.

.

# *Testing Forecast Optimality under Unknown Loss*

.

Andrew Patton
London School of Economics

Allan Timmermann
University of California, San Diego

.

.

May 2006

# 1 Forecast evaluation in economics and finance

- Tests of market efficiency and investor rationality are usually done by testing properties of forecast errors:
  - *Relating to the efficient markets hypothesis:* Cargill and Meyer (JF, 1980), De Bondt and Thaler (JFQA, 1992), Mishkin (AER, 1981), *inter alia.*
  - *Relating to rationality of decision-makers:* Brown and Maital (EMA, 1981), Figlewski and Watchel (REStat, 1981), Keane and Runkle (AER, 1990), Lakonishok (JF, 1980), *inter alia.*
- Forecasts of key economic variables (inflation, GDP growth, unemployment, etc) are an important input to decisions by central banks, fiscal authorities, international organisations.
- A large majority of past work on testing forecast optimality assumes squared error loss.

# 2 Is MSE the right loss function?

- The assumption of MSE loss in economics has been questioned in the (mostly) recent literature:

  - Granger (1969), Granger and Newbold (1986), West, Edison and Cho (1996), Granger and Pesaran (2000), Pesaran and Skouras (2001).

- For example: equity analysts' forecasts have been found to be biased upwards

  - A result of analyst irrationality, or simply that the analyst is penalised more heavily for under-predictions than over-predictions?

  - Elliott et al. (2004) show that macroeconomic forecasters may be rational, but not under MSE loss.

- Patton and Timmermann (2004) we show that none of the standard properties of optimal forecasts withstand the move to asymmetric loss, in general.

# 3 What we do in this paper

1. We present new testable properties of optimal forecasts that can be obtained when $L$ is *unknown*, subject to restrictions on the data generating process (DGP):

   (a) the variable exhibits dynamics only in its conditional mean and variance

   (b) the loss function can be approximated using a spline of some sort (extending Elliott, et al., 2004).

2. We study the Federal Reserve's "greenbook" quarterly forecasts of GDP growth over the period 1968Q4 to 1999Q4.

   (a) These forecasts are rejected for a broad class of loss functions based solely on the forecast error (including MSE)

   (b) We find optimality of the Fed forecasts only when we allow for a "conservative" loss function *and* when the degree of conservatism increases when growth is moderate or low.

# 4 Standard properties of optimal forecasts

- The properties of optimal forecasts in the standard framework are, see Diebold and Lopez (1996) for example:

1. Optimal forecasts are unbiased: $E_t\left[e^*_{t+h,t}\right] = 0$

2. Optimal $h$-step forecast errors are serially correlated only to lag $(h-1)$

    - So one-step forecast errors are white noise

3. Optimal forecast error variance is a non-decreasing function of the forecast horizon

# 5 Testable implications under MSE loss

- The most common tests of forecast optimality under MSE loss is the "Mincer-Zarnowitz" regression:

$$Y_{t+h} = \alpha + \beta \hat{Y}_{t+h,t} + u_{t+h}$$

$$\begin{aligned} H_0 &: \alpha = 0 \cap \beta = 1 \\ \text{vs } H_a &: \alpha \neq 0 \cup \beta \neq 1 \end{aligned}$$

- The results of this test for the Fed forecasts were:

$$\begin{aligned} Y_t &= \underset{(0.451)}{1.253} + \underset{(0.106)}{0.710}\, \hat{Y}_t + u_t. \\ R^2 &= 0.218 \end{aligned}$$

A Wald test of the joint restriction on the intercept and the slope yields a $\chi^2_2$-statistic of 8.57 and a $p$-value of 0.01. Thus we reject the null of forecast optimality under MSE loss.

