{N-1}\frac{1}{g_1-g_2}\right)\mathbf{1}'_{N-1} & \left(\frac{1}{N}g_1\frac{1}{g_1-g_2} + \frac{1}{N}\frac{g_1-Ng_2}{N-1}\frac{1}{g_1-g_2}\right)\mathbf{1}'_{N-1} \end{bmatrix} \\
= &\begin{bmatrix} \frac{1}{N} & \frac{1}{N}\mathbf{1}'_{N-1} & \mathbf{0}'_{N-1} \\ \frac{1}{N} & -\frac{1}{N(N-1)}\mathbf{1}'_{N-1} & \frac{1}{N-1}\mathbf{1}'_{N-1} \end{bmatrix}.
\end{aligned}$$

A.3 FGLS Estimation

The estimation of the covariance matrix used for the FGLS estimation is implemented as follows: First the OLS estimates of forecast uncertainty are used to calculate the vector of demeaned squared forecast errors $\mathbf{u} = \left(u_1, u_2, \ldots, u_{\frac{1}{2}H(2N-H+1)}\right)$ as

$$u_i = e^2_{t+h,h} - \hat{\sigma}^2_{OLS,h}$$

where the elements are ordered as in (6). Then, the matrix $\mathbf{u}'\mathbf{u}$ is calculated. The lower triangular elements of this matrix and the restrictions derived for the case of optimal forecasts are used to estimate the covariance matrix of the squared forecast errors. Using the example with $N = 3$ and $H = 2$, the lower triangular elements of the matrix $\mathbf{u}'\mathbf{u}$ are given by

$$\begin{bmatrix} u_1^2 & & & & \\ u_1u_2 & u_2^2 & & & \\ u_1u_3 & u_2u_3 & u_3^2 & & \\ u_1u_4 & u_2u_4 & u_3u_4 & u_4^2 & \\ u_1u_5 & u_2u_5 & u_3u_5 & u_4u_5 & u_5^2 \end{bmatrix}.$$

Due to the restrictions implied by (13), the estimated covariance matrix $\hat{\Omega}$ used for the GLS estimation is determined by

$$\hat{\mathbf{\Omega}} = \begin{bmatrix} \hat{\omega}_1 & 0 & 0 & \hat{\omega}_3 & 0 \\ 0 & \hat{\omega}_1 & 0 & \hat{\omega}_1 & \hat{\omega}_3 \\ 0 & 0 & \hat{\omega}_1 & 0 & \hat{\omega}_1 \\ \hat{\omega}_3 & \hat{\omega}_1 & 0 & \hat{\omega}_2 & \hat{\omega}_3 \\ 0 & \hat{\omega}_3 & \hat{\omega}_1 & \hat{\omega}_3 & \hat{\omega}_2 \end{bmatrix}$$

with

$$\begin{aligned} \hat{\omega}_1 &= \frac{1}{5-1}\left(u_1^2 + u_2^2 + u_3^2 + u_2u_4 + u_3u_5\right) \\ \hat{\omega}_2 &= \frac{1}{2-1}\left(u_4^2 + u_5^2\right) \\ \hat{\omega}_3 &= \frac{1}{3-1}\left(u_1u_4 + u_2u_5 + u_4u_5\right). \end{aligned}$$

Based on $\hat{\boldsymbol{\Omega}}$, the GLS estimator $\hat{\boldsymbol{\sigma}}^2_{GLS}$ can be obtained. In principle, one could start an iterative process by calculating new u_i's based on the GLS estimates $\hat{\sigma}^2_{GLS,h}$ instead of $\hat{\sigma}^2_{OLS,h}$ and repeating this until $\hat{\boldsymbol{\sigma}}^2_{GLS}$ converges. However, since the first estimator $\hat{\boldsymbol{\sigma}}^2_{GLS}$ can sometimes deviate strongly from $\hat{\boldsymbol{\sigma}}^2_{OLS}$ and from the true vector $\boldsymbol{\sigma}^2$, this approach does not seem too promising and is therefore not pursued here.

