Zum statistischen Vorhersagefehler von statistischen Modellen

Jürgen Grieser

17.02.1998

Als statistisches Modell wird hier ein Modell verstanden, das eine beste Anpassung zwischen i Einflußzeitreihen $X_{i,t}$ und einer Zielzeitreihe Y_t macht. Dieses Modell kann dann dazu verwendet werden, um die Zielzeitreihe aufgrund vorhergehender oder weiterer Werte der Einflußzeitreihe zu rekonstruieren bzw. vorherzusagen. Je nachdem wie gut die Anpassung ist, wird man einen mehr oder weniger hohen statistischen Fehler machen. Diesen gilt es abzuschätzen. Der statistische Fehler muß dabei klar vom systematischen Fehler der Modellanpaßung unterschieden werden. Letzterer kann als vernachlässigbar angesehen werden, wenn das Residuum aus aufgrund der Anpassung erklärtem Anteil an der Zielzeitreihe $\hat{Y}_t(X_{i,t})$ und der Zielzeitreihe selbst als Rauschen aufgefaßt werden kann. Dann trägt nur noch dieses statistische Rauschen zum Fehler der Vorhersage oder Rekonstruktion bei. Hier soll nun angegeben werden, wie unter dieser Vorraussetzung der statistische Fehler der Vorhersage bzw. Rekonstruktion von der erklärten Varianz der Anpassung und der Standardabweichung der Rekonstruktion abhängt. Dazu wird zunächst die Varianz der Zielzeitreihe σ_y^2 in zwei Anteile aufgespalten:

$$\sigma_y^2 = \sigma_{\hat{y}}^2 + \sigma_{y-\hat{y}}^2 \tag{1}$$

wobei $\sigma_{\hat{y}}^2$ der Anteil der Varianz ist, der durch die Einflußzeitreihen erklärt wird, während $\sigma_{y-\hat{y}}^2$ der Anteil der Varianz von Y_t ist, der nicht durch die $X_{i,t}$ erklärt ist. Teilt man Gleichung (1) durch $\sigma_{\hat{y}}^2$ so erhält man folgenden Ausdruck

$$\frac{\sigma_y^2}{\sigma_{\hat{y}}^2} = 1 + \frac{\sigma_{y-\hat{y}}^2}{\sigma_{\hat{y}}^2}. \tag{2}$$

Nun ist der Ausdruck auf der linken Seite von Gleichung (2) gleich dem Reziproken des Bestimmtheitsmaßes r^2 (oft auch erklärte Varianz genannt, obwohl es keine Varianz ist, sondern ein Varianzanteil). Setzt man r^2 in Gleichung (2) ein, stellt nach dem rechten Term auf der rechten Seite um und zieht die Wurzel, so erhält man

$$\frac{\sigma_{y-\hat{y}}}{\sigma_{\hat{y}}} = \sqrt{\frac{1}{r^2} - 1}. \tag{3}$$

Gleichung (3) gibt demnach die Abhängigkeit des Verhältnisses der Standardabweichung des nicht erklärten Anteils zu der des erklärten Anteils als Funktion des Bestimmtheitsmaßes an. Damit ist man in der Lage aus der Standardabweichung der Vorhersage oder Rekonstruktion auf die Standardabweichung des zugehörigen statistischen Fehlers zu schließen. Wie sensibel diese Abhängigkeit insbesondere für nicht sehr große erklärte Varianzen ist, zeigt Abbildung 1. So ist z.B. bei einer erklärten Varianz von 80 % die Standardabweichung des statistischen Fehlers noch halb so groß wie die der Rekonstruktion. Aus der Kenntnis von $\sigma^2_{y-\hat{y}}$ können dann Vertrauensgrenzen für die Vorhersage- bzw. Rekonstruktionswerte angegeben werden. Unter der Bedingung, daß der statistische Fehler normalverteilt ist, folgt

90% der wahren Werte liegen weniger als $1.64\,\sigma_{y-\hat{y}}$ vom Modellwert entfernt,

95% der wahren Werte liegen weniger als $1.96\,\sigma_{y-\hat{y}}$ vom Modellwert entfernt und

99% der wahren Werte liegen weniger als $2.58\,\sigma_{y-\hat{y}}$ vom Modellwert entfernt.

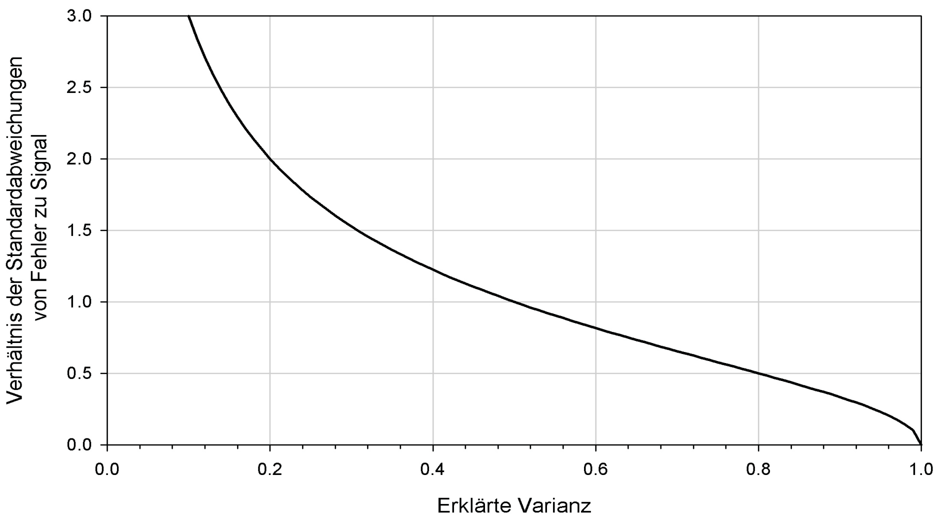

Abbildung 1: Verhältnis der Standardabweichung des statistischen Fehlers zur Standardabweichung der Rekonstruktion in Abhängigkeit von der erklärten Varianz der Anpassung (s. Glg. (3)).