

Higher-Order Perturbation & Penalty Functions

Wouter J. Den Haan
with contributions by Joris de Wind & Ken Judd

August 15, 2010

Outline

- ➊ Reasons why nonlinearities matter more when modelling idiosyncratic risk
- ➋ Problems with higher-order perturbation solutions
- ➌ Using penalty functions instead of inequality constraints

Non-linearities more important for individual

Reasons:

- ① Higher variance state variables
- ② Frictions
- ③ Inequality constraints matter

Need for higher-order perturbation solutions?

- for risk to matter \implies need *at least* 2nd-order
- welfare comparison \implies need *at least* 2nd-order
- for risk premiums to be cyclical \implies need *at least* 3th-order
- idiosyncratic risk \implies need *at least* ?th-order
- models with interesting frictions \implies need *at least* ?th-order
- models about the financial crisis \implies need *at least* ?th-order

Problems of higher-order perturbation

- Well-known problem for lots of model solvers
- Higher-order perturbation solutions are often explosive
- Standard solution is pruning:
 - this creates an ugly *distortion* of underlying perturbation solution
- Perturbation solutions have more problems
 - for example weird shapes
- What can be done?

Outline

- Polynomial approximations and its problems
- Pruning and its problems
- Understanding what perturbation is
- Understanding the flexibility of perturbation
- Some ideas on how to exploit this flexibility

Polynomial approximations

$$x_{+1} = h(x) \approx p_N(x; \alpha_N)$$

How to find α_N ?

- Perturbation, Taylor series expansion around \bar{x}
- Projection method

Problems of higher-order polynomials

- oscillating patterns \implies not shape preserving
- often explosive behavior

$$x_{+1} = h(x) \approx p_N(x)$$

$$\lim_{x \rightarrow \infty} \frac{\partial p_N(x)}{\partial x} = \pm \infty$$

$$\lim_{x \rightarrow +\infty} \frac{\partial p_N(x)}{\partial x} = +\infty \implies \text{no global convergence}$$

$$\lim_{x \rightarrow +\infty} \frac{\partial p_N(x)}{\partial x} = -\infty \implies \text{function must turn negative}$$

Is convergence guaranteed?

- Projection methods:
 - even *uniform* convergence (with Chebyshev nodes)
 - of course only within the grid
- Taylor series expansion
 - limited radius of convergence
 - unless function is analytic
- **Huge** difference!!!
 - grid is controlled by model solver
 - radius of convergence is not

Couple examples

- sometimes you get great global approximations
- Sometimes you do not. We will look at
 - limited radius of convergence
 - problems with weird/undesirable shapes
 - stability problems

Example with simple Taylor expansion

Truth is a polynomial:

$$f(x) = -690.59 + 3202.4x - 5739.45x^2 \\ + 4954.2x^3 - 2053.6x^4 + 327.10x^5$$

defined on $[0.7, 2]$

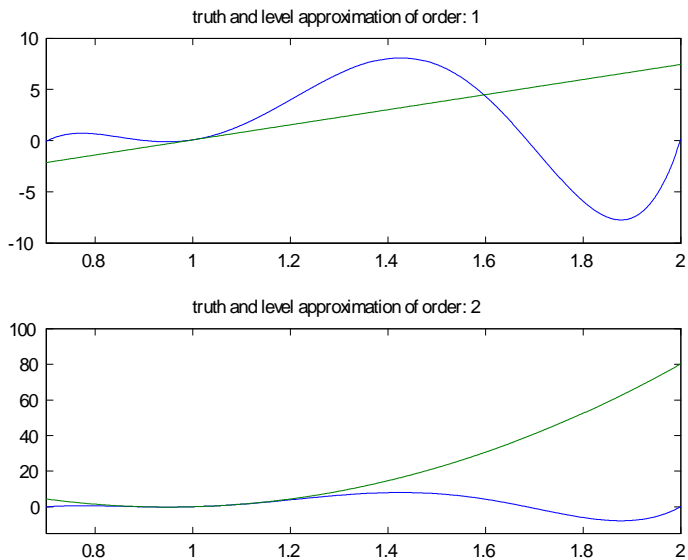


Figure: Level approximations

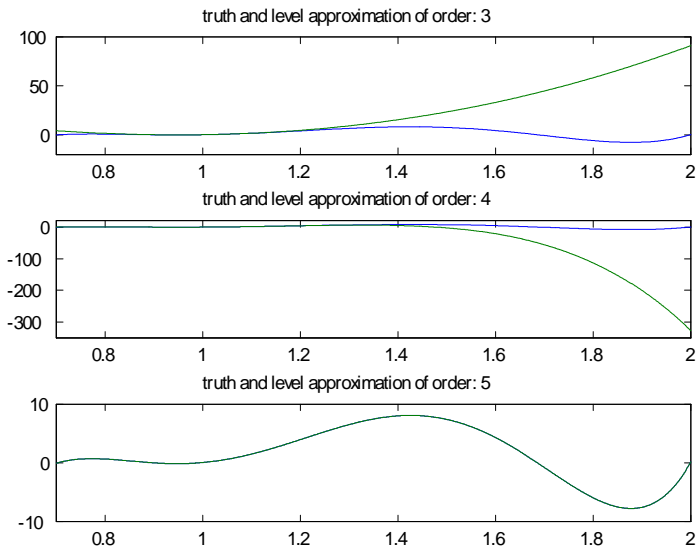


Figure: Level approximations continued

Approximation in log levels

Truth is no a polynomial.

Think of $f(x)$ as a function of $z = \log(x)$. Thus,

$$\begin{aligned} f(x) = & -690.59 + 3202.4 \exp(z) - 5739.45 \exp(2z) \\ & + 4954.2 \exp(3z) - 2053.6 \exp(4z) + 327.10 \exp(5z). \end{aligned}$$

