Conference on

Forecasting and Monetary Policy

Berlin, 23-24 March 2009

Malte Knüppel
Deutsche Bundesbank

„Efficient Estimation of Forecast Uncertainty“

Efficient Estimation of Forecast Uncertainty

Malte Knüppel

Deutsche Bundesbank

23 March 2009

Introduction

- Many institutions publishing macroeconomic forecasts give assessments of forecast uncertainties
- In most cases, forecasts are made for several horizons
- Forecasting uncertainty is difficult $\Rightarrow$ Many institutions rely on past forecast errors
- Apparently, past forecast uncertainty is estimated for each horizon separately
- Sample of forecast errors might be small, especially for large horizons, and forecast errors of large horizons are often strongly autocorrelated $\Rightarrow$ possibly large uncertainty about past forecast uncertainty, especially for large horizons
- Question adressed here:
  Can efficiency gains be achieved by joint estimation of forecast uncertainty for all horizons?

Optimal Forecasts

- Error of optimal forecast $y_{t+h,t}$ made in period t for period $t+h$ has MA($h-1$)-representation

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

with $E[\varepsilon_t] = 0$, $E[\varepsilon_t^2] = \sigma^2$, $b_0 = 1$, $E[\varepsilon_t^4] < \infty$.
The ε_t's are the shocks to the data-generating process (DGP), the b_i's depend on the DGP.

- Forecast uncertainty for horizon h measured by $E[e^2_{t+h,t}]$ equals

$$E[e^2_{t+h,t}] = \sigma^2_h = \sigma^2 \sum_{i=0}^{h-1} b_i^2.$$

- OLS estimator of forecast uncertainty: regression of $e^2_{t+h,t}$ on a constant for each horizon h, yields sample means of $e^2_{t+h,t}$

Covariances of Squared Forecast Errors and GLS

- Define $\omega := cov\left(e^2_{j,i} e^2_{l,k}\right)$

	$e_{3,0}$	$e_{5,1}$	$e_{3,0}$	$e_{5,3}$	$e_{3,0}$	$e_{3,1}$
$t=1$	$= b_2\varepsilon_1$		$= b_2\varepsilon_1$		$= b_2\varepsilon_1$	
$t=2$	$+b_1\varepsilon_2$	$= b_3\varepsilon_2$	$+b_1\varepsilon_2$		$+b_1\varepsilon_2$	$= b_1\varepsilon_2$
$t=3$	$+\varepsilon_3$	$+b_2\varepsilon_3$	$+\varepsilon_3$		$+\varepsilon_3$	$+\varepsilon_3$
$t=4$		$+b_1\varepsilon_4$		$= b_1\varepsilon_4$		
$t=5$		$+\varepsilon_5$		$+\varepsilon_5$		
ω	$\neq 0$		$= 0$		$\neq 0$	

- (Squared) forecast errors with overlapping shocks are correlated $\Rightarrow$ efficiency gains of GLS over OLS estimation
- Non-zero covariances depend on b_i's, σ, and kurtosis of ε_t

Covariances of Squared Forecast Errors and SUR

- SUR estimation: Assume $\omega = cov\left(e^2_{j,i} e^2_{l,k}\right) = 0$ unless $j = l$

	$e_{3,0}$	$e_{5,1}$	$e_{3,0}$	$e_{5,3}$	$e_{3,0}$	$e_{3,1}$
$t=1$	$= b_2\varepsilon_1$		$= b_2\varepsilon_1$		$= b_2\varepsilon_1$	
$t=2$	$+b_1\varepsilon_2$	$= b_3\varepsilon_2$	$+b_1\varepsilon_2$		$+b_1\varepsilon_2$	$= b_1\varepsilon_2$
$t=3$	$+\varepsilon_3$	$+b_2\varepsilon_3$	$+\varepsilon_3$		$+\varepsilon_3$	$+\varepsilon_3$
$t=4$		$+b_1\varepsilon_4$		$= b_1\varepsilon_4$		
$t=5$		$+\varepsilon_5$		$+\varepsilon_5$		
ω	$= 0$		$= 0$		$\neq 0$	

- SUR supposedly works well if $b_1, b_2, \ldots$ are small, i.e. if persistence is small

