

DSGE-Modelle

Verfahren vollständiger Information

Kalman-Filter, Maximum Likelihood und Bayesianische Methoden

Dr. Andrea Beccarini

Willi Mutschler, M.Sc.

Institut für Ökonometrie und Wirtschaftsstatistik Münster
willi.mutschler@uni-muenster.de

Sommersemester 2012

Verfahren vollständiger Information

- 1 Idee
- 2 Kalman-Filter
 - Notation
 - Initialisierung
 - Rekursion
 - Zusammenfassung und Likelihood
- 3 Maximum-Likelihood
 - Idee
 - Aufgabe 4: An und Schorfheide (2007) mit Maximum-Likelihood
- 4 Bayesianische Schätzung
 - Idee
 - Metropolis-Hastings-Algorithmus
 - Hinweise
 - Aufgabe 5: Schätzung mit Bayesianischen Methoden
 - Inferenz, Prognose, Modellvergleich und Identifikation
- 5 Diskussion

Verfahren vollständiger Information

Idee

- Verfahren vollständiger Information bauen auf einer kompletten Charakterisierung des Daten generierenden Prozesses auf (keine Beschränkung auf bestimmte Momente).
- Betrachte die lineare *state-space* Repräsentation des Modells:

$$\mathbf{x}_t = \mathbf{F}(\mu)\mathbf{x}_{t-1} + \mathbf{G}(\mu)\mathbf{v}_t, \quad \text{mit } E[\mathbf{v}_t\mathbf{v}_t'] = \mathbf{V}, \quad E[\mathbf{v}_t\mathbf{v}_s'] = 0 \quad (1)$$

$$\mathbf{X}_t = \mathbf{H}(\mu)'\mathbf{x}_t + \mathbf{e}_t, \quad \text{mit } E[\mathbf{e}_t\mathbf{e}_t'] = \mathbf{\Sigma}_e, \quad E[\mathbf{e}_t\mathbf{e}_s'] = 0 \quad (2)$$

- Matrix \mathbf{H} verknüpft die Variablen des Modells \mathbf{x}_t mit den beobachtbaren Daten \mathbf{X}_t .
- Gleichung (1): *state- oder transition-equation*
 - Entspricht der Lösung des Modells.
 - \mathbf{v}_t sind die stochastischen Innovationen.
- Gleichung (2): *observation-equation*
 - Entspricht den Messgleichungen,
 - unter Berücksichtigung eventueller Messfehler \mathbf{e}_t in den Daten.

Verfahren vollständiger Information

Idee

- Trifft man nun Annahmen über die Verteilung von \mathbf{v}_t und \mathbf{e}_t , so lässt sich die Log-Likelihood-Funktion, $\log L(\mathbf{X}|\boldsymbol{\mu})$, analytisch bzw. numerisch herleiten.
- Im linearen Fall und unter der Annahme normalverteilter Variablen, kann die Likelihood mithilfe des Kalman-Filters berechnet werden.
- Im nichtlinearen Fall betrachtet man $\mathbf{s}_t = s(\mathbf{s}_{t-1}, \mathbf{v}_t)$, $\mathbf{c}_t = c(\mathbf{s}_t)$ und $\mathbf{X}_t = \tilde{h}(\mathbf{s}_t, \mathbf{c}_t, \mathbf{v}_t, \mathbf{e}_t) \equiv h(\mathbf{s}_t, \mathbf{e}_t)$, wobei c, s und h Funktionen des Parametervektors $\boldsymbol{\mu}$ sind. Für die Herleitung der Likelihood wird dann der Partikelfilter oder das *efficient importance sampling* verwendet.
- Bei der Analyse und Auswertung der Log-Likelihood lassen sich zwei Herangehensweisen unterscheiden:
 - ① **die klassische, frequentistische Maximum-Likelihood-Methode,**
 - ② **die bayesianische Methode.**

Verfahren vollständiger Information

- 1 Idee
- 2 Kalman-Filter
 - Notation
 - Initialisierung
 - Rekursion
 - Zusammenfassung und Likelihood
- 3 Maximum-Likelihood
 - Idee
 - Aufgabe 4: An und Schorfheide (2007) mit Maximum-Likelihood
- 4 Bayesianische Schätzung
 - Idee
 - Metropolis-Hastings-Algorithmus
 - Hinweise
 - Aufgabe 5: Schätzung mit Bayesianischen Methoden
 - Inferenz, Prognose, Modellvergleich und Identifikation
- 5 Diskussion

Kalman-Filter

Notation

Wir beschränken uns auf den linearen Fall und ignorieren, vereinfachend, eventuelle Messfehler in den Daten:

- $\mathbf{X}_t = \mathbf{H}'\mathbf{x}_t$
- $\mathbf{x}_{t+1} = \mathbf{F}\mathbf{x}_t + \mathbf{G}\mathbf{v}_{t+1}$
- $\mathbf{v}_i \stackrel{iid}{\sim} \mathcal{N}(\mathbf{0}, \mathbf{V}), \mathbf{V} = E(\mathbf{v}_i\mathbf{v}_i'), E(\mathbf{v}_i\mathbf{v}_j') = 0$

Notation für die lineare Projektion

$$\hat{\mathbf{x}}_{t|t-j} = E(\mathbf{x}_t | \mathbf{X}_{t-j}, \mathbf{X}_{t-j-1}, \dots, \mathbf{X}_1)$$

$$\boldsymbol{\Sigma}_{t|t-j} = E(\mathbf{x}_t - \hat{\mathbf{x}}_{t|t-j})(\mathbf{x}_t - \hat{\mathbf{x}}_{t|t-j})'$$

$$\hat{\mathbf{X}}_{t|t-j} = E(\mathbf{X}_t | \mathbf{X}_{t-j}, \mathbf{X}_{t-j-1}, \dots, \mathbf{X}_1)$$

$$\mathbf{u}_t = \mathbf{x}_t - \hat{\mathbf{x}}_{t|t-1} = \mathbf{H}'(\mathbf{x}_t - \hat{\mathbf{x}}_{t|t-1})$$

$$E(\mathbf{u}_t\mathbf{u}_t') = \mathbf{H}'\boldsymbol{\Sigma}_{t|t-1}\mathbf{H}$$

für $t = 1, 2, \dots, T$ und $j = 0, 1, \dots, T$.