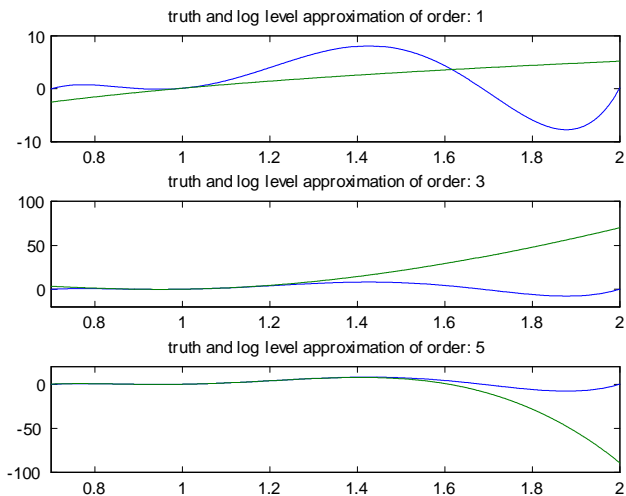


Figure: Log level approximations

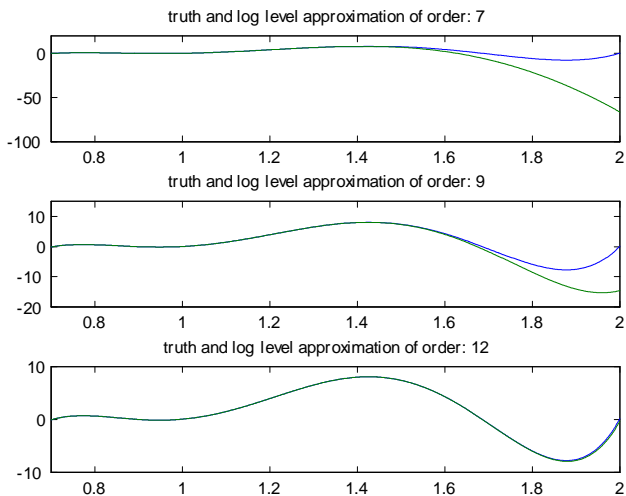


Figure: Log level approximations continued

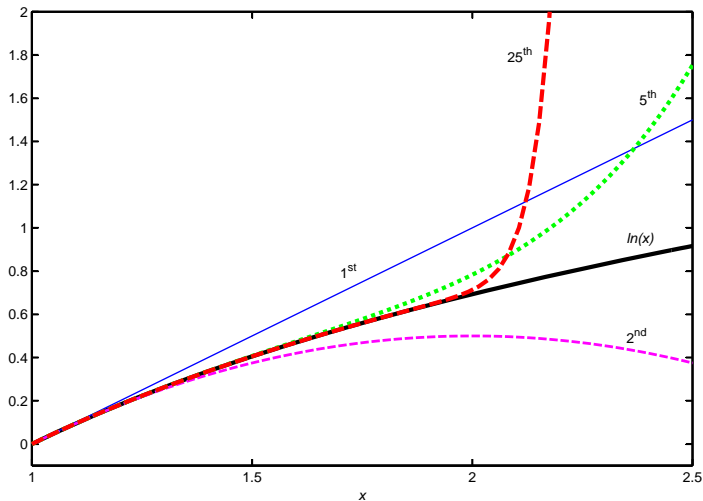
$\ln(x)$ & Taylor series expansion

$$\begin{aligned} \ln(x) - \ln(\bar{x}) &\approx \\ \frac{\tilde{x}}{\bar{x}} - \frac{1}{2!} \left(\frac{\tilde{x}}{\bar{x}}\right)^2 + \frac{2!}{3!} \left(\frac{\tilde{x}}{\bar{x}}\right)^3 - \frac{3!}{4!} \left(\frac{\tilde{x}}{\bar{x}}\right)^4 + \dots + (-1)^{N-1} \frac{(N-1)!}{N!} \left(\frac{\tilde{x}}{\bar{x}}\right)^N \\ &= \\ \frac{\tilde{x}}{\bar{x}} - \frac{1}{2} \left(\frac{\tilde{x}}{\bar{x}}\right)^2 + \frac{1}{3} \left(\frac{\tilde{x}}{\bar{x}}\right)^3 - \frac{1}{4} \left(\frac{\tilde{x}}{\bar{x}}\right)^4 + \dots + (-1)^{N-1} \frac{1}{N} \left(\frac{\tilde{x}}{\bar{x}}\right)^N \end{aligned}$$

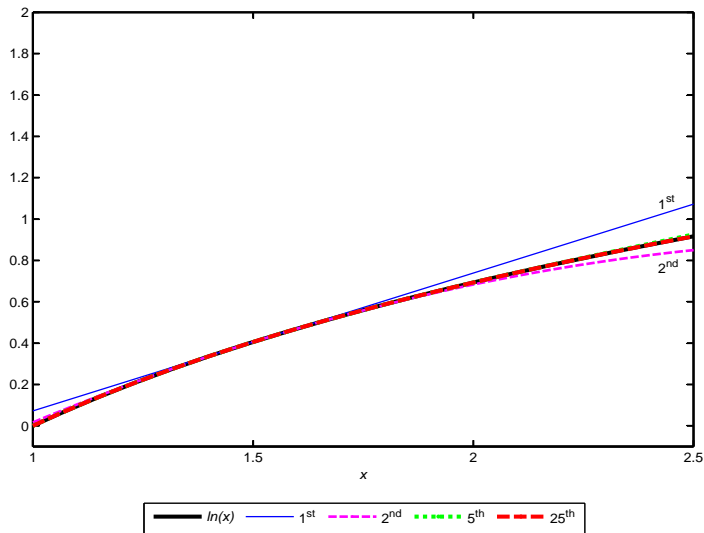
with $\tilde{x} = x - \bar{x}$

For which \tilde{x} can we expect things to go wrong?