The following Discussion Papers have been published since 2008:

Series 1: Economic Studies

01	2008	Can capacity constraints explain asymmetries of the business cycle?	Malte Knüppel
02	2008	Communication, decision-making and the optimal degree of transparency of monetary policy committees	Anke Weber
03	2008	The impact of thin-capitalization rules on multinationals' financing and investment decisions	Buettner, Overesch Schreiber, Wamser
04	2008	Comparing the DSGE model with the factor model: an out-of-sample forecasting experiment	Mu-Chun Wang
05	2008	Financial markets and the current account – emerging Europe versus emerging Asia	Sabine Herrmann Adalbert Winkler
06	2008	The German sub-national government bond market: evolution, yields and liquidity	Alexander Schulz Guntram B. Wolff
07	2008	Integration of financial markets and national price levels: the role of exchange rate volatility	Mathias Hoffmann Peter Tillmann
08	2008	Business cycle evidence on firm entry	Vivien Lewis
09	2008	Panel estimation of state dependent adjustment when the target is unobserved	Ulf von Kalckreuth
10	2008	Nonlinear oil price dynamics – a tale of heterogeneous speculators?	Stefan Reitz Ulf Slopek
11	2008	Financing constraints, firm level adjustment of capital and aggregate implications	Ulf von Kalckreuth

12	2008	Sovereign bond market integration: the euro, trading platforms and globalization	Alexander Schulz Guntram B. Wolff
13	2008	Great moderation at the firm level? Unconditional versus conditional output volatility	Claudia M. Buch Jörg Döpke Kerstin Stahn
14	2008	How informative are macroeconomic risk forecasts? An examination of the Bank of England's inflation forecasts	Malte Knüppel Guido Schultefrankenfeld
15	2008	Foreign (in)direct investment and corporate taxation	Georg Wamser
16	2008	The global dimension of inflation – evidence from factor-augmented Phillips curves	Sandra Eickmeier Katharina Moll
17	2008	Global business cycles: convergence or decoupling?	M. Ayhan Kose Christopher Otrok, Ewar Prasad
18	2008	Restrictive immigration policy in Germany: pains and gains foregone?	Gabriel Felbermayr Wido Geis Wilhelm Kohler
19	2008	International portfolios, capital accumulation and foreign assets dynamics	Nicolas Coeurdacier Robert Kollmann Philippe Martin
20	2008	Financial globalization and monetary policy	Michael B. Devereux Alan Sutherland
21	2008	Banking globalization, monetary transmission and the lending channel	Nicola Cetorelli Linda S. Goldberg
22	2008	Financial exchange rates and international currency exposures	Philip R. Lane Jay C. Shambaugh

23	2008	Financial integration, specialization and systemic risk	F. Fecht, H. P. Grüner P. Hartmann
24	2008	Sectoral differences in wage freezes and wage cuts: evidence from a new firm survey	Daniel Radowski Holger Bonin
25	2008	Liquidity and the dynamic pattern of price adjustment: a global view	Ansgar Belke Walter Orth, Ralph Setzer
26	2008	Employment protection and temporary work agencies	Florian Baumann Mario Mechtel, Nikolai Stähler
27	2008	International financial markets' influence on the welfare performance of alternative exchange rate regimes	Mathias Hoffmann
28	2008	Does regional redistribution spur growth?	M. Koetter, M. Wedow
29	2008	International financial competitiveness and incentives to foreign direct investment	Axel Jochem
30	2008	The price of liquidity: bank characteristics and market conditions	Falko Fecht Kjell G. Nyborg, Jörg Rocholl
01	2009	Spillover effects of minimum wages in a two-sector search model	Christoph Moser Nikolai Stähler
02	2009	Who is afraid of political risk? Multinational firms and their choice of capital structure	Iris Kesternich Monika Schnitzer
03	2009	Pooling versus model selection for nowcasting with many predictors: an application to German GDP	Vladimir Kuzin Massimiliano Marcellino Christian Schumacher