Kalman-Filter

Initialisierung

- Da \mathbf{x}_t schwach-stationär ist, gilt für die Varianz:

$$\begin{aligned}\underbrace{E(\mathbf{x}_t \mathbf{x}_t')}_{\equiv \Sigma} &= E[(\mathbf{F} \mathbf{x}_{t-1} + \mathbf{G} \mathbf{v}_t)(\mathbf{F} \mathbf{x}_{t-1} + \mathbf{G} \mathbf{v}_t)'] \\ &= \mathbf{F} \underbrace{E(\mathbf{x}_{t-1} \mathbf{x}_{t-1}')}_{\equiv \Sigma} \mathbf{F}' + \mathbf{G} \underbrace{E(\mathbf{v}_t \mathbf{v}_t')}_{=\mathbf{V}} \mathbf{G}' \\ \Leftrightarrow \Sigma &= \mathbf{F} \Sigma \mathbf{F}' + \mathbf{G} \mathbf{V} \mathbf{G}'\end{aligned}$$

Vektorisierung

vec -Operator stapelt die Spalten einer $m \times n$ Matrix \mathbf{M} in einem $mn \times 1$ Vektor $\text{vec}(\mathbf{M})$. Dann gilt für beliebige Matrizen $\underset{m \times n}{A}$, $\underset{n \times p}{B}$ und $\underset{p \times k}{C}$:

$$\text{vec}(ABC) = (C' \otimes A) \text{vec}(B), \quad \text{wobei } \otimes : \text{Kroneckerprodukt.}$$

- Angewandt ergibt sich somit folgende Lösung für Σ :

$$\begin{aligned} \text{vec}(\Sigma) &= (\mathbf{F} \otimes \mathbf{F}) \text{vec}(\Sigma) + \text{vec}(\mathbf{G}\mathbf{V}\mathbf{G}') \\ \Leftrightarrow \text{vec}(\Sigma) &= (\mathbf{I} - \mathbf{F} \otimes \mathbf{F})^{-1} \text{vec}(\mathbf{G}\mathbf{V}\mathbf{G}') \end{aligned}$$

- Die Initialisierung des Kalman-Filters erfolgt mit dem unbedingten Erwartungswert von \mathbf{x}_1 , da noch keine Beobachtungen zur Verfügung stehen:

$$\begin{aligned} \hat{\mathbf{x}}_1 &= \underbrace{E(\mathbf{x}_1)}_{=E(\mathbf{x})} = \mathbf{F} \underbrace{E(\mathbf{x}_0)}_{=E(\mathbf{x})} + \mathbf{G} \underbrace{E(\mathbf{v}_1)}_{=0} \Leftrightarrow \hat{\mathbf{x}}_1 = \mathbf{0}, \\ \text{vec}(\Sigma_{1|0}) &= E(\mathbf{x}_1 - \mathbf{0})(\mathbf{x}_1 - \mathbf{0})' = (\mathbf{I} - \mathbf{F} \otimes \mathbf{F})^{-1} \text{vec}(\mathbf{G}\mathbf{V}\mathbf{G}'). \end{aligned}$$

Die weiteren Rekursionsschritte orientieren sich an:

$$\hat{\mathbf{x}}_{t+1|t} = \mathbf{F}\hat{\mathbf{x}}_{t|t}$$

Formel für die Aktualisierung einer linearen Projektion (Hamilton (1994, S.99 und S.379))

$$\begin{aligned}\hat{\mathbf{x}}_{t|t} &= \hat{\mathbf{x}}_{t|t-1} + \left[E(\mathbf{x}_t - \hat{\mathbf{x}}_{t|t-1})(\mathbf{X}_t - \hat{\mathbf{X}}_{t|t-1})' \right] \left[E(\mathbf{X}_t - \hat{\mathbf{X}}_{t|t-1})(\mathbf{X}_t - \hat{\mathbf{X}}_{t|t-1})' \right]^{-1} \mathbf{u}_t \\ \Leftrightarrow \hat{\mathbf{x}}_{t|t} &= \hat{\mathbf{x}}_{t|t-1} + \boldsymbol{\Sigma}_{t|t-1} \mathbf{H} (\mathbf{H}' \boldsymbol{\Sigma}_{t|t-1} \mathbf{H})^{-1} \mathbf{u}_t \\ \Rightarrow \hat{\mathbf{x}}_{t+1|t} &= \mathbf{F}\hat{\mathbf{x}}_{t|t} = \mathbf{F}\hat{\mathbf{x}}_{t|t-1} + \mathbf{F}\boldsymbol{\Sigma}_{t|t-1} \mathbf{H} (\mathbf{H}' \boldsymbol{\Sigma}_{t|t-1} \mathbf{H})^{-1} \mathbf{u}_t, \\ &\quad \text{mit } \mathbf{u}_t = \mathbf{X}_t - \hat{\mathbf{X}}_{t|t-1} = \mathbf{H}'(\mathbf{x}_t - \hat{\mathbf{x}}_{t|t-1}).\end{aligned}$$

Kalman-Filter

Rekursion

- $\mathbf{x}_{t+1} - \hat{\mathbf{x}}_{t+1|t} = \mathbf{F} (\mathbf{x}_t - \hat{\mathbf{x}}_{t|t-1}) + \mathbf{G}v_{t+1} - \mathbf{F}\Sigma_{t|t-1}\mathbf{H} (\mathbf{H}'\Sigma_{t|t-1}\mathbf{H})^{-1} \mathbf{u}_t$
- Der *MSE*: $\Sigma_{t+1|t} = E (\mathbf{x}_{t+1} - \hat{\mathbf{x}}_{t+1|t}) (\mathbf{x}_{t+1} - \hat{\mathbf{x}}_{t+1|t})'$ ist dann gegeben durch:

$$\begin{aligned} \Sigma_{t+1|t} = & \mathbf{F}\Sigma_{t|t-1}\mathbf{F}' + \mathbf{G}\mathbf{V}\mathbf{G}' - \underbrace{\mathbf{F}\Sigma_{t|t-1}\mathbf{H} (\mathbf{H}'\Sigma_{t|t-1}\mathbf{H})^{-1} \underbrace{E(\mathbf{u}_t\mathbf{u}_t')}_{=\mathbf{H}'\Sigma_{t|t-1}\mathbf{H}} \mathbf{H}'\Sigma_{t|t-1}\mathbf{F}'}_{=I} \end{aligned}$$