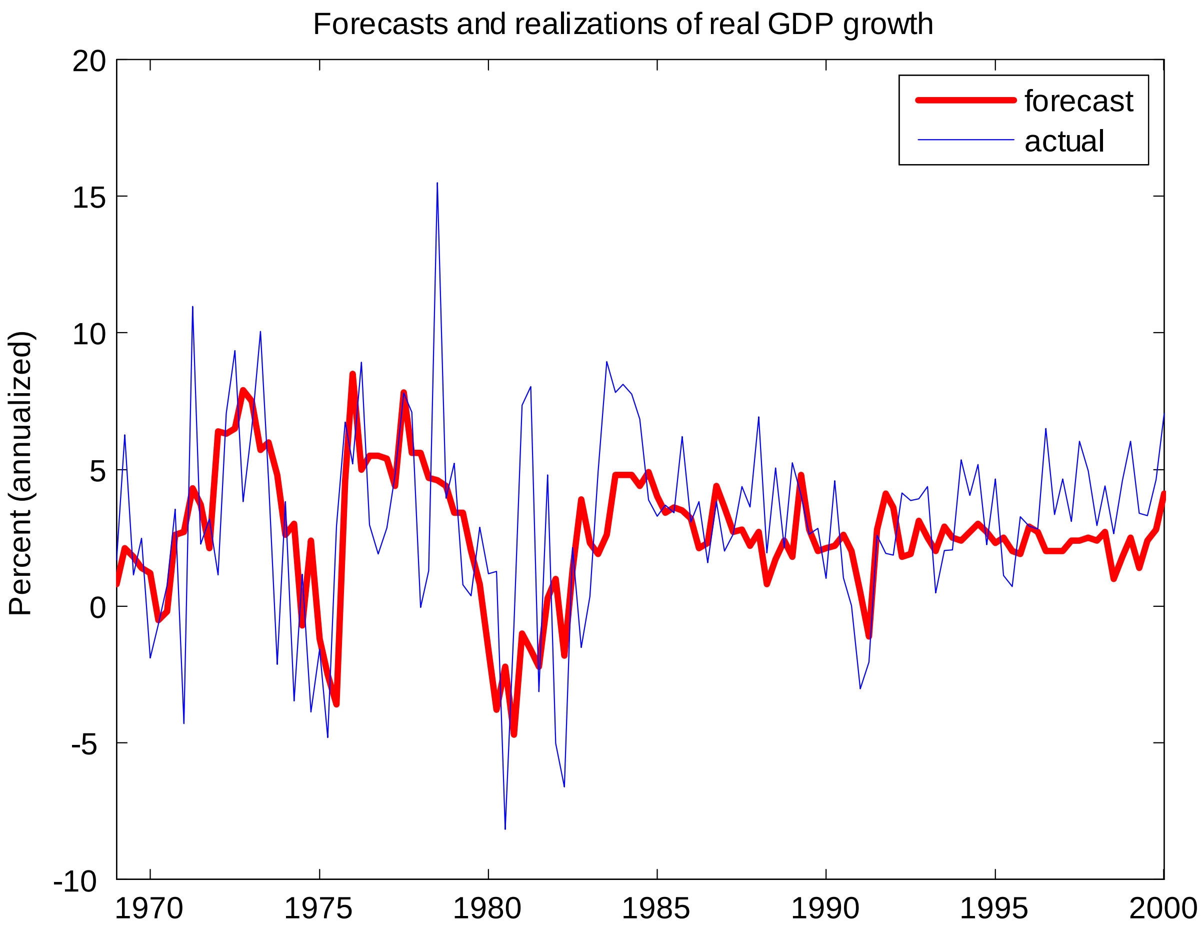


Figure 1: *Real GDP growth (annualized), and the Federal Reserve's "Greenbook" forecasts of real GDP growth, over the period 1968Q4 to 1999Q4.*

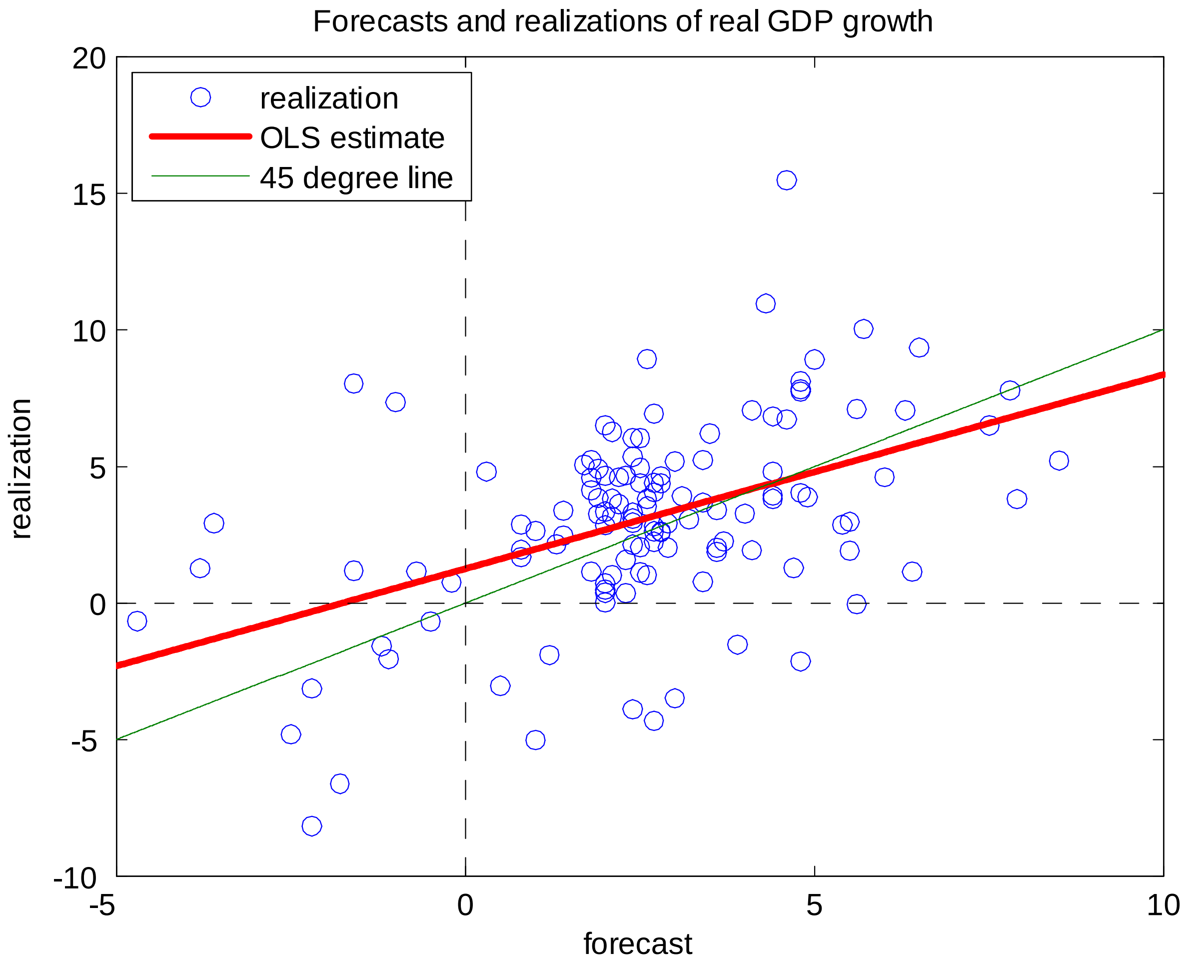


Figure 2: *Forecast and realized real GDP growth (annualized) over the period 1968Q4 to 1999Q4, with OLS fitted line.*

