

# Prognosen mit dynamischen Strukturgleichungsmodellen – ein Ausblick

M. Ruge\*

## Zusammenfassung

Die vorliegende Untersuchung gibt einen Ausblick auf Prognosemöglichkeiten mit dynamischen Strukturgleichungsmodellen. Die Analyse komplexer Systeme mit umfangreichen Datensätzen und die Erkennung relevanter Muster erfordern die Verwendung moderner statistischer Verfahren. DPLS-Modelle, eine Variante der Strukturgleichungsmodelle mit Latenten Variablen, werden methodisch erweitert, um mehrere zeitliche Verzögerungsstufen gleichzeitig modellieren zu können. Die Modelle versuchen, zahlreiche latente Einflussfaktoren und ihre Wechselwirkungen zu identifizieren. Als Daten werden rund 80 Indikatoren aus 20 Quellen verwendet, um Stimmungen, Erwartungen und wirtschaftlich relevanten Größen zu operationalisieren und zeitliche Prognosemöglichkeiten zu evaluieren. Für kürzere Zeiträume von sechs Monaten sind Stimmungen und Erwartungen die besten verfügbaren Prognosevariablen. Dieser Beitrag ist entstanden im Rahmen eines Vortrages im gemeinsamen Forschungsseminar mit der Staatlichen Universität für Wirtschaft und Finanzen Sankt Petersburg Finec im Dezember 2010.

## 1 Einleitung

Die zunehmende Vernetzung und Digitalisierung der modernen Welt führen zu wachsenden Datenmengen und erfordern neue statistische Verfahren, um die volatilen und unübersichtlichen Entwicklungen adäquat abbilden zu können. Die entwickelten statistischen Methoden bieten vielfältige Anwendungsmöglichkeiten im Bereich der Prognose und der Analyse komplexer Systeme – hier liegt der langfristige Nutzen des Forschungsgebietes, weit über den Horizont einer einzelnen Anwendung hinaus.

---

\*Universität Potsdam, Wirtschafts- und Sozialwissenschaftliche Fakultät, August-Bebel-Straße. 89, 14482 Potsdam, Deutschland. E-mail: marcus.ruge@uni-potsdam.de.

Die verwendeten Strukturgleichungsmodelle ermöglichen es, mehrere ähnliche beobachtbare Indikatoren zu Latenten Variablen zusammenzufassen und die Zusammenhänge zwischen diesen zu untersuchen. Latente Variable umfassen nicht direkt messbare oder komplexe Größen, die sinnvoll durch mehrere Indikatoren approximiert werden können. Der verwendete Partial-Least-Squares (PLS) Ansatz zählt zur zweiten Generation der komplexeren multivariaten Verfahren. Nach aktuellem Wissensstand bietet PLS bei kleineren Beobachtungsumfängen genauere Schätzungen, höhere Teststärken und bessere Prognosemöglichkeiten als verwandte Verfahren. Das dynamische PLS für Zeitreihenanalysen (DPLS) von Strohe (1997) wird methodisch erweitert, um Abhängigkeiten mit mehreren zeitlichen Verzögerungsstufen gleichzeitig modellieren zu können.

Die zusammengestellte Datenbank umfasst ca. 80 Indikatoren aus ca. 20 Quellen für den Zeitraum Januar 1991 bis Juni 2010, um Stimmungen, Erwartungen und wirtschaftlich relevante Größen im In- und Ausland zu operationalisieren. Zu den Herausgebern zählen Forschungsinstitute, wie das Ifo-Institut, das Zentrum für Europäische Wirtschaftsforschung und die Gesellschaft für Konsumforschung sowie öffentliche Institutionen, wie das Statistische Bundesamt, und internationale Zentralbanken.

## **2 Einführung in dynamische Strukturgleichungsmodelle**

Innerhalb der Welt der Strukturgleichungsmodelle existiert eine Vielzahl an Ansätzen, für diese Arbeit wird die DPLS-Pfadmodellierung gewählt. Der ursprüngliche PLS-Ansatz (Partial-Least-Squares) wurde vor allem vom norwegischen Statistiker und Wirtschaftswissenschaftler Herman Wold (1985) in den 1960–1980er Jahren entwickelt, basierend auf seinen vorherigen Arbeiten. Eine aktuelle, ausführliche Erläuterung zum Thema Strukturgleichungsmodelle und PLS findet sich in Vinzi et al. (2010). Das Prinzip des einfachen DPLS von Strohe (1997) kann erweitert werden, um mehrere zeitliche Verzögerungen gleichzeitig zu nutzen. Das Verfahren wird als erweitertes DPLS bezeichnet. Die verwendeten DPLS-Modelle erlauben die Nutzung zahlreicher Indikatoren, welche thematisch in Blöcke gruppiert werden, um sogenannte Latente Variable zu operationalisieren. Latente Variable umfassen nicht direkt messbare oder komplexe Größen, die sinnvoll durch mehrere Indikatoren approximiert werden können.