Mean-Squared-Error (MSE)

$$\Sigma_{t+1} \equiv \Sigma_{t+1|t} = \mathbf{F}\Sigma_{t|t-1}\mathbf{F}' + \mathbf{G}\mathbf{V}\mathbf{G}' - \mathbf{F}\Sigma_{t|t-1}\mathbf{H} (\mathbf{H}'\Sigma_{t|t-1}\mathbf{H})^{-1} \mathbf{H}'\Sigma_{t|t-1}\mathbf{F}'$$

Bezeichne:

- $\mathbf{K}_t = \mathbf{F}\boldsymbol{\Sigma}_t\mathbf{H}(\mathbf{H}'\boldsymbol{\Sigma}_t\mathbf{H})^{-1}$ die sogenannte *Gain-Matrix*,
- $\hat{\mathbf{x}}_t = \hat{\mathbf{x}}_{t|t-1}$ die lineare Projektion,
- wobei $\boldsymbol{\Sigma}_t = \boldsymbol{\Sigma}_{t|t-1}$ der zugehörige *Mean Squared Error* ist.

Dann lässt sich der Kalman-Filter folgendermaßen zusammenfassen:

① Initialisierung mit

- $\hat{\mathbf{x}}_1 = \mathbf{0}$,
- $\text{vec}(\boldsymbol{\Sigma}_{1|0}) = (\mathbf{I} - \mathbf{F} \otimes \mathbf{F})^{-1} \text{vec}(\mathbf{G}\mathbf{V}\mathbf{G}')$.

② Rekursion mit

- $\mathbf{u}_t = \mathbf{X}_t - \hat{\mathbf{X}}_t = \mathbf{X}_t - \mathbf{H}'\hat{\mathbf{x}}_t, \quad E(\mathbf{u}_t\mathbf{u}_t') = \mathbf{H}'\boldsymbol{\Sigma}_t\mathbf{H} \equiv \boldsymbol{\Omega}_t,$
- $\hat{\mathbf{x}}_{t+1} = \mathbf{F}\hat{\mathbf{x}}_t + \mathbf{K}_t\mathbf{u}_t,$
- $\boldsymbol{\Sigma}_{t+1} = \mathbf{F}\boldsymbol{\Sigma}_t\mathbf{F}' + \mathbf{G}\mathbf{V}\mathbf{G}' - \mathbf{K}_t\mathbf{H}'\boldsymbol{\Sigma}_t\mathbf{F}'.$

Log-Likelihood

Mithilfe der Normalverteilungsannahme des Prognosefehlers \mathbf{u}_t lässt sich nun die Verteilung der Daten \mathbf{X}_t bedingt auf $(\mathbf{x}_t, \mathbf{X}_{t-1}, \mathbf{X}_{t-2}, \dots)$ bestimmen und die Log-Likelihood Funktion aufstellen:

Log-Likelihood

$$\begin{aligned}\log \mathcal{L}(\mathbf{X}|\boldsymbol{\mu}) &= \sum_{t=1}^T \log \mathcal{L}(\mathbf{X}_t|\boldsymbol{\mu}) \\ &= -\frac{nT}{2} \log(2\pi) - \frac{1}{2} \sum_{t=1}^T \log |\boldsymbol{\Omega}_t| - \frac{1}{2} \sum_{t=1}^T \mathbf{u}_t' \boldsymbol{\Omega}_t^{-1} \mathbf{u}_t.\end{aligned}$$

Verfahren vollständiger Information

- 1 Idee
- 2 Kalman-Filter
 - Notation
 - Initialisierung
 - Rekursion
 - Zusammenfassung und Likelihood
- 3 **Maximum-Likelihood**
 - Idee
 - Aufgabe 4: An und Schorfheide (2007) mit Maximum-Likelihood
- 4 Bayesianische Schätzung
 - Idee
 - Metropolis-Hastings-Algorithmus
 - Hinweise
 - Aufgabe 5: Schätzung mit Bayesianischen Methoden
 - Inferenz, Prognose, Modellvergleich und Identifikation
- 5 Diskussion

Maximum-Likelihood

Idee

- **Betrachtungsweise:** Parametervektor μ ist fix und die Daten sind eine zufällige Realisation unter vielen.
- Der *Maximum-Likelihood*-Schätzer $\hat{\mu}_{\text{ML}}$ ist dann definiert als

$$\hat{\mu}_{\text{ML}} = \underset{\mu}{\operatorname{argmax}} \left\{ \sum_{t=1}^T \log \mathcal{L}(\mathbf{X}_t | \mu) \right\}.$$

- Unter geeigneten Regularitätsbedingungen ist der *ML*-Schätzer konsistent, asymptotisch effizient und asymptotisch normalverteilt.
- Die Unsicherheit bzw. Möglichkeit der Inferenz beruht auf der Annahme, dass zu **jeder anderen Realisation der Daten ein etwas anderer Parametervektor gehört, der die Likelihood maximiert.**
- Hinweis für die Schätzung der Parameter eines DSGE-Modells:
 - Die Dimension von \mathbf{X}_t muss mindestens so groß sein, wie die Dimension der strukturellen Schocks v_t , da es sonst zu einer singulären Varianz-Kovarianz-Matrix des Residualterms führt.
 - Falls nicht: entweder Messfehler oder weitere Schocks hinzufügen.