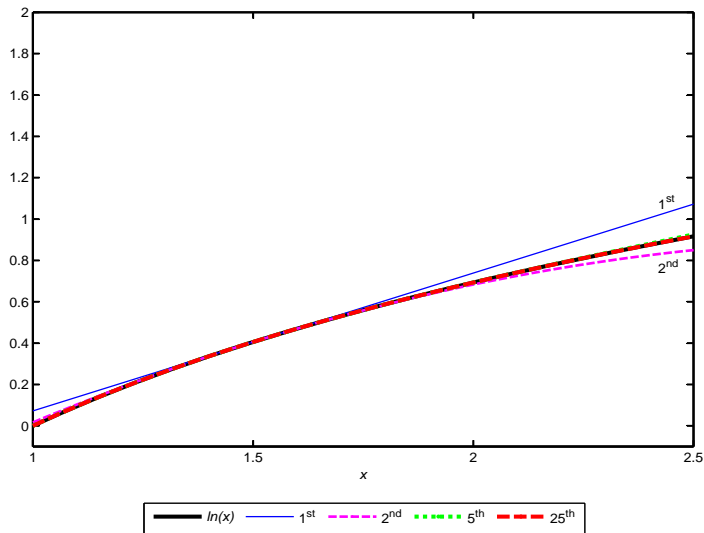
$\ln(x)$ & Taylor series expansions at $x = 1$



$\ln(x)$ & Taylor series expansions at $x = 1.5$



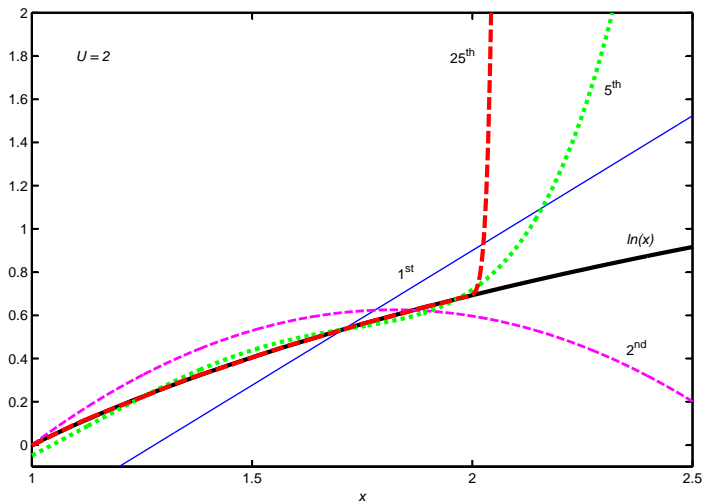
$\ln(x)$ & Taylor series expansions at $x = 1.5$



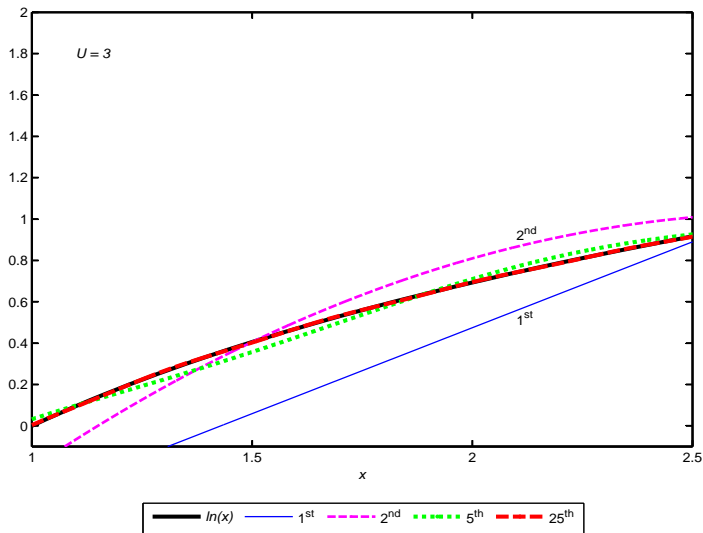
Perturbation versus projection

- Projection methods \implies uniform convergence within the grid
- You control the grid