Efficiency Gains and Sample Size - Examples

- Regression on a constant ⇒ OLS estimation asymptotically efficient (Grenander and Rosenblatt, 1957)
  ⇒ GLS and SUR estimation only interesting in small samples

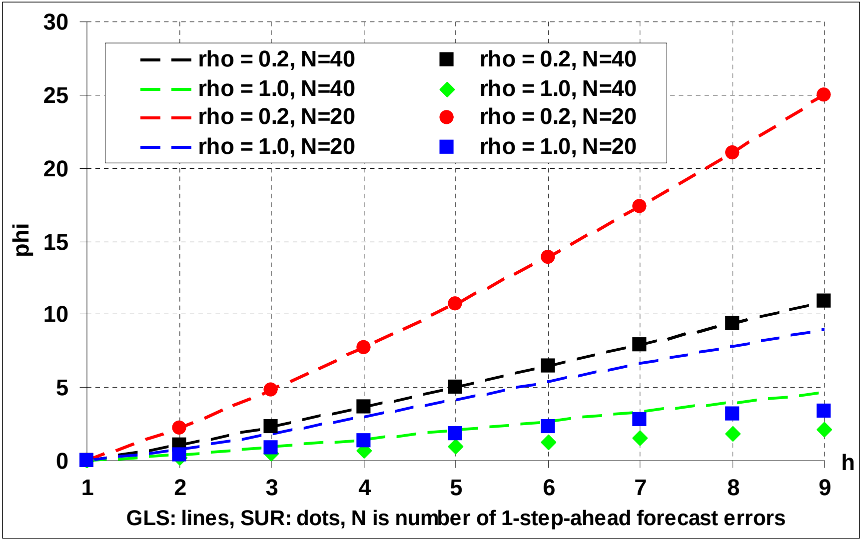
How small is a small sample? Consider following example:

- b_i's and kurtosis assumed to be known ⇒ covariance matrix of squared forecast errors known (except for σ)
- Measure of efficiency gains based on ratio of standard errors

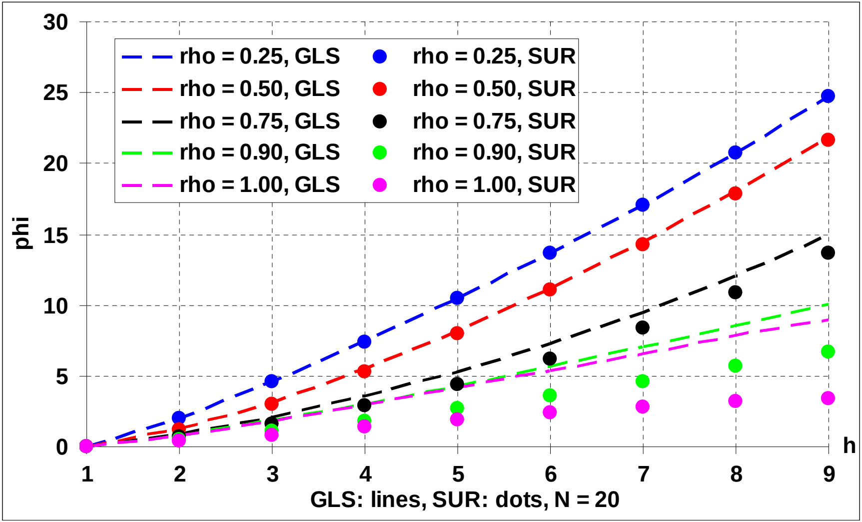
$$\varphi_h := 100 \ln \left(se\left(\hat{\sigma}^2_{OLS,h}\right) / se\left(\hat{\sigma}^2_{GLS \text{ or } SUR,h}\right)\right)$$

  with $se\left(\hat{\sigma}^2_{j,h}\right)$ the standard error of the estimate obtained with method j for σ^2_h, $h = 1, 2, \ldots H$
- DGP is AR(1)-process $y_t = \rho y_{t-1} + \varepsilon_t$, ε_t i.i.d. $N\left(0, \sigma^2\right)$
- Largest forecast horizon $H = 9$
- Data structure investigated in the following: For forecast horizon h, 1 more forecast error available than for forecast horizon $h + 1$

Efficiency Gains and Sample Size - Examples



Efficiency Gains and Sample Size - Examples



Efficiency Gains

Efficiency gains...