Aufgabe 4: An und Schorfheide (2007) mit ML

Betrachten Sie folgendes linearisierte neukeynesianische Modell:

$$\hat{y}_t = E_t[\hat{y}_{t+1}] + \hat{g}_t - E_t[\hat{g}_{t+1}] - \frac{1}{\tau}(\hat{R}_t - E_t[\hat{\pi}_{t+1}] - E_t[\hat{z}_{t+1}]),$$

$$\hat{\pi}_t = \beta E_t[\hat{\pi}_{t+1}] + \kappa(\hat{y}_t - \hat{g}_t),$$

$$\hat{c}_t = \hat{y}_t - \hat{g}_t,$$

$$\hat{R}_t = \rho_R \hat{R}_{t-1} + (1 - \rho_R)\psi_1 \hat{\pi}_t + (1 - \rho_R)\psi_2 (\hat{y}_t - \hat{g}_t) + \epsilon_{R,t}$$

$$\hat{g}_t = \rho_g \hat{g}_{t-1} + \epsilon_{g,t},$$

$$\hat{z}_t = \rho_z \hat{z}_{t-1} + \epsilon_{z,t},$$

wobei alle Variablen mit $\hat{\cdot}$, die logarithmierte Abweichungen vom steady-state bezeichnen, d.h. $\hat{x}_t = \log(x_t) - \log(x)$.

Die stochastischen Schocks sind normalverteilt mit $E[\epsilon_{i,t}] = 0$ und $E[\epsilon_{i,t}^2] = \sigma_i^2$.

Aufgabe 4: An und Schorfheide (2007) mit ML

- Ihnen liegen Daten auf Quartalsebene zu folgenden Größen vor:
 - Quartalswachstumsraten des pro-Kopf BIPs in Prozent (YGR_t),
 - Annualisierte Inflationsraten in Prozent ($INFL_t$),
 - Annualisierte Nominalzinssätze in Prozent (INT_t).
- Modellvariablen und beobachtbare Daten sind folgendermaßen miteinander verknüpft:

$$YGR_t = \gamma^{(Q)} + 100(\hat{y}_t - \hat{y}_{t-1} + \hat{z}_t),$$

$$INFL_t = \pi^{(A)} + 400\hat{\pi}_t,$$

$$INT_t = \pi^{(A)} + r^{(A)} + 4\gamma^{(Q)} + 400\hat{R}_t.$$

- Die Parameter $\gamma^{(Q)}$, $\pi^{(A)}$ und $r^{(A)}$ haben folgende Beziehung zu den *steady-state* Werten:

$$\gamma = \frac{A_{t+1}}{A_t} = e^{\frac{\gamma Q}{100}} \approx 1 + \frac{\gamma^{(Q)}}{100}, \quad \pi = e^{\frac{\pi^{(A)}}{400}} \approx 1 + \frac{\pi^{(A)}}{400}, \quad r = e^{\frac{r^{(A)}}{400}} \approx 1 + \frac{r^{(A)}}{400},$$

$$\beta = e^{-\frac{r^{(A)}}{400}} \approx \frac{1}{1 + r^{(A)}/400}.$$

Aufgabe 4: An und Schorfheide (2007) mit ML

Erstellen Sie eine mod-Datei für das Modell, mit der Sie die Parameter mithilfe der Maximum-Likelihood-Methode schätzen können.

- (a) Verwenden Sie zur Schätzung den simulierten Datensatz `simdat1.mat`. Dabei wurden folgende Parameter verwendet:

$$\begin{aligned}\tau &= 2.000, & \kappa &= 0.150, & \psi_1 &= 1.500, & \psi_2 &= 1.000, \\ \rho_R &= 0.600, & \rho_z &= 0.650, & \rho_g &= 0.950, \\ \sigma_R &= 0.2/100, & \sigma_g &= 0.8/100, & \sigma_z &= 0.45/100, \\ \pi^{(A)} &= 4.000, & \gamma^{(Q)} &= 0.500, & r^{(A)} &= 0.400.\end{aligned}$$

- ① Schätzen Sie alle Parameter mit ML. Warum gelingt dies nicht?

Hinweis: Das nichtlineare Modell impliziert, dass $\beta = \frac{\gamma}{r} = \frac{e^{\frac{\gamma^{(Q)}}{100}}}{e^{\frac{r^{(A)}}{400}}}$.

- ② Kalibrieren Sie nun die Parameter ψ_1 und $r^{(A)}$ auf ihren wahren Wert und schätzen Sie die restlichen Parameter. Warum gelingt dies nun?

Aufgabe 4: An und Schorfheide (2007) mit ML

- (b) Verwenden Sie zur Schätzung nun den simulierten Datensatz `simdat2.mat`. Bei diesem Datensatz wurde $r^{(A)} = 4$ verwendet.
- ① Schätzen Sie alle Parameter mit ML. Warum gelingt dies immer noch nicht?
 - ② Kalibrieren Sie nun den Parameter ψ_1 auf seinen wahren Wert und schätzen Sie die restlichen Parameter. Beurteilen Sie ihr Ergebnis.

Verfahren vollständiger Information

- 1 Idee
- 2 Kalman-Filter
 - Notation
 - Initialisierung
 - Rekursion
 - Zusammenfassung und Likelihood
- 3 Maximum-Likelihood
 - Idee
 - Aufgabe 4: An und Schorfheide (2007) mit Maximum-Likelihood
- 4 **Bayesianische Schätzung**
 - Idee
 - Metropolis-Hastings-Algorithmus
 - Hinweise
 - Aufgabe 5: Schätzung mit Bayesianischen Methoden
 - Inferenz, Prognose, Modellvergleich und Identifikation
- 5 Diskussion

- Die Erfahrung zeigt, dass es ziemlich schwer sein kann ein DSGE-Modell mit Maximum-Likelihood zu schätzen.
- Daten sind oft nicht ausreichend informativ, d.h. die Likelihood ist flach in einigen Richtungen (Identifizierung).
- DSGE-Modell sind an sich misspezifiziert. Dies kann zu absurden Parameterwerten führen.