$\ln(x)$ & projection approximation in $[0,2]$



$\ln(x)$ & projection approximation in $[0,3]$

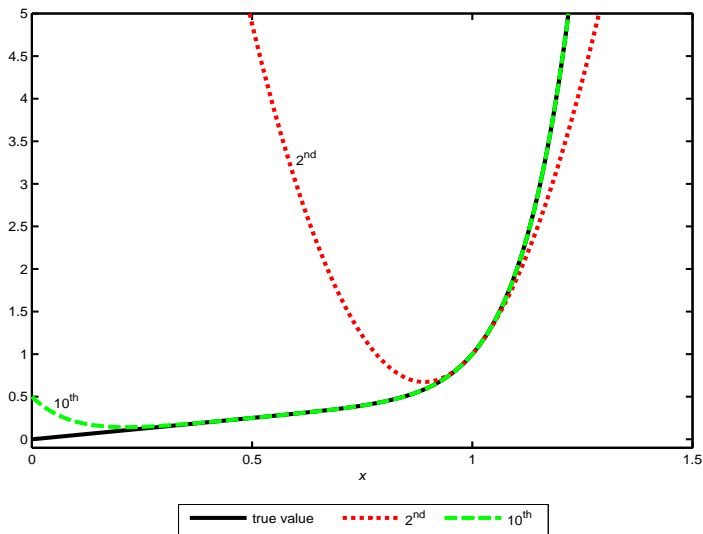


Problems with preserving shape

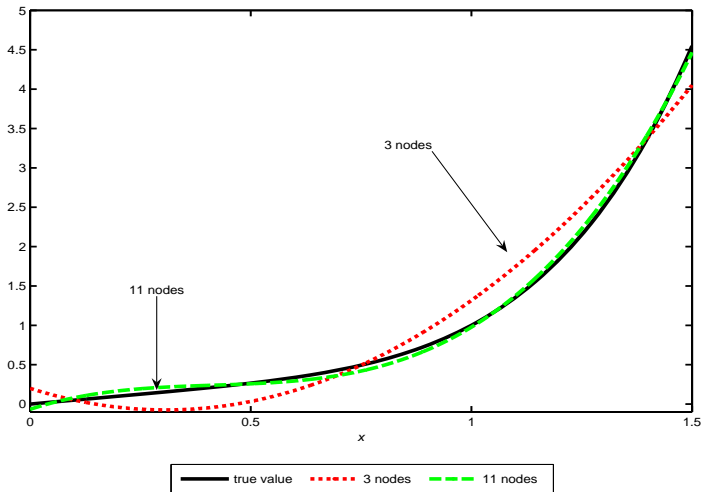
$$h(x) = 0.5x^\alpha + 0.5x$$

- α is an integer $\implies h(x)$ is a polynomial
- α is odd $\implies \partial h(x) / \partial x > 0$

Perturbation approximation & preserving shape



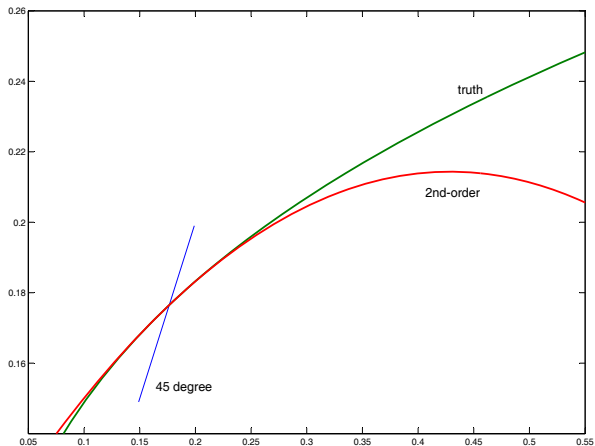
Projection approximation & preserving shape



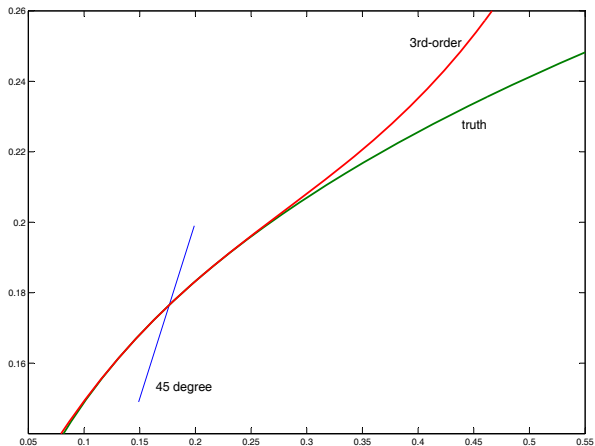
Problems with preserving shape

- nonlinear finite-order polynomials *always* have "weird" shapes
- weirdness may occur close to or far away from steady state
- thus also in the standard growth model