Die Struktur eines erweiterten DPLS-Modells kann in Matrixgleichungen zusammengefasst werden. Im dynamischen inneren Modell (1) wird für jede Zeitverzögerungsstufe  $\tau$  eine Koeffizientenmatrix  $\Gamma^{t-\tau}$  benötigt bis zur maximalen Verzögerungsstufe  $\tau_{max}$ . Die

Gewichts- (2) und Ladungsbeziehungen (3) im äußeren Modell bestimmen die Zusammenhänge der Indikatoren zu den zugehörigen Latenten Variablen und entsprechen weiterhin dem klassischen PLS-Modell.

$$\eta_t = B\eta_t + \Gamma\eta_{t-1} + \Gamma^2\eta_{t-2} + \dots + \Gamma^{\tau_{max}}\eta_{t-\tau_{max}} + v_t \quad (1)$$

$$\eta_t = W'x_t \quad (2)$$

$$x_t = P\eta_t + \varepsilon_t \quad (3)$$

Wobei  $\eta_t$  der Vektor der Latenten Variablen ist und  $v_t$  der zugehörige Störvektor mit Erwartungswerten von Null im inneren Modell. Der Index  $t$  bezeichnet die Zeit von  $t = 1$  bis  $T$ . Die Koeffizientenmatrix  $B$  (Beta) enthält die Pfadkoeffizienten des inneren Modells ohne Zeitverzögerungen. Die verzögerte Koeffizientenmatrix  $\Gamma^{\tau}$  oder vereinfacht  $\Gamma$  (Gamma) bei nur einer Verzögerungsstufe im einfachen DPLS enthält die Koeffizienten für die verzögerten Abhängigkeiten und wird wie  $B$  in unterer Dreiecksform gebildet. Die Diagonalelemente können hier aber von Null verschieden sein, um Autoregression zu ermöglichen. Die Struktur von  $B$  und  $\Gamma$  wird entsprechend der angenommenen Abhängigkeitsstruktur zwischen den Latenten Variablen in Analogie zum klassischen PLS geregelt. Der Vektor  $x_t$  enthält die Indikatoren.  $W$  ist die Gewichtungsmatrix und  $P$  (Rho) die Ladungsmatrix, beide definieren die Zusammenhänge der Indikatoren zu den zugehörigen Latenten Variablen.

In der Modellstruktur wird zwischen dem äußeren und inneren Modell unterschieden. Die Zuordnung der Gruppen von Indikatoren durch die jeweiligen Latenten Variablen wird als das äußere Modell bzw. Messmodell bezeichnet, da hier versucht wird, die unbeobachtbaren Latenten Variablen durch ihre Indikatoren zu messen. Das innere Modell bzw. Strukturmodell beschreibt dagegen die unbeobachtbaren Beziehungen zwischen den Latenten Variablen. (D)PLS-Modelle kombinieren somit Elemente der Hauptkomponenten- bzw. Faktoranalyse und der multiplen linearen Regression. Das Pfadsystem des inneren Modells beschreibt die linearen Beziehungen der Latenten Variablen als Gleichungssystem.

Die Parameterschätzungen im (D)PLS-Verfahren basieren auf einem partiellen, iterativen Algorithmus. In der äußeren Approximationsstufe werden wiederholt lineare Regressio-

nen durchgeführt, in der inneren Approximationsstufe erfolgt ein Austausch der Werte der Latenten Variablen nach Gewichtungsschemen. Die Latenten Variablen sollen somit sowohl das innere als auch das äußere Modell sinnvoll abbilden. Nach Konvergenz werden in einer zweiten Stufe die Pfadkoeffizienten und Bestimmtheitsmaße im inneren Modell mit OLS geschätzt.