Bayesianische Schätzung

Idee

- Basiert auch auf der Likelihood: der kompletten Charakterisierung des Daten generierenden Prozesses.
- **Betrachtungsweise:** Parametervektor μ ist eine Zufallsvariable und die Daten \mathbf{X} sind fix.
- Die Idee ist, bekannte Informationen (die Daten) mit zusätzlichen Auffassungen (*prior-believes*) über die Parameter zu kombinieren und Aussagen über die bedingte Wahrscheinlichkeitsverteilung der Parameter zu treffen.
- Bei der Schätzung kann so mehr Gewicht auf den vermuteten Bereich des Parameterraums gelegt werden.
- Bayesianische Methoden bilden somit eine Brücke zwischen der Kalibrierung und dem *Maximum-Likelihood*-Verfahren:

„Bayesian Inference is a Way of Thinking, Not a Basket of Methods“

Bayesianische Schätzung

Idee

- Likelihood-Funktion $\mathcal{L}(\mathbf{X}|\mu)$ wird als auf die Parameter bedingte Dichte der beobachtbaren Daten interpretiert: $\wp(\mathbf{X}|\mu) = \mathcal{L}(\mathbf{X}|\mu)$.
- Bezeichne $\wp(\mu)$ die bekannte Priori-Dichtefunktion des Parametervektors, dann folgt nach der Regel von Bayes

$$\wp(\mu|\mathbf{X}) = \frac{\mathcal{L}(\mathbf{X}|\mu)\wp(\mu)}{\wp(\mathbf{X})} = \frac{\mathcal{L}(\mathbf{X}|\mu)\wp(\mu)}{\int \wp(\mu)\mathcal{L}(\mathbf{X}|\mu) d\mu} \propto \mathcal{L}(\mathbf{X}|\mu)\wp(\mu),$$

wobei \propto „proportional zu“ bedeutet.

- $\wp(\mathbf{X})$ ist die *marginale Likelihood* der Daten und letztlich nur eine Normierungskonstante, die nicht vom Parametervektor abhängt.
- Lässt man diese weg, so bleibt die Gestalt der Posteriori-Dichte $\wp(\mu|\mathbf{X})$ erhalten, sie integriert sich lediglich nicht zu eins.
- Diese nicht normalisierte Dichte wird auch *Posterior-Kernel* oder, logarithmiert, als *Log-Posterior-Kernel* bezeichnet.

Bayesianische Schätzung

Idee

- Der Modus ist dann der bayesianische Schätzer $\hat{\mu}_{\mathbf{B}}$ für den wahren Parametervektor:

$$\hat{\mu}_{\mathbf{B}} = \operatorname{argmax}_{\mu} \{ \log \wp(\mu | \mathbf{X}) \} = \operatorname{argmax}_{\mu} \{ \log \mathcal{L}(\mathbf{X} | \mu) + \log \wp(\mu) \}$$

- Vorgehen: Log-Likelihood mithilfe des Kalman-Filters berechnen und *Log-Posterior-Kernel* mit *Sampling-* bzw. *Monte-Carlo-*Methoden simulieren.
- In der Literatur – und in Dynare – findet hierzu bevorzugt der *Metropolis-Hastings-Algorithmus* Verwendung.
- Inferenz lässt sich dann mithilfe der Eigenschaften der Posteriori-Verteilung durchführen.

An und Schorfheide (2007, S. 132)

The algorithm constructs a Gaussian approximation around the posterior mode and uses a scaled version of the asymptotic covariance matrix as the covariance matrix for the proposal distribution. This allows for an efficient exploration of the posterior distribution at least in the neighborhood of the mode.

- Der Algorithmus nutzt aus, dass unter recht allgemeinen Bedingungen die Momente der Verteilung asymptotisch normalverteilt sind.
- Es wird eine Sequenz von Ziehungen, sogenannte Markov-Ketten, aus einer Vorschlagsdichte erstellt.
- Diese braucht nicht identisch mit der Posteriori-Dichte zu sein, sondern gewährleistet, dass der Algorithmus Stichproben aus der ganzen Breite der Posteriori-Dichte ziehen kann.

Bayesianische Schätzung

Metropolis-Hastings-Algorithmus

- Die aktuelle Ziehung μ^* ist dabei abhängig von der vergangenen Ziehung $\mu^{(s-1)}$.
- Alle Ziehungen werden zwar gleich gewichtet, allerdings nur mit einer bestimmten Wahrscheinlichkeit α auch akzeptiert. Diese wird dabei aus dem Quotient der *Posterior-Kernel* des neuen und des aktuellen Kandidaten berechnet.
- Dieses Konstrukt ermöglicht es, dass der Algorithmus tendenziell die Ziehung von Bereichen geringer Posteriori-Wahrscheinlichkeit in Bereiche hoher Wahrscheinlichkeit verschiebt:
 - Ist $\mu^{(s-1)}$ in einem Bereich hoher Posteriori-Wahrscheinlichkeit, tendiert der Algorithmus dazu, dort zu verweilen.
 - Befindet sich $\mu^{(s-1)}$ in einem Bereich niedriger Posteriori-Wahrscheinlichkeit, werden neue Kandidaten bevorzugt akzeptiert.
- Die Wahl der Varianz-Kovarianz-Matrix der Vorschlagsdichte spielt hierbei eine entscheidende Rolle, damit α nicht zu groß, aber auch nicht zu klein gewählt wird.

Bayesianische Schätzung

Metropolis-Hastings-Algorithmus

- ① Spezifiziere c_0 , c und S .
- ② Maximiere $\log \mathcal{L}(\mathbf{X}|\boldsymbol{\mu}) + \log \wp(\boldsymbol{\mu})$ mithilfe numerischer Verfahren.
 $\hat{\boldsymbol{\mu}}_{\mathbf{B}}$ bezeichnet den Modus.
- ③ Berechne die Inverse der Hessematrix ausgewertet am Modus,
bezeichne diese mit $\boldsymbol{\Sigma}_{\mathbf{B}}$.
- ④ Spezifiziere einen Startwert $\boldsymbol{\mu}^{(0)}$ oder ziehe diesen aus $\mathcal{N}(\hat{\boldsymbol{\mu}}_{\mathbf{B}}, c_0^2 \boldsymbol{\Sigma}_{\mathbf{B}})$.