- increase with forecast horizon
- are zero for smallest forecast horizon
- decrease with persistence (but not linearly)
- of SUR are close to those of GLS if persistence is low
- of SUR are positive even for random walk
- hardly exceed 10% for $N = 40$ and $H = 9$

- do not depend on the inclusion of errors from larger horizons
  $\Rightarrow$ estimate for horizon h only employs information of g-step-ahead forecast errors with $g \leq h$

All results can change when different data structures are studied

Problems in Practice
Unknown covariance matrix

- Notation: Estimates determined by $\hat{\sigma}_i^2 = \mathbf{A}_i \mathbf{e}^2$, with $i \in \{GLS, SUR, OLS\}$, and $\mathbf{e}^2$ a vector containing all squared forecast errors
- GLS
  - b_i's can be calculated (estimated) from forecast errors if forecasts are (almost) optimal
  - $\mathbf{A}_{GLS}$ turns out to be independent of other parameters
    $\Rightarrow$ (Almost) same results as with known covariance matrix
  - Direct estimation of covariance matrix in general leads to large efficiency losses
- SUR
  - $\mathbf{A}_{SUR}$ turns out to be independent of all parameters
    $\Rightarrow$ *Covariance matrix of squared forecast errors does not have to be estimated, just a certain structure needs to be imposed (like OLS)*
  - Direct estimation of covariance matrix unnecessary
  - Efficiency gains depend only on DGP in the case of optimal forecasts

- If forecasts are strongly non-optimal, use SUR estimator based on covariance matrix derived for optimal forecasts (no estimation uncertainty)
- Monte Carlo studies conducted with an AR(1) forecasting model and several DGPs, implying the following misspecifications
    - dynamic misspecifications
    - bias
    - conditional heteroskedasticity

Problems in Practice

Strongly non-optimal forecasts

- Reason for non-optimality: simple dynamic misspecification
- Efficiency gains of SUR estimator with $N = 20$ and
  - DGP: $y_t = \rho y_{t-1} + \varepsilon_t$, ε_t i.i.d. $N\left(0, \sigma^2\right)$
  - forecasts: $\hat{y}_{t+h,t} = \hat{\rho}^h y_t$

	$\rho = 1$ $\hat{\rho} = 0.5$	$\rho = 2$ $\hat{\rho} = 0.5$	$\rho = 1$ $\hat{\rho} = 0.3$	$\rho = 0.5$ $\hat{\rho} = 0.71$	$\rho = 0.5$ $\hat{\rho} = 0.88$
$h = 1$	0.0	0.0	0.0	0.0	0.0
$h = 2$	0.3	3.3	0.4	0.9	0.5
$h = 3$	0.6	7.8	0.9	1.8	1.4
$h = 4$	1.0	13.2	1.5	3.5	1.4
$h = 5$	1.5	19.1	2.1	5.6	1.1
$h = 6$	2.0	25.5	2.7	7.7	0.5
$h = 7$	2.6	32.4	3.4	10.6	0.0
$h = 8$	3.1	39.8	4.0	14.2	0.4
$h = 9$	3.7	47.8	4.7	18.4	1.4

Problems in Practice

Strongly non-optimal forecasts

- Reason for non-optimality: bias
- Efficiency gains with $N = 20$, $\rho = 0.5$ and
  - DGP: $y_t = \mu + \rho y_{t-1} + \varepsilon_t$, ε_t i.i.d. $N(0,1)$
  - forecasts: $\hat{y}_{t+h,t} = \rho^h y_t$

	$\mu = 0$	$\mu = 0.3$	$\mu = 0.5$	$\mu = 100$
$h = 1$	0.0	0.0	0.0	0.0
$h = 2$	1.2	0.9	1.0	0.9
$h = 3$	3.2	2.5	2.5	2.4
$h = 4$	5.7	4.6	4.5	4.4
$h = 5$	8.6	7.0	7.1	6.5
$h = 6$	11.7	9.9	9.9	9.1
$h = 7$	14.7	13.1	12.8	12.0
$h = 8$	18.2	16.7	16.0	15.2
$h = 9$	21.8	20.3	19.8	18.8