Die Konvergenz des Algorithmus ist praktisch immer gegeben – mit sehr seltenen Ausnahmen. Simulationsstudien von Reinartz et al. (2009) haben ergeben, dass die Genauigkeit der PLS-Parameterschätzungen für Stichproben unter 250–500 Beobachtungen höher ist als bei vergleichbaren kovarianzorientierten Verfahren wie LISREL.

Abbildung 1 verdeutlicht die Zusammenhänge in einem beispielhaften erweiterten DPLS-Modell. Sobald eine Latente Variable verzögert auftritt, wird der entsprechende Pfeil im Pfaddiagramm gestrichelt dargestellt. Die jeweiligen Koeffizienten der verzögerten Beziehung erhalten eine zusätzliche Kennzeichnung  $t - \tau$  mit  $\tau$  als Stufe der zeitlichen Verzögerung.

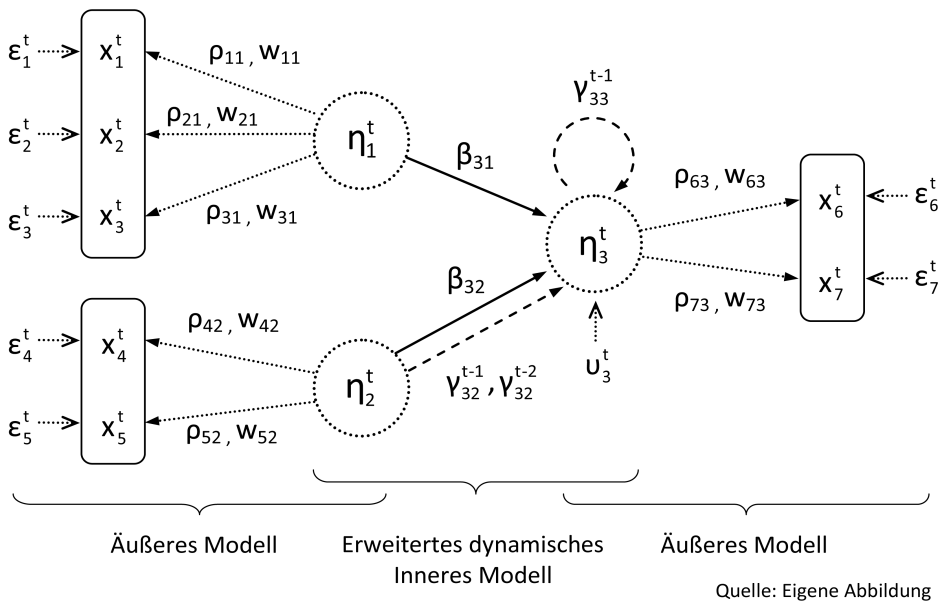


Abb. 1: Erweitertes DPLS-Pfadmodell

Die dynamische Erweiterung des DPLS-Schätzverfahrens ermöglicht die Verwendung verzögerter Latenter Variablen mit mehreren Verzögerungsstufen gleichzeitig. Dazu wird die innere Approximationsstufe erweitert für alle verzögerten Latenten Umgebungsvariablen

in Analogie zum einfachen DPLS. Die Äußere Approximation bleibt unverändert. Eine ausführlichere Darstellung zum erweiterten DPLS ist in Ruge (2011) zu finden.

Die dynamischen Strukturgleichungsmodelle wie DPLS bieten gegenüber der weit verbreiteten multivariaten Regression den Vorteil, eine wesentliche höhere Anzahl an Variablen gleichzeitig nutzen zu können. Gegenüber der dynamischen Faktoranalyse besteht der Vorteil, dass besser für die Prognose geeignete Indikatoren automatisch höhere Gewichte und Ladungen erhalten und sich somit die Prognosekraft erhöht. Gegenüber den Vektorautoregressiven Modellen (VAR) besteht der Vorteil, ein einheitliches Modell zu schätzen, ohne die Zerlegung in zahlreiche Teilmodelle.