# 6 Testable implications under unknown loss

- In the vast majority of empirical situations, we have only a sequence of forecasts and realisations, with little idea of the loss function of the forecaster. Without knowledge of the forecaster's loss function testing forecast optimality may appear an impossible task, but this is not the case.

- We will now present results that provide testable implications of forecast optimality that may be applied when the forecaster's loss function is unknown, but testable restrictions may be placed on the DGP.

  - A rejection of optimality in this section constitutes a rejection of forecast optimality not merely under a single loss function, but under any member of an entire class of loss functions.

# 7 Conditional mean dynamics only

**Assumption D1:** The DGP satisfies: $Y_{t+h} = \mu_{t+h,t} + \varepsilon_{t+h}$, $\varepsilon_{t+h}|\mathcal{F}_t \sim F_{\varepsilon,h}\left(0, \sigma^2_{\varepsilon,h}\right)$.

- Though restrictive, notice that this class of DGPs includes all types of conditional mean specifications, linear or nonlinear. Also note that it makes no statement about stationarity (though this will need to be considered when implementing a test).

**Assumption L1:** The loss function is a function solely of the forecast error:

$$L(y, \hat{y}) = L(y - \hat{y}) = L(e)$$

- Assumption L1 is satisfied for many common loss functions: MSE, MAE, lin-ex, lin-lin, asymmetric quadratic.

# 8 Conditional mean dynamics only, cont'd

**Proposition 1:** Let the DGP and loss function satisfy D1 and L1. Then:

**(1)** The optimal forecast takes the form:

$$\hat{Y}^*_{t+h,t} = \mu_{t+h,t} + \alpha^*_h$$

where $\alpha^*_h$ is a constant that depends on $L$ and $F_{\varepsilon,h}$, but not on $\mathcal{F}_t$. (This is due to Granger 1969.)

**(2)** The optimal forecast error $e^*_{t+h,t}$ is independent of all $Z_t \in \mathcal{F}_t$. In particular,

$$Cov\left[e^*_{t+h,t}, e^*_{t+h-j,t-j}\right] = 0 \;\forall\; j \geq h$$

if the covariance exists.

# 9 Implications

- Without knowledge of the loss function we cannot (and should not) test that the forecast errors have mean zero.

- But under assumptions D1 and L1 we can still test for serial correlation in the forecast errors, and for correlation between the forecast errors and the forecasts, *without* needing to know $L$:

$$e_{t+h,t} = \alpha + \beta Z_t + u_{t+h}$$

$$\begin{aligned} H_0 &: \beta = 0 \\ H_a &: \beta \neq 0 \end{aligned}$$

# 10 Conditional mean and variance dynamics

**Assumption D2:** The DGP satisfies: $Y_{t+h} = \mu_{t+h,t} + \sigma_{t+h,t}\eta_{t+h}$, $\eta_{t+h}|\mathcal{F}_t \sim F_{\eta,h}(0,1)$.

- This class of DGPs is very broad and includes most common volatility proceses, eg ARCH, stochastic volatility, etc.

**Assumption L2:** The loss function is a homogeneous function solely of the forecast error:

$$\begin{aligned} L(y,\hat{y}) &= L(y-\hat{y}) = L(e) \\ \text{and } L(ae) &= g(a) \cdot L(e) \ \forall\ a \neq 0, \text{ for some positive function } g \end{aligned}$$

- L2 is satisfied by many common loss function families: MSE, MAE, lin-lin, asymmetric quadratic, but not lin-ex.

# 11 Conditional mean and variance dynamics

**Proposition 2:** Let the DGP and loss function satisfy D2 and L2, and define $d^*_{t+h,t} \equiv e^*_{t+h,t}/\sigma_{t+h,t}$. Then:

**(1)** The optimal forecast takes the following form:

$$\hat{Y}^*_{t+h,t} = \mu_{t+h,t} + \sigma_{t+h,t} \cdot \gamma^*_h$$

where $\gamma^*_h$ is a constant, depending only on $F_{\eta,h}$ and the loss function $L$.

**(2)** $d^*_{t+h,t}$ is independent of all $Z_t \in \mathcal{F}_t$. In particular,

$$Cov\left[d^{*r}_{t+h,t}, d^{*s}_{t+h-j,t-j}\right] = 0 \ \forall j \geq h$$

for all $(r, s)$ such that the covariance exists.