Standard growth model and odd shapes due to perturbation (log utility)



Standard growth model and odd shapes due to perturbation (log utility)



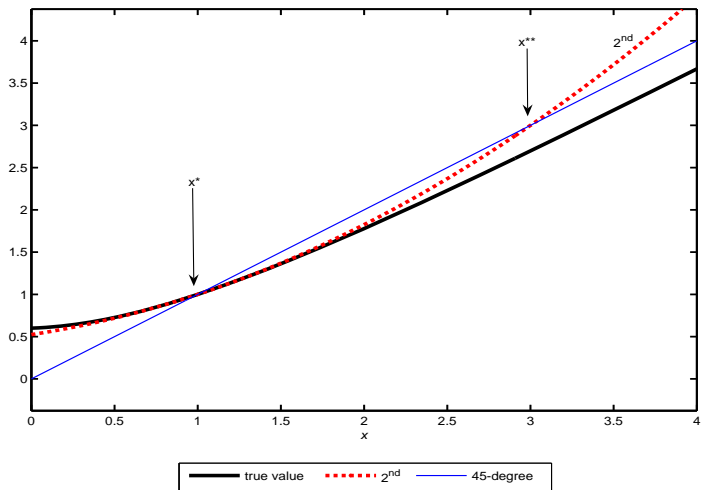
Problems with stability

$$h(x) = \alpha_0 + x + \alpha_1 e^{-\alpha_2 x}$$

$$x_{+1} = h(x) + \text{shock}_{+1}$$

- Unique globally stable fixed point

Perturbation approximation & stability



Model

$$\max_{\{c_t, a_t\}_{t=1}^{\infty}} \mathbb{E} \sum_{t=1}^{\infty} \beta^{t-1} \frac{c_t^{1-\nu} - 1}{1-\nu} - P(a_t)$$

s.t.

$$c_t + \frac{a_t}{1+r} = a_{t-1} + \theta_t$$

$$\theta_t = \bar{\theta} + \varepsilon_t \text{ and } \varepsilon_t \sim N(0, \sigma^2)$$

a_0 given.

Penalty function

Standard inequality constraint

$$a \geq 0$$

corresponds to

$$P(a) = \begin{cases} \infty & \text{if } a < 0 \\ 0 & \text{if } a \geq 0 \end{cases}$$

Flexible alternative:

$$P(a) = \frac{\eta_1}{\eta_0} \exp(-\eta_0 a) - \eta_2 a.$$

Our penalty function

- can be approximated *globally* with Taylor series expansion
- linear part, $-\eta_2 a$
 - not necessary
 - steady state can be equal to the one without penalty function

Interpreting the penalty function

- 1 penalty function *implements* inequality constraint
 - η_0 must be very high
- 2 penalty function is alternative to penalty function
 - η_0 could be high or low

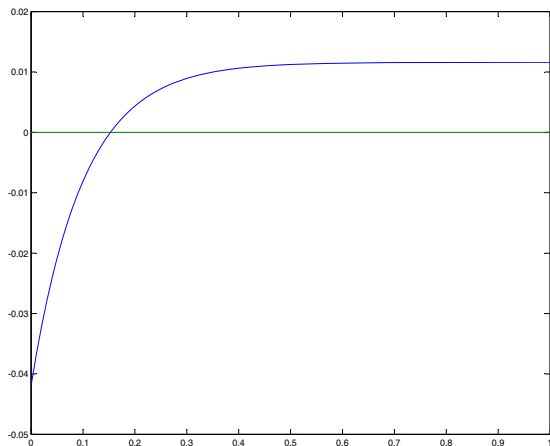
Calibrating the penalty function

- η_0 , η_1 , and η_2 can be chosen to match data characteristics
- Here:
 - different values for curvature parameter, η_0
 - η_1 and η_2 chosen to match mean and