### 3 Prognosemodelle

Die Prognosekraft der Indikatoren und Latenten Variablen auf zukünftige wirtschaftliche Entwicklungen soll ausführlicher untersucht werden. Die Wirtschaftssubjekte werden von den Forschungsinstituten nach ihrer Einschätzung der zukünftigen wirtschaftlichen Entwicklung befragt, allerdings ist dabei unklar, wie die wirtschaftliche Entwicklung zu definieren ist. Es handelt sich um eine Latente Variable, die durch ausgewählte Indikatoren approximiert werden kann. Es ist ebenfalls möglich, lediglich einen Indikator zu verwenden, wie das Bruttoinlandsprodukt oder einen ausgewählten Aktienindex. Allerdings beziehen sich die Umfragen nicht auf eine einzelne konkrete Größe, sondern auf das abstrakte Konstrukt einer wirtschaftlichen Situation. Die Modelle werden mit Wachstumsraten geschätzt, um die Probleme gemeinsamer Trends und Scheinkorrelation zu verhindern.

Zur Modellprognose werden elf erklärende Latente Variable genutzt. Die zu prognostizierende Latente Variable *Wirtschaftliche Entwicklung*<sub>*t*</sub> zum Zeitpunkt *t* wird mit zehn reflektiven Indikatoren spezifiziert, u.a. *Ausrüstungsinvestitionen*, *DAX*, *Erwerbslosenquote*, um möglichst repräsentativ die allgemeine wirtschaftliche Entwicklung abzubilden. Die Ladungen der wirtschaftlichen Entwicklung<sub>*t*</sub> liegen meist im hohen Bereich über 0,8, so dass die Latente Variable relativ eindeutig spezifiziert ist. Eine Ausnahme stellt der Indikator *Erwerbslosenquote gesamt* dar, der tendenziell antizyklisch zur wirtschaftlichen Entwicklung verläuft und eine niedrige Ladung erhält. Die hohe Faktorreliabilität von 0,93 bestätigt die Aussagen.

Auf die vollständige graphische Darstellung des Pfadmodells soll aus Platzgründen verzichtet werden. Das innere Modell wird dynamisch spezifiziert als Eingleichung mit der

erklärten Variable  $\eta_1$  *Wirtschaftliche Entwicklung*<sub>t</sub> und  $\eta_2$  bis  $\eta_{12}$  als verzögerte erklärende Latente Variablen. Die verzögerten Pfadkoeffizienten  $y_i$  quantifizieren die Zusammenhänge. Die zu schätzende Regressionsfunktion des inneren Modells lautet:

$$\begin{aligned}
 \text{Wirtschaftliche Entwicklung}_t = & y_1 \text{Erwartungen}_{t-\tau} + y_2 \text{Stimmungen}_{t-\tau} \\
 & + y_3 \text{Investitionen}_{t-\tau} + y_4 \text{Nachfrage}_{t-\tau} \\
 & + y_5 \text{Auftragseingänge}_{t-\tau} + y_6 \text{Aktien}_{t-\tau} \\
 & + y_7 \text{Anleihen}_{t-\tau} + y_8 \text{Zinsen}_{t-\tau} \\
 & + y_9 \text{Inflation}_{t-\tau} + y_{10} \text{Beschäftigung}_{t-\tau} \\
 & + y_{11} \text{Arbeitskosten}_{t-\tau} + v_t
 \end{aligned} \quad (4)$$

Auf die Zuordnung aller 80 Indikatoren auf die insgesamt 12 Latenten Variablen soll in diesem Überblick nicht detailliert eingegangen werden. Beispielhaft sei genannt, dass die Latente Variable Aktien aus insgesamt 14 deutschen und internationalen Aktienindizes konstruiert wird. Da die einzelnen Aktienindizes hoch korreliert sind, können diese gut zu einer Latenten Variablen zusammengefasst werden.

Zum Vergleich der Prognosekraft im zeitlichen Verlauf wird der Zeithorizont verschoben von  $\tau = 1$  bis 24 Monaten. Abbildung 2 fasst die Ergebnisse zusammen, auf der Abszisse sind alle Verzögerungsstufen abgetragen, auf der Ordinate die Bestimmtheitsmaße der jeweiligen inneren Einzelmodelle.