04	2009	Fiscal sustainability and policy implications for the euro area	Balassone, Cunha, Langenus Manzke, Pavot, Prammer Tommasino
05	2009	Testing for structural breaks in dynamic factor models	Jörg Breitung Sandra Eickmeier
06	2009	Price convergence in the EMU? Evidence from micro data	Christoph Fischer
07	2009	MIDAS versus mixed-frequency VAR: nowcasting GDP in the euro area	V. Kuzin, M. Marcellino C. Schumacher
08	2009	Time-dependent pricing and New Keynesian Phillips curve	Fang Yao
09	2009	Knowledge sourcing: legitimacy deficits for MNC subsidiaries?	Tobias Schmidt Wolfgang Sofka
10	2009	Factor forecasting using international targeted predictors: the case of German GDP	Christian Schumacher
11	2009	Forecasting national activity using lots of international predictors: an application to New Zealand	Sandra Eickmeier Tim Ng
12	2009	Opting out of the great inflation: German monetary policy after the breakdown of Bretton Woods	Andreas Beyer, Vitor Gaspar Christina Gerberding Otmar Issing
13	2009	Financial intermediation and the role of price discrimination in a two-tier market	Stefan Reitz Markus A. Schmidt, Mark P. Taylor
14	2009	Changes in import pricing behaviour: the case of Germany	Kerstin Stahn

15	2009	Firm-specific productivity risk over the business cycle: facts and aggregate implications	Ruediger Bachmann Christian Bayer
16	2009	The effects of knowledge management on innovative success – an empirical analysis of German firms	Uwe Cantner Kristin Joel Tobias Schmidt
17	2009	The cross-section of firms over the business cycle: new facts and a DSGE exploration	Ruediger Bachmann Christian Bayer
18	2009	Money and monetary policy transmission in the euro area: evidence from FAVAR- and VAR approaches	Barno Blaes
19	2009	Does lowering dividend tax rates increase dividends repatriated? Evidence of intra-firm cross-border dividend repatriation policies by German multinational enterprises	Christian Bellak Markus Leibrecht Michael Wild
20	2009	Export-supporting FDI	Sebastian Krautheim
21	2009	Transmission of nominal exchange rate changes to export prices and trade flows and implications for exchange rate policy	Mathias Hoffmann Oliver Holtemöller
22	2009	Do we really know that flexible exchange rates facilitate current account adjustment? Some new empirical evidence for CEE countries	Sabine Herrmann
23	2009	More or less aggressive? Robust monetary policy in a New Keynesian model with financial distress	Rafael Gerke Felix Hammermann Vivien Lewis
24	2009	The debt brake: business cycle and welfare consequences of Germany's new fiscal policy rule	Eric Mayer Nikolai Stähler

25	2009	Price discovery on traded inflation expectations: Does the financial crisis matter?	Alexander Schulz Jelena Stapf
26	2009	Supply-side effects of strong energy price hikes in German industry and transportation	Thomas A. Knetsch Alexander Molzahn
27	2009	Coin migration within the euro area	Franz Seitz, Dietrich Stoyan Karl-Heinz Tödter
28	2009	Efficient estimation of forecast uncertainty based on recent forecast errors	Malte Knüppel

Series 2: Banking and Financial Studies

01	2008	Analyzing the interest rate risk of banks using time series of accounting-based data: evidence from Germany	O. Entrop, C. Memmel M. Wilkens, A. Zeisler
02	2008	Bank mergers and the dynamics of deposit interest rates	Ben R. Craig Valeriya Dinger
03	2008	Monetary policy and bank distress: an integrated micro-macro approach	F. de Graeve T. Kick, M. Koetter
04	2008	Estimating asset correlations from stock prices or default rates – which method is superior?	K. Düllmann J. Küll, M. Kunisch
05	2008	Rollover risk in commercial paper markets and firms' debt maturity choice	Felix Thierfelder
06	2008	The success of bank mergers revisited – an assessment based on a matching strategy	Andreas Behr Frank Heid
07	2008	Which interest rate scenario is the worst one for a bank? Evidence from a tracking bank approach for German savings and cooperative banks	Christoph Memmel
08	2008	Market conditions, default risk and credit spreads	Dragon Yongjun Tang Hong Yan
09	2008	The pricing of correlated default risk: evidence from the credit derivatives market	Nikola Tarashev Haibin Zhu
10	2008	Determinants of European banks' engagement in loan securitization	Christina E. Bannier Dennis N. Hänsel
11	2008	Interaction of market and credit risk: an analysis of inter-risk correlation and risk aggregation	Klaus Böcker Martin Hillebrand