# 12 Conditional mean and variance dynamics

**Proof:** Let us represent a forecast as $\hat{Y}_{t+h,t} = \mu_{t+h,t} + \sigma_{t+h,t} \cdot \hat{\gamma}_{t+h,t}$.

$$\begin{aligned}
\hat{Y}^*_{t+h,t} &\equiv \arg\min_{\hat{y}} \int L\left(y-\hat{y}\right) dF_{t+h,t}\left(y\right) \\
&= \arg\min_{\hat{y}} \int g\left(\frac{1}{\sigma_{t+h,t}}\right) L\left(\frac{1}{\sigma_{t+h,t}}\left(y-\hat{y}\right)\right) dF_{t+h,t}\left(y\right) \\
&= \arg\min_{\hat{y}} \int L\left(\frac{1}{\sigma_{t+h,t}}\left(y-\hat{y}\right)\right) dF_{t+h,t}\left(y\right) \\
&= \mu_{t+h,t} + \sigma_{t+h,t} \cdot \arg\min_{\hat{\gamma}} \int L\left(\eta_{t+h} - \hat{\gamma}\right) dF_{\eta,h}\left(\eta\right) \\
&\equiv \mu_{t+h,t} + \sigma_{t+h,t} \cdot \gamma^*_h
\end{aligned}$$

$$\text{So } \; d^*_{t+h,t} \equiv \frac{Y_{t+h} - \hat{Y}^*_{t+h,t}}{\sigma_{t+h,t}} = \eta_{t+h} - \gamma^*_h$$

which is independent of all elements in $\mathcal{F}_t$ though not mean zero. ■

# 13 Conditional mean and variance dynamics

- We can test optimality by estimating the following for example:

$$
\begin{aligned}
d^*_{t+h,t} &= \beta_0 + \beta_1 d^*_{t,t-h} + u_{t+h} \\
u_{t+h} &= \sigma^2_{u,t+h} v_{t+h},\ v_{t+h} \sim (0,1) \\
\sigma^2_{u,t+h} &= \omega_0 + \omega_1 u^2_{t-1}, \text{ and then testing} \\
H_0 &: \omega_0 = 1 \cap \beta_1 = \omega_1 = 0
\end{aligned}
$$

- This test is easily computed, but requires that an estimate of $\sigma^2_{t+h,t}$ is available. An estimate of this conditional variance may be obtained from the observed $Y_t$ process either by means of a parametric model (eg, a GARCH-type model) or by nonparametric methods, using a realised volatility estimator.

- No estimate of $\mu_{t+h,t}$ is required. This is helpful particularly in finance.

# 14 Linked conditional mean and variance

- Often no reliable estimate of $\sigma_{t+h,t}$ is available, and so it is of interest to establish results that do not require such an estimate.

**Assumption D2':** The DGP satisfies: $Y_{t+h} = \beta\sigma_{t+h,t} + \sigma_{t+h,t}\eta_{t+h}$, where $\beta \in \mathbb{R}$, and $\eta_{t+h}|\mathcal{F}_t \sim F_{\eta,h}(0,1)$.

**Corollary 1:** Let the DGP and loss function satisfy assumptions D2' and L2, and assume that $\beta \neq -\gamma_h^*$. Define $\hat{d}^*_{t+h,t} \equiv \left(Y_{t+h} - \hat{Y}^*_{t+h,t}\right)/\hat{Y}^*_{t+h,t}$.

Then $\hat{d}^*_{t+h,t}$ is independent of any element $Z_t \in \mathcal{F}_t$.

# 15 A conditional quantile representation

- Under the conditions given for Propositions 1 or 2, we can show that the optimal forecast is a conditional quantile of the variable of interest.

**Proposition 3:** Let the DGP and loss function satisfy D1 and L1 **or** assumptions D2 and L2. Then:

**(1)** The optimal forecast is such that, for all $t$,

$$F_{t+h,t}\left(\hat{Y}^*_{t+h,t}\right) = q_h$$

where $q_h \in (0,1)$ depends only on the forecast horizon and the loss function. If $F_{t+h,t}$ is continuous and strictly increasing then we obtain:

$$\hat{Y}^*_{t+h,t} = F^{-1}_{t+h,t}(q_h)$$

**(2)** $I^*_{t+h,t} \equiv \mathbf{1}\left(Y_{t+h} \leq \hat{Y}^*_{t+h,t}\right)$ is independent of all $Z_t \in \mathcal{F}_t$.

# 16 A conditional quantile representation

**Proof:** (1) Under assumptions D1 and L1, or assumptions D2 and L2, we know from above that

$$Y^*_{t+h,t} = \mu_{t+h,t} + \sigma_{t+h,t} \cdot \gamma^*_h$$

with $\sigma_{t+h,t}$ constant under assumption D1. $\gamma^*_h$ depends only upon the loss function and the forecast horizon. Now notice that

$$\begin{aligned} F_{t+h,t}\left(\hat{Y}^*_{t+h,t}\right) &\equiv \text{Pr}\left[Y_{t+h} \leq \hat{Y}^*_{t+h,t} | \mathcal{F}_t\right] \\ &= \text{Pr}\left[\mu_{t+h,t} + \sigma_{t+h,t}\eta_{t+h} \leq \mu_{t+h,t} + \sigma_{t+h,t} \cdot \gamma_h | \mathcal{F}_t\right] \\ &= \text{Pr}\left[\eta_{t+h} \leq \gamma_h | \mathcal{F}_t\right] \\ &\equiv q^*_h \ \forall\ t \end{aligned}$$

Thus $\hat{Y}^*_{t+h,t}$ is the $q^*_h$ conditional quantile of $Y_{t+h}|\mathcal{F}_t \ \forall\ t$. Note that $q^*_h$ is only a function of the loss function and the forecast horizon.