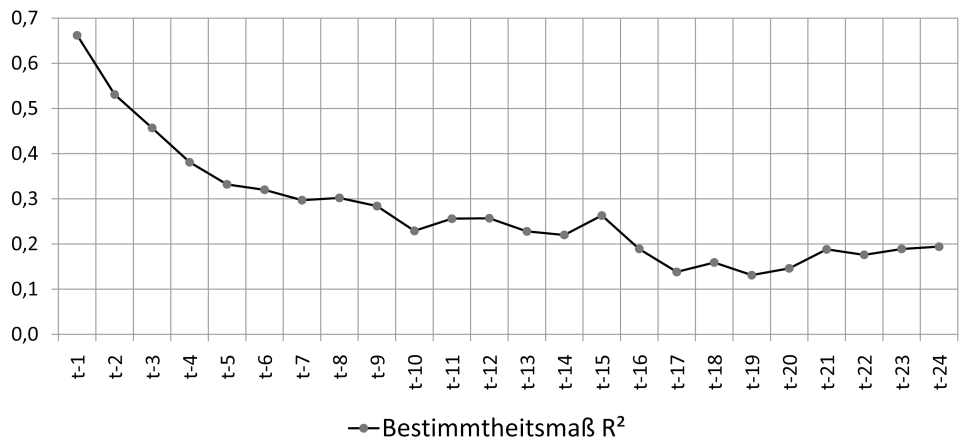


Abb. 2: Prognosekraft für wirtschaftliche Entwicklung

Die Modelle zeigen bei einem Monat Verzögerung die bestmögliche Prognose mit dem höchsten Bestimmtheitsmaß von 0,66. Die Prognosekraft fällt mit zunehmendem Zeithorizont deutlich ab und erreicht ab 16 Monaten einen schwachen Bereich von  $R^2 = 0,15$  bis 0,2. Zukunftsprognosen sind somit nur für kurze Zeiträume sinnvoll möglich. Für 36 bzw. 48 Monate ergeben sich Bestimmtheitsmaße von 0,28 bzw. 0,21. Somit verbleibt eine Restprognosekraft für langfristige Prognosen, allerdings ist diese in Anbetracht der hohen Anzahl an Latenten Variablen und Indikatoren als schwach zu beurteilen. Die Ergebnisse bestätigen grundsätzlich bisherige Forschungsarbeiten im Bereich der Konjunkturforschung und zeigen die Grenzen der Prognosenmöglichkeiten auf. Es scheint unwahrscheinlich, dass die Prognose durch weitere Variable oder veränderte Methoden wesentlich verbessert werden kann.

Für kürzere Zeiträume von sechs Monaten sind Stimmungen und Erwartungen die besten Latenten Prognosevariablen. Für längere Prognosen über zwölf Monate verlieren Stimmungen und Erwartungen deutlich an Einfluss.

## **4 Fazit**

Untersucht wurden Prognosemöglichkeiten mit dynamischen Strukturgleichungsmodellen. Diese Modelle ermöglichen es, eine Vielzahl an Variablen zu nutzen, um komplexe Systeme mit umfangreichen Datensätzen zu analysieren. Als Daten werden rund 80 Indikatoren aus 20 Quellen verwendet. Strukturgleichungsmodelle bieten einen wertvollen Beitrag zur Verbesserung der Prognosemöglichkeiten gegenüber einfachen Regressionen oder dynamischen Faktoranalysen. Die Modellergebnisse zeigen, dass für kürzere Zeiträume von sechs Monaten Stimmungen und Erwartungen der befragten Wirtschaftssubjekte die besten verfügbaren Prognosevariablen sind. Insgesamt nimmt die Prognosekraft mit zunehmendem Zeithorizont deutlich ab, die Zukunft verbleibt letztendlich überwiegend ungewiss. Eine umfassendere Darstellung der Thematik findet sich Ruge (2011).

## **Literatur**

- Reinartz, W., Haenlein, M. und Henseler, J. (2009). 'An Empirical Comparison of the Efficacy of Covariance-Based and Variance-Based SEM.' Working Paper 44, INSEAD.
- Ruge, M. (2011). 'Stimmungen und Erwartungen im System der Märkte – eine Analyse mit DPLS-Modellen.'

- Strohe, H. (1997). 'Partial Least Squares Approach to Estimating Dynamic Path Models.' In M.R.S.M. Balderjahn I. (Hg.), *Classification, Data Analysis and Data Highways*, S. 192–202. Berlin, Heidelberg.
- Vinzi, V., Chin, W., J., H. et al. (2010). *Handbook of partial least squares: Concepts, methods, and applications*. Heidelberg.
- Wold, H. (1985). 'Partial least squares.' In S. Kotz (Hg.), *Encyclopedia of statistical sciences*, Bd. 6, S. 581–591. New York.