⑤ Für $s = 1, \dots, S$:

- Ziehe μ^* aus der Kandidaten-generierenden-Verteilung $\mathcal{N}(\mu^{(s-1)}, c^2 \Sigma_B)$.
- Berechne die Akzeptanz-Wahrscheinlichkeit α :

$$\alpha \equiv \alpha(\mu^{(s-1)}, \mu^*) = \frac{\mathcal{L}(\mu^* | \mathbf{X}) \varphi(\mu^*)}{\mathcal{L}(\mu^{(s-1)} | \mathbf{X}) \varphi(\mu^{(s-1)})}$$

- Mit der Wahrscheinlichkeit $\min\{\alpha, 1\}$ wird der Sprung von $\mu^{(s-1)}$ auf μ^* akzeptiert. D.h. falls $\alpha \geq 1$, setze $\mu^{(s)} = \mu^*$.
- Mit der Gegenwahrscheinlichkeit wird der Sprung nicht akzeptiert. D.h. ziehe eine gleichverteilte Zufallsvariable r zwischen 0 und 1:
 - Falls $r \leq \alpha$ setze $\mu^{(s)} = \mu^*$.
 - Falls $r > \alpha$ setze $\mu^{(s)} = \mu^{(s-1)}$.

- ⑥ Schätze den Posteriori-Erwartungswert einer Funktion $\tilde{h}(\mu)$ mit $\frac{1}{S} \sum_{s=1}^S \tilde{h}(\mu^{(s)})$.
- ⑦ Falls die durchschnittliche Akzeptanzwahrscheinlichkeit nicht einen erwünschten Wert hat (üblicherweise zwischen 20% – 30%) oder der Algorithmus nicht konvergiert, ändere c_0 , c oder S .

Bayesianische Schätzung

Hinweise

- Bayesianische Schätzung eines DSGE-Modells erfordert, dass die Anzahl der Schocks mit der Anzahl beobachtbarer Variablen übereinstimmt (nicht eindeutig identifizierbare Parameter).
- Übliche Priori-Verteilungen: Normalverteilung, (normale, verschobene oder inverse) Gammaverteilung, Betaverteilung oder Gleichverteilung.
- Bei der Wahl einer geeigneten Priori-Verteilung spielen Überlegungen zur unteren bzw. oberen Grenzen, sowie zur Schiefe und Kurtosis der Verteilung eine Rolle.
- Die Ergebnisse können stark abhängen von der Wahl der Priori oder ihrer Parametrisierung.
- Deshalb muss die Robustheit der Resultate sicher gestellt werden:
 - Unterschiedliche Parametrisierungen ausprobieren.
 - Allgemeinere Priori-Verteilungen verwenden.
 - Nichtinformative Priori-Verteilungen.
 - Sensitivitätsanalysen.

Aufgabe 5: Schätzung mit Bayesianischen Methoden

Betrachten Sie folgendes einfache RBC-Modell (soziale Planer-Problem);

$$\begin{aligned} \max_{\{c_{t+j}, l_{t+j}, k_{t+j}\}_{j=0}^{\infty}} W_t &= \sum_{j=0}^{\infty} \beta^j u(c_{t+j}, l_{t+j}) \\ \text{s.t. } y_t &= c_t + i_t, & A_t &= Ae^{a_t}, \\ y_t &= A_t f(k_{t-1}, l_t), & a_t &= \rho a_{t-1} + \varepsilon_t, \\ k_t &= i_t + (1 - \delta)k_{t-1}, & \varepsilon_t &\sim N(0, \sigma_\varepsilon^2), \end{aligned}$$

wobei für Präferenzen und Technologie folgendes gilt:

$$u(c_t, l_t) = \frac{[c_t^\theta (1 - l_t)^{1-\theta}]^{1-\tau}}{1 - \tau}, \quad f(k_{t-1}, l_t) = [\alpha k_{t-1}^\psi + (1 - \alpha)l_t^\psi]^{1/\psi}.$$

Die Optimalitätsbedingungen sind gegeben durch:

$$\begin{aligned} u_c(c_t, l_t) - \beta E_t \{u_c(c_{t+1}, l_{t+1}) [A_{t+1} f_k(k_t, l_{t+1}) + 1 - \delta]\} &= 0, \\ -\frac{u_l(c_t, l_t)}{u_c(c_t, l_t)} - A_t f_l(k_{t-1}, l_t) &= 0, \\ c_t + k_t - A_t f(k_{t-1}, l_t) - (1 - \delta)k_{t-1} &= 0. \end{aligned}$$

Aufgabe 5: Schätzung mit Bayesianischen Methoden

- (a) Schreiben Sie eine mod-Datei für dieses Modell (mit einer vernünftigen Kalibrierung der Parameter und einem steady-state Block).
- (b) Simulieren Sie einen Datensatz von 10000 Beobachtungen für c_t , l_t und y_t mit `stoch_simul` und speichern Sie diesen in einer mat-Datei.
- (c) Definieren Sie Priori-Verteilungen für α , θ und τ (oder einer anderen Auswahl an Parametern).
- (d) Schätzen Sie den Posteriori-Modus (mithilfe des `estimation` Kommandos). Beschränken Sie Ihren Datensatz auf 100 Beobachtungen. Wie viele beobachtbare Variablen benötigen Sie? Überprüfen Sie den Posteriori-Modus mithilfe des `mode_check` Kommandos. Falls Sie Warnmeldungen aufgrund einer nicht positiv-definiten Hessematrix erhalten, verwenden Sie einen anderen Optimierungsalgorithmus oder verändern Sie ihre Anfangswerte.

Aufgabe 5: Schätzung mit Bayesianischen Methoden

- (e) Wenn Sie zufrieden sind mit dem Wert Ihres Posteriori-Modus, approximieren Sie die Posteriori-Verteilung mithilfe des Metropolis-Hastings-Algorithmus mit 3×5000 Iterationen. Falls dies nicht zu der (ergodischen) Posteriori-Verteilung konvergiert, wiederholen Sie den Algorithmus ohne die vorherigen Ziehungen zu verwerfen.
- (f) Wie steht es um die Robustheit ihres Ergebnisses bezüglich der Spezifikation ihrer Priori-Verteilungen? Wiederholen Sie die Schätzung des Posteriori-Modus für verschiedene Priori-Verteilungen.