(2) Since $I^*_{t+h,t}$ is a binary random variable and $\text{Pr}\left[I^*_{t+h,t} = 1|\mathcal{F}_t\right] = q^*_h \ \forall\ t$, we thus have that $I^*_{t+h,t}$ is independent of all $Z_t \in \mathcal{F}_t$. ■

# 17 A conditional quantile representation

- The conditional quantile representation of the optimal forecast enables us to test forecast optimality in this case *without* the need for an estimate of $\mu_{t+h,t}$ or $\sigma^2_{t+h,t}$.

- We can obtain a simple test of forecast optimality in two ways:

1. A regression of $I^*_{t+h,t}$ on elements of $\mathcal{F}_t$ :

$$\begin{aligned} I^*_{t+h,t} &= \alpha + \boldsymbol{\beta}'\mathbf{Z}_t + u_{t+h} \\ H_0 &: \boldsymbol{\beta} = \mathbf{0} \text{ vs. } H_a : \boldsymbol{\beta} \neq \mathbf{0} \end{aligned}$$

2. Via the test of Christoffersen (1998), who proposes modelling $I^*_{t+h,t}$ as a first-order Markov process, with transition matrix

$$\begin{bmatrix} 1-\pi_{01} & \pi_{01} \\ 1-\pi_{11} & \pi_{11} \end{bmatrix}$$

$$H_0 : \pi_{01} = \pi_{11} \left( = q^*_h \right) \text{ vs. } H_a : \pi_{01} \neq \pi_{11}$$

# 18 General DGPs and flexible parametric loss

- We now consider a test that may be implemented when the assumptions required for the above tests do not hold.

- We base this test on a flexible parametric estimate of the first derivative of the loss function, using the FOC: $0 = E_t \left[ \partial L \left( Y_{t+h}, \hat{Y}^*_{t+h,t} \right) / \partial \hat{y} \right]$.

$$\text{Let} \quad \lambda(y, \hat{y}) \equiv \frac{\partial L(y, \hat{y})}{\partial \hat{y}}$$

- We will approximate $\lambda$ using a linear spline.

- Say $\lambda = \lambda(e)$, let $(\zeta_1, ..., \zeta_K)$ be the nodes and impose that one of the nodes is zero. We impose that the spline is continuous, though not necessarily differentiable, except possibly at zero. We could allow further discontinuities in $\lambda$ at the cost of introducing more parameters to estimate.

## 19 A linear spline for $\partial L / \partial \hat{y}$

- With just a few nodes this class is quite flexible, and nests both MSE and MAE, as well as the "quad-quad", "lin-lin", and the symmetric, non-convex loss function of Granger (1969).

- If we impose that the spline is continuous at zero then MSE loss is nested without the boundary of the parameter space being hit. In this case the resulting estimated loss function is a quadratic spline, and is continuous and differentiable everywhere:

$$\frac{\partial \lambda\left(e;\boldsymbol{\theta}\right)}{\partial e} = \begin{cases} \gamma_1, & \text{for } e \leq \zeta_1 \\ \gamma_i, & \text{for } \zeta_{i-1} < e \leq \zeta_i,\ i = 2, ..., K \\ \gamma_{K+1}, & \text{for } e > \zeta_K \end{cases} .$$

- $\lambda\left(e;\boldsymbol{\theta}\right)$ and $L\left(e;\boldsymbol{\theta}\right)$ are constructed from the above specification by imposing that $\lambda\left(0;\boldsymbol{\theta}\right) = L\left(0;\boldsymbol{\theta}\right) = 0$ and that both $\lambda\left(e;\boldsymbol{\theta}\right)$ and $L\left(e;\boldsymbol{\theta}\right)$ are continuous in $e$.

# 20 A linear spline for $\partial L/\partial\hat{y}$, cont'd

- $\lambda(e;\boldsymbol{\theta})$ is only identified up to a multiplicative constant, so some normalisation is required to identify the parameters.

- Further, we must impose constraints on $\boldsymbol{\theta}$ so that the resulting estimate of $\lambda$ satisfies the assumptions required for it to be the first derivative of some valid loss function; eg: $\lambda(y,\hat{y}) \leq (\geq)\ 0$ for $y \geq (\leq)\ \hat{y}$.

- If L1 does not hold, then we must approximate $\lambda(y,\hat{y})$ rather than $\lambda(e)$. We choose to model $\lambda(e,y)$ rather than $\lambda(y,\hat{y})$ as it is simpler to impose the required conditions.

$$\frac{\partial\lambda(e,y;\boldsymbol{\theta})}{\partial e} = \begin{cases} \gamma_1 \equiv \Gamma(\varphi_{01} + \varphi_{11}y - \ln K), \\ \gamma_i \equiv \left(1 - \sum_{j=1}^{i-1}\gamma_j\right)\cdot\Gamma(\varphi_{0i} + \varphi_{1i}y - \ln K), \\ \gamma_{K+1} = 1 - \sum_{j=1}^{K}\gamma_j, \end{cases}$$

where $\Gamma(x) \equiv \left(1 + e^{-x}\right)^{-1}$ is the logistic transformation.