Problems in Practice

Strongly non-optimal forecasts and the SUR estimator

Simulation results:

- Bias reduces efficiency gains, does not cause efficiency losses
- Simple dynamic misspecification can cause efficiency losses if misspecification is severe
    - Note: Severe dynamic misspecification is a necessary, but not a sufficient condition for efficiency losses
    - Similar results for other forms of dynamic misspecification (wrong lag order, undetected non-linearities, etc.)
- Conditional heteroskedasticity does apparently not lead to efficiency losses

Problems in Practice

Structural break of forecast uncertainty within the sample

- If σ_h^2 in the sample is function of time, e.g. due to structural break, SUR (and GLS) estimation more problematic than OLS
- Reason: Consider following sample of forecast errors

		forecast horizon h	
		1	2
forecast	1	$e_{1,0}$	
for	2	$e_{2,1}$	$e_{2,0}$
period	3	$e_{3,2}$	$e_{3,1}$

- Suppose structural break of σ_h^2 occurs in period 2. SUR and GLS estimator use $e_{1,0}^2$ to estimate σ_2^2
  $\Rightarrow$ SUR and GLS estimator biased. OLS estimator unbiased.
- In principle, GLS and SUR estimator could be used employing only $e_{2,1}^2, e_{3,2}^2, e_{2,0}^2, e_{3,1}^2$, but in this case SUR is identical to OLS, and GLS yields only minor efficiency gains even if covariances are known

Application

Inflation forecasts of the Bank of England

- Since 2004q3, Bank of England has published inflation nowcast for current quarter and inflation forecasts for next 12 quarters ($\Rightarrow H = 13$) once every quarter
- Using data until 2008q4, only 6 $(7, 8, 9)$ forecast errors are available for forecast horizon 13 $(12, 11, 10)$
- Plausibility of OLS and SUR estimates might be checked by estimating forecast uncertainty for the sample 1998q1 until 2008q4.
- For this larger sample, $H = 9$ until 2004q2, so forecast uncertainty for $h = 10, 11, 12, 13$ cannot be estimated. Nevertheless, results for larger sample should give indications about plausible uncertainty for these horizons.

  Note: Measure of inflation and inflation target changed in 2004q1, but forecast errors show no signs of structural break at this date.

Application

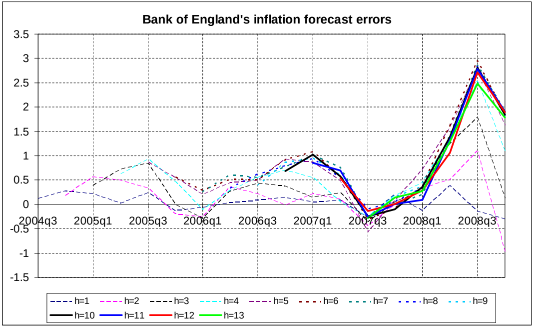
Inflation forecasts of the Bank of England

- Standard deviations of inflation forecast errors:

	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**			

Application

Inflation forecasts of the Bank of England



- Inflation was very high in 2008, leading to large forecast errors especially at longer forecast horizons
  $\Rightarrow$ OLS probably overestimates forecast uncertainty
- SUR estimation uses information of forecast errors since 2004q3 also for longer horizons, yields more plausible results (unless struc. break)

Conclusions and Outlook

- Squared multi-step forecast errors are correlated
- OLS estimation of forecast uncertainty is asymptotically efficient
- If sample is small and if forecasts are (almost) optimal, GLS estimation yields efficiency gains
- If sample is small and certain data structure is given
  - SUR estimation does not require estimation of a covariance matrix
  - SUR estimation often yields efficiency gains even if forecasts are strongly non-optimal
- If sample contains structural break of forecast uncertainty, OLS is less problematic than GLS and SUR
- For inflation forecasts of the Bank of England from 2004q3 to 2008q4, SUR and OLS estimator yield strongly differing results at long horizons. OLS estimator appears to overestimate forecast uncertainty
- Uncertainty about forecast uncertainty still needs to be investigated