Verfahren vollständiger Information

- 1 Idee
- 2 Kalman-Filter
 - Notation
 - Initialisierung
 - Rekursion
 - Zusammenfassung und Likelihood
- 3 Maximum-Likelihood
 - Idee
 - Aufgabe 4: An und Schorfheide (2007) mit Maximum-Likelihood
- 4 **Bayesianische Schätzung**
 - Idee
 - Metropolis-Hastings-Algorithmus
 - Hinweise
 - Aufgabe 5: Schätzung mit Bayesianischen Methoden
 - Inferenz, Prognose, Modellvergleich und Identifikation
- 5 Diskussion

Inferenz, Prognose, Modellvergleich und Identifikation

Eigenschaften der Posteriori-Verteilung

- Die Posteriori-Verteilung verknüpft alle zur Verfügung stehenden Daten über μ : sowohl Informationen der Daten als auch derjenigen, bevor man die Daten beobachtet hat.
- Bayesianische Inferenz funktioniert für jeden Beobachtungszeitraum, hat zusätzlich aber auch folgende asymptotische Eigenschaften:
 - ① Die Priori-Verteilung wird bei zunehmenden Beobachtungszeitraum irrelevant für die Bestimmung der Posteriori-Verteilung.
 - ② Die Posteriori-Verteilung konvergiert zu einer degenerierten Verteilung um den wahren Wert.
 - ③ Die Posteriori-Verteilung ist approximativ normal-verteilt.
- Mithilfe der Posteriori-Verteilung lassen sich nun
 - Bayesianische Konfidenzintervalle (credibility sets) aufstellen,
 - Prognosen erstellen mithilfe der predictive-density:
$$\mathcal{L}(\mathbf{X}_f|\mathbf{X}) = \int \mathcal{L}(\mathbf{X}_f|(\mu|\mathbf{X}))d\mu = \int \mathcal{L}(\mathbf{X}_f|\mu, \mathbf{X})\wp(\mu|\mathbf{X})d\mu$$
 - Modellvergleiche durchführen.

- Modelle können sich in ihrer Priori-Verteilung, ihrer Likelihood oder ihren Parametern unterscheiden.
- Bayesianische Herangehensweise: Berechne die Wahrscheinlichkeit, dass Modell i das wahre Modell ist, gegeben der Daten.
- O.B.d.A.: Angenommen es gibt $i = 1, 2$ Modelle M_i mit jeweiliger Priori-Wahrscheinlichkeit $p_i = P(M_i)$, dass Modell M_i das wahre Modell ist.
- Jedes Modell besitzt eine Menge von Parametern μ_i , einer Priori-Verteilung für diese $\wp_i(\mu_i)$ und einer Likelihood $\mathcal{L}_i(\mathbf{X}|\mu)$.
- Dann ist die Wahrscheinlichkeit, dass Modell 1 das wahre Modell ist, gegeben der Daten:

$$\begin{aligned} P(M_1|\mathbf{X}) &= \frac{P(M_1)\mathcal{L}_1(\mathbf{X}|M_1)}{\mathcal{L}(\mathbf{X})} = \frac{p_1 \int \mathcal{L}_1(\mathbf{X}, \mu_1|M_1)d\mu_1}{\mathcal{L}(\mathbf{X})} \\ &= \frac{p_1 \int \mathcal{L}_1(\mathbf{X}|\mu_1, M_1)\wp_1(\mu_1|M_1)d\mu_1}{\mathcal{L}(\mathbf{X})} \end{aligned}$$

$$\text{mit } \mathcal{L}(\mathbf{X}) = p_1 \int \mathcal{L}_1(\mathbf{X}|\mu_1, M_1)\wp_1(\mu_1|M_1)d\mu_1 + p_2 \int \mathcal{L}_2(\mathbf{X}|\mu_2, M_2)\wp_2(\mu_2|M_2)d\mu_2$$

- Der erwartete Wert der Likelihood unter Berücksichtigung der Priori-Verteilung ist die *Marginal-Likelihood* für Modell i :

$$m_i(\mathbf{X}) = \int \mathcal{L}_i(\mathbf{X}|\boldsymbol{\mu}_i, M_i) \varphi_i(\boldsymbol{\mu}_i|M_i) d\boldsymbol{\mu}_i$$

- Daraus lassen sich nun die *Posterior-Odds* berechnen:

$$PO_{12} = \frac{P(M_1|\mathbf{X})}{P(M_2|\mathbf{X})} = \underbrace{\frac{p_1}{p_2}}_{\text{Prior-Odds-Ratio}} \cdot \underbrace{\frac{m_1(\mathbf{X})}{m_2(\mathbf{X})}}_{\text{Bayes-Faktor}}$$

- Zusammen mit $P(M_1|\mathbf{X}) + P(M_2|\mathbf{X}) = 1$ ergeben sich die *Posterior-Model-Probabilities*:

$$P(M_1|\mathbf{X}) = \frac{PO_{12}}{1 + PO_{12}}, \quad P(M_2|\mathbf{X}) = 1 - P(M_1|\mathbf{X}).$$

- Die *Marginal Likelihood* misst die Qualität eines Modells die Daten zu charakterisieren.
- Die *Posterior-Odds* geben keinen Hinweis auf das wahre Modell, sie beschreiben lediglich, welches Modell im Vergleich zu anderen Modellen die höchste bedingte Wahrscheinlichkeit besitzt.
- Ein großes $PO_{12} \gg 1$ spricht dafür, dass die Daten und die Priori-Verteilungen das Modell 1 bevorzugen.
- Guidelines von Jeffrey (1961):
 - $1 : 1 - 3 : 1$ sehr schwache Evidenz für Modell 1,
 - $10 : 1 - 100 : 1$ starke Evidenz für Modell 1,
 - $> 100 : 1$ entscheidende Evidenz für Modell 1.
- Die Implementation und Berechnung der Integrale geschieht mit numerischen MCMC- und Sampling-Methoden, sowie der Laplace- bzw. Harmonic-Mean-Approximation.