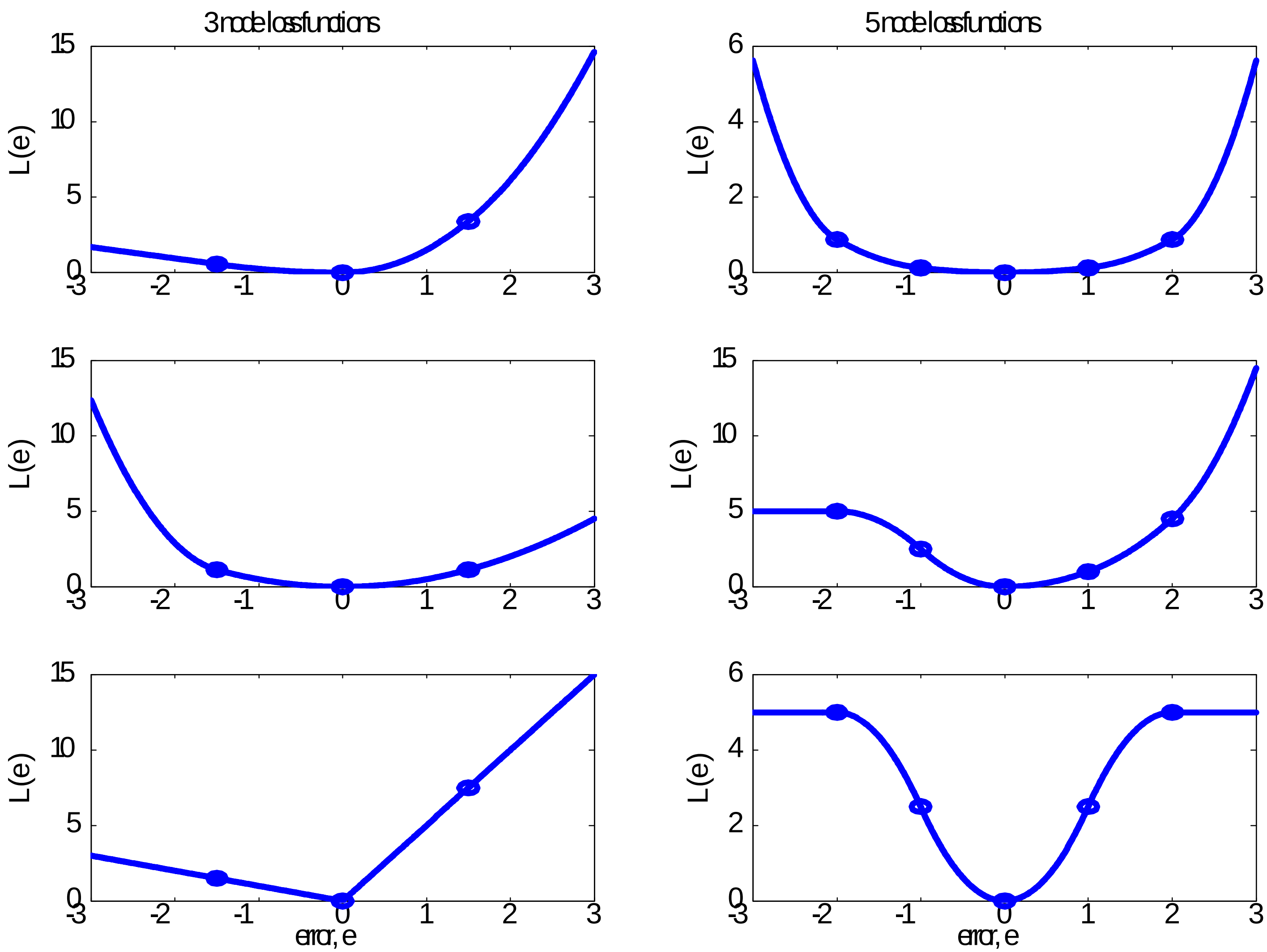


Figure 3: *Spline-based loss functions*

# 21 A test from over-identifying restrictions

- Under standard conditions the parameter vector of the approximating function can be estimated via GMM, in a similar fashion to Elliott, *et al.* (2002):

$$\hat{\boldsymbol{\theta}}_T \equiv \arg\min_{\boldsymbol{\theta}\in\Theta} \mathbf{g}_T(\boldsymbol{\theta})' \mathbf{W} \mathbf{g}_T(\boldsymbol{\theta})$$

$$\mathbf{g}_T(\boldsymbol{\theta}) \equiv \frac{1}{T}\sum_{t=1}^{T} \lambda\left(e_{t+h,t}, Y_{t+h}; \boldsymbol{\theta}\right) \cdot \mathbf{Z}_t,$$

where $\mathbf{W}$ is a weighting matrix, $\Theta$ is a compact set, and $\mathbf{Z}_t \in \mathcal{F}_t$.

- A test of forecast optimality can be obtained from a test of over-identifying restrictions if we ensure we have more moment restrictions, $k$, than parameters, $p$:

$$T\mathbf{g}_T\left(\hat{\boldsymbol{\theta}}_n\right)' \hat{\mathbf{W}}_T^* \mathbf{g}_T\left(\hat{\boldsymbol{\theta}}_T\right) \Rightarrow \chi^2_{k-p}, \text{ as } T \to \infty$$

where $\hat{\mathbf{W}}_T^*$ is a consistent estimate of the optimal weight matrix, c.f. Newey and McFadden (1994).