12	2008	A value at risk analysis of credit default swaps	B. Raunig, M. Scheicher
13	2008	Systemic bank risk in Brazil: an assessment of correlated market, credit, sovereign and inter-bank risk in an environment with stochastic volatilities and correlations	Theodore M. Barnhill, Jr. Marcos Rietti Souto
14	2008	Regulatory capital for market and credit risk inter-action: is current regulation always conservative?	T. Breuer, M. Jandačka K. Rheinberger, M. Summer
15	2008	The implications of latent technology regimes for competition and efficiency in banking	Michael Koetter Tigran Poghosyan
16	2008	The impact of downward rating momentum on credit portfolio risk	André Güttler Peter Raupach
17	2008	Stress testing of real credit portfolios	F. Mager, C. Schmieder
18	2008	Real estate markets and bank distress	M. Koetter, T. Poghosyan
19	2008	Stochastic frontier analysis by means of maxi-mum likelihood and the method of moments	Andreas Behr Sebastian Tente
20	2008	Sturm und Drang in money market funds: when money market funds cease to be narrow	Stehpan Jank Michael Wedow
01	2009	Dominating estimators for the global minimum variance portfolio	Gabriel Frahm Christoph Memmel
02	2009	Stress testing German banks in a downturn in the automobile industry	Klaus Düllmann Martin Erdelmeier
03	2009	The effects of privatization and consolidation on bank productivity: comparative evidence from Italy and Germany	E. Fiorentino A. De Vincenzo, F. Heid A. Karmann, M. Koetter

04	2009	Shocks at large banks and banking sector distress: the Banking Granular Residual	Sven Blank, Claudia M. Buch Katja Neugebauer
05	2009	Why do savings banks transform sight deposits into illiquid assets less intensively than the regulation allows?	Dorothee Holl Andrea Schertler
06	2009	Does banks' size distort market prices? Evidence for too-big-to-fail in the CDS market	Manja Völz Michael Wedow
07	2009	Time dynamic and hierarchical dependence modelling of an aggregated portfolio of trading books – a multivariate nonparametric approach	Sandra Gaisser Christoph Memmel Rafael Schmidt Carsten Wehn
08	2009	Financial markets' appetite for risk – and the challenge of assessing its evolution by risk appetite indicators	Birgit Uhlenbrock
09	2009	Income diversification in the German banking industry	Ramona Busch Thomas Kick
10	2009	The dark and the bright side of liquidity risks: evidence from open-end real estate funds in Germany	Falko Fecht Michael Wedow
11	2009	Determinants for using visible reserves in German banks – an empirical study	Bornemann, Homölle Hubensack, Kick, Pfingsten

Visiting researcher at the Deutsche Bundesbank

The Deutsche Bundesbank in Frankfurt is looking for a visiting researcher. Among others under certain conditions visiting researchers have access to a wide range of data in the Bundesbank. They include micro data on firms and banks not available in the public. Visitors should prepare a research project during their stay at the Bundesbank. Candidates must hold a PhD and be engaged in the field of either macroeconomics and monetary economics, financial markets or international economics. Proposed research projects should be from these fields. The visiting term will be from 3 to 6 months. Salary is commensurate with experience.

Applicants are requested to send a CV, copies of recent papers, letters of reference and a proposal for a research project to:

Deutsche Bundesbank
Personalabteilung
Wilhelm-Epstein-Str. 14

60431 Frankfurt
GERMANY