Aufgabe 5: Schätzung mit Bayesianischen Methoden

- (g) Benutzen Sie denselben Datensatz, um die Parameter eines fehlspezifizierten Modells zu schätzen. Verwenden Sie hierzu das selbe Modell, jedoch entweder mit
- **einer Cobb-Douglas Produktionsfunktion** \Leftarrow
 - oder einer separierbaren Nutzenfunktion
 - oder einem Modell, indem der Haushalt unelastisch genau eine Einheit Arbeit anbietet.
- Hinweis:** *Vergessen Sie nicht sowohl die Modellgleichungen als auch die steady-state Gleichungen entsprechend anzupassen.*
- (h) Vergleichen Sie nun die Schätzung der gemeinsamen Parameter sowie die marginalen Dichten der verschiedenen Modelle. Berechnen Sie auch die *Posterior-Odds* bzw. die *Posterior-Model-Probabilities*.

- Betrachte zwei Parametervektoren μ_1 und μ_2 für die folgendes gilt:

$$\mathcal{L}(\mathbf{X}|\mu_1) = \mathcal{L}(\mathbf{X}|\mu_2)$$

- Falls $\mu_1 = \mu_2$ spricht man von Identifikation, falls dies jedoch für $\mu_1 \neq \mu_2$ gilt, so weiß man nicht, welcher Parametervektor die Daten generiert hat.
- Spezialfall: $\mathcal{L}(\mathbf{X}|\mu_1, \mu_2) = \mathcal{L}(\mathbf{X}|\mu_1)$.
 - μ_2 ist durch die Daten nicht identifiziert.
 - Was gilt für die Posteriori-Verteilung $\wp(\mu_2|\mathbf{X})$?

$$\begin{aligned}\wp(\mu_2|\mathbf{X}) &= \int \wp(\mu_1, \mu_2|\mathbf{X}) d\mu_1 \\ &= \left[\int \mathcal{L}(\mathbf{X}|\mu_1, \mu_2) \wp(\mu_1) \wp(\mu_2|\mu_1) d\mu_1 \right] \cdot [\mathcal{L}(\mathbf{X})]^{-1} \\ &= \left[\int \mathcal{L}(\mathbf{X}|\mu_1) \wp(\mu_1) \wp(\mu_2|\mu_1) d\mu_1 \right] \cdot [\mathcal{L}(\mathbf{X})]^{-1} \\ &= \int \wp(\mu_1|\mathbf{X}) \wp(\mu_2|\mu_1) d\mu_1\end{aligned}$$

- Falls die Priori-Verteilung von μ_2 unabhängig von μ_1 ist, d.h. $\wp(\mu_2|\mu_1) = \wp(\mu_2)$, dann folgt $\wp(\mu_2|\mathbf{X}) = \wp(\mu_2)$. Beliefs über μ_2 werden durch das Beobachten der Daten nicht modifiziert.
- Falls die Unabhängigkeit der Priori-Verteilungen nicht gilt, so beeinflusst die Information der Daten die Beliefs über μ_2 über den Einfluss auf μ_1 .
- Falls also eine Identifikation mithilfe der Daten nicht möglich ist, ist jeglicher Unterscheid zwischen der Priori- und Posteriori-Verteilung auf die Priori-Verteilung zurückzuführen.
- Praktisch muss man sich damit auseinander setzen, ob Parameter durch die Daten oder durch die Priori-Verteilungen identifiziert sind.

Verfahren vollständiger Information

- 1 Idee
- 2 Kalman-Filter
 - Notation
 - Initialisierung
 - Rekursion
 - Zusammenfassung und Likelihood
- 3 Maximum-Likelihood
 - Idee
 - Aufgabe 4: An und Schorfheide (2007) mit Maximum-Likelihood
- 4 Bayesianische Schätzung
 - Idee
 - Metropolis-Hastings-Algorithmus
 - Hinweise
 - Aufgabe 5: Schätzung mit Bayesianischen Methoden
 - Inferenz, Prognose, Modellvergleich und Identifikation
- 5 Diskussion

- Restriktivere Annahmen als bei den Verfahren begrenzter Information: Spezifikation der Verteilung der Schocks, die sogenannte Likelihood.
- Vorteile einer *Maximum-Likelihood*-Schätzung liegen in der vollen Charakterisierung des Daten generierenden Prozesses und der genaueren, konsistenten sowie effizienteren Schätzung der Parameter.
- „Dilemma of absurd parameter estimates“: Problem der ML-Schätzung, aufgrund fehlerhaften Verteilungsannahmen, Problemen im Optimierungsalgorithmus oder bei nicht separat identifizierbaren Parametern.
- Auch Transformationen, obere und untere Grenzen, etc. können nur bedingt helfen, wenn die Likelihood-Funktion sehr flach verläuft.

Grundproblem

Für einen mathematischen Ausdruck mit vielen Bedingungen und Parametern, aber nur einem begrenzten Beobachtungszeitraum, können verschiedene Parameterkonstellationen existieren, die zum selben Ergebnis und somit auch zu einem ähnlichen Datensatz führen.

- An diesem Punkt setzen bayesianische Methoden an und bilden eine Brücke zwischen der Kalibrierung und dem *ML-Prinzip*.
 - Durch das Berücksichtigen von Priors lassen sich auch nicht modellierte Informationen im Modell verarbeiten.
 - „Dilemma of absurd parameter estimates“: Auch mit bayesianischen Mitteln werden diese Parameter nicht geschätzt (die Priori-Verteilung unterscheidet sich kaum von der Posteriori-Verteilung), sondern durch eine Belegung mit Wahrscheinlichkeit festgelegt.
- ⇒ Durch die Wahl von Priors lassen sich diese absurd erscheinenden Werte ausschließen.
- Trotzdem bleibt die Frage nach Robustheit und der Identifizierbarkeit der Parameter kritisch.

An und Schorfheide (2006, S.124)

Once one acknowledges that the DSGE model provides merely an approximation to the law of motion of the time series (...), then it seems reasonable to assume that there need not exist a single parameter vector (...), that delivers, say, the „true“ intertemporal substitution elasticity or price adjustment costs and, simultaneously, the most precise impulse responses to a technology or monetary policy shock. Each estimation method is associated with a particular measure of discrepancy between the „true“ law of motion and the class of approximating models.