# 22 Empirical tests of Fed forecast optimality

- We study the Federal Reserve's "greenbook" quarterly forecasts of GDP growth over the period 1968Q4 to 1999Q4: 125 observations.

- Homoskedasticity (assumption D1) was rejected for this series, but assumption D2 was not rejected by some simple tests. Thus the quantile-based test from Proposition 3 was employed. Recall $I_t \equiv \mathbf{1}\left(Y_t \leq \hat{Y}_t\right)$.

$$I_t = \underset{(0.071)}{0.346} + \underset{(0.016)}{0.036}\,\hat{Y}_t + u_t$$

- The $t$-statistic on $\hat{Y}_t$ is 2.27, which is significant at the 0.05 level, indicating that forecast optimality under *any* loss function satisfying assumption L2 is rejected.

- This suggests that the assumption that $L(y, \hat{y}) = L(e)$ is not reasonable for the Fed's GDP forecasts.

# 23 Flexible estimation of the loss function

- Finally, we estimate the loss function using the linear spline model discussed previously.

- We used three nodes: $[-2, 0, 2]$, which correspond to the $[0.17, 0.44, 0.70]$ empirical quantiles of the forecast errors, and thus have 3 parameters to estimate.

- As instruments we used a constant, the forecast, and one lag each of the forecast error, the realisation and the generalised forecast error.

- We estimated the parameters by GMM, iterated using an estimate of the optimal weighting function.

- The test of over-identifying restrictions is $\chi_2^2$ under the null. Our test stat and p-value were $5.57$ and $0.06$, and so we marginally fail to reject optimality.

# 24 Flexible estimation of the loss function

- The above results suggest that the restriction that $L(y, \hat{y}) = L(e)$ is not reasonable for these forecasts. We next estimate the more general loss function to gain a better understanding of the forecasters' behaviour.

- We used the same three nodes, and thus have 6 parameters to estimate.

- As instruments we used a constant, the forecast, and two lags each of the forecast error, the realisation and the generalised forecast error.

- The test of over-identifying restrictions is $\chi^2_2$ under the null. Our test stat and p-value were 0.02 and 0.99, and so we fail to reject optimality.

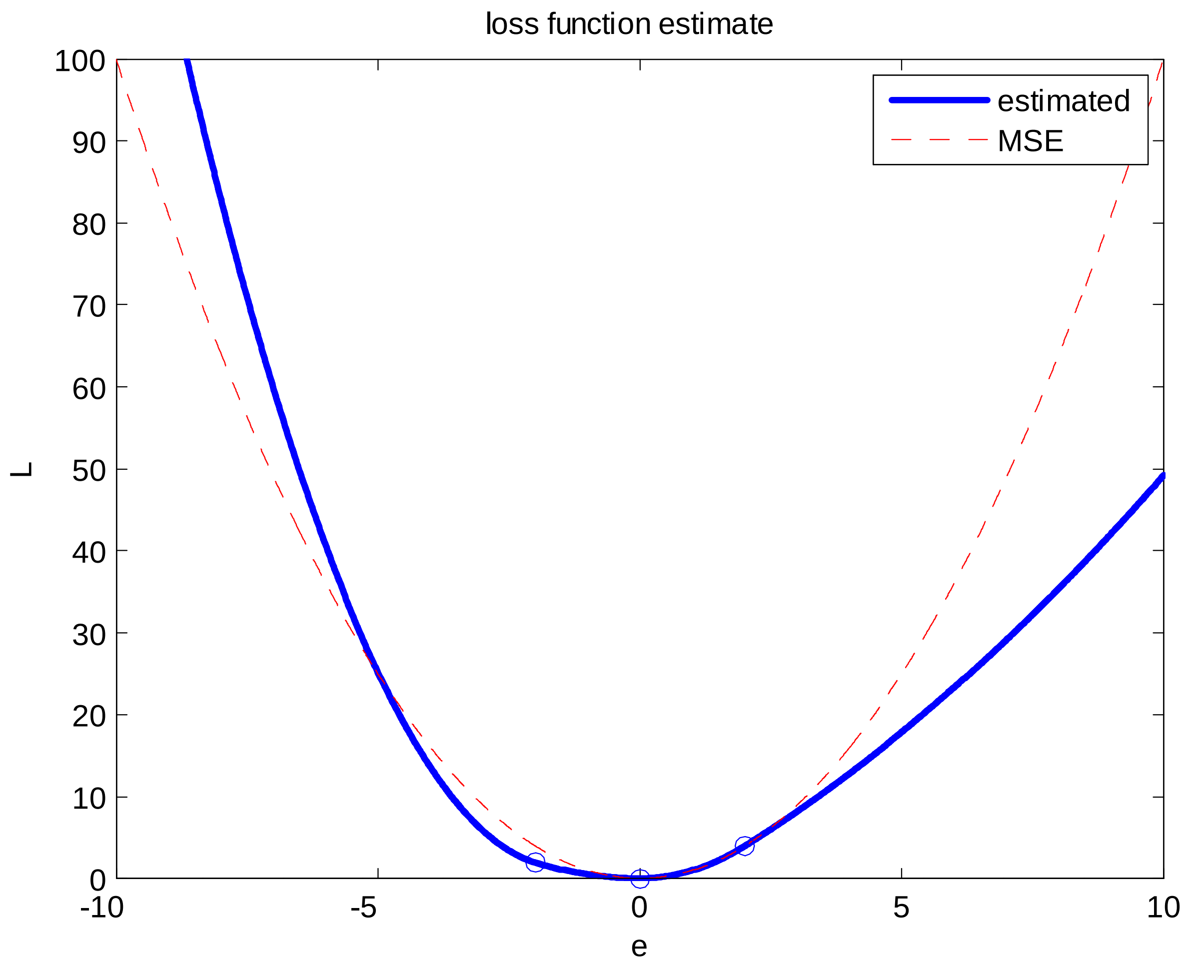


Figure 4: *The estimated loss function of the Federal Reserve for real GDP growth forecasts, based on a quadratic spline.*

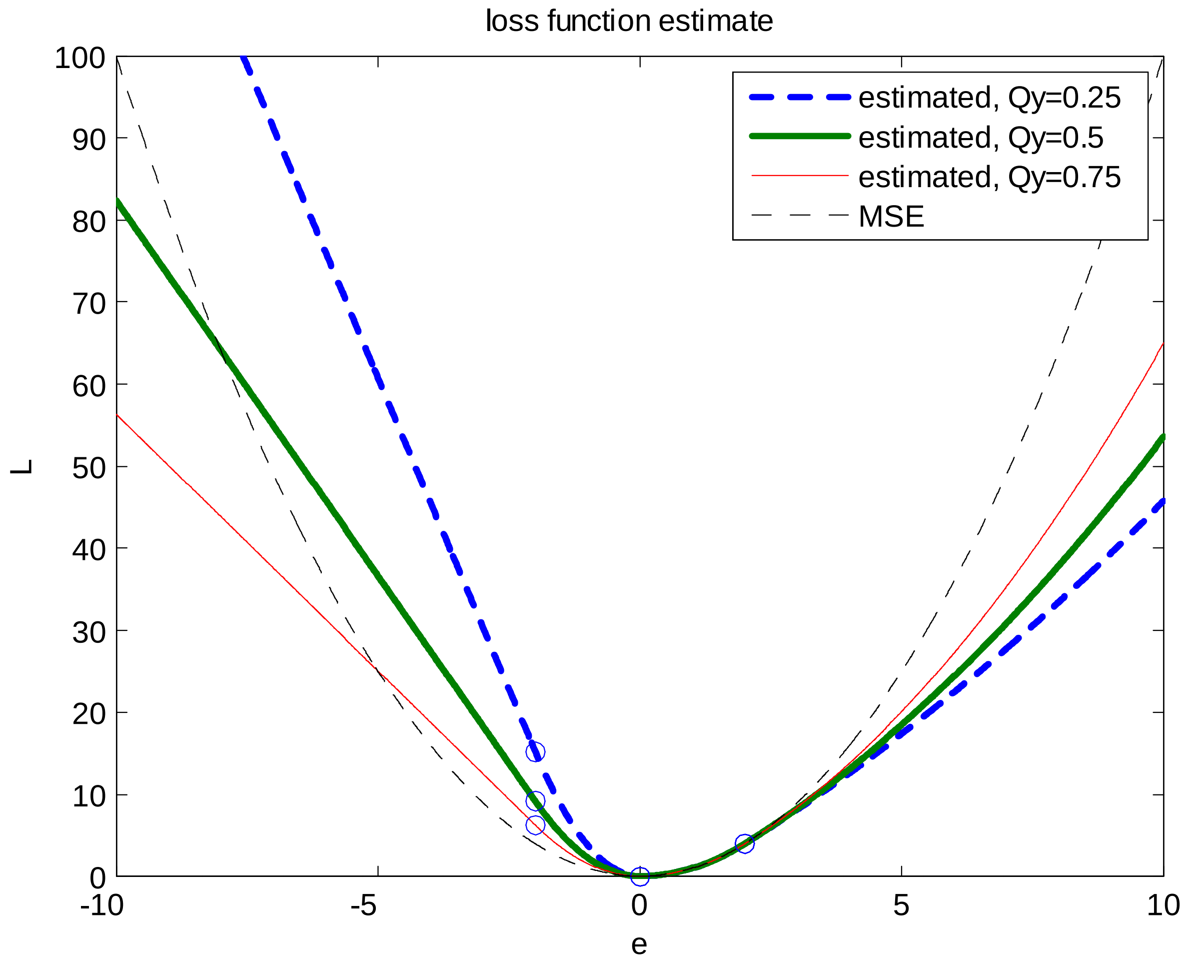


Figure 5: *The estimated loss function of the Federal Reserve for real GDP growth forecasts, based on quadratic splines. The estimated loss function is evaluated for GDP growth equal to its 0.25, 0.5 and 0.75 quantiles.*

# 25 Conclusion

- This paper provides new tests of forecast optimality that are applicable when the loss function is of the forecaster is unknown.
  - Our first set of tests trades off restrictions on the loss function with testable restrictions on the DGP.
  - Our second set of tests are based on flexible parametric approximations of the unknown loss function estimated via GMM. These tests extend Elliott, et al. (2004).

- We applied these tests to the Fed's "greenbook" forecasts of GDP growth and found optimality of the Fed forecasts only when we allow for a "conservative" loss function *and* when the degree of conservatism increases when growth is moderate or low.
  - The Fed's forecasts are not optimal under MSE, or any loss function that is homogeneous in the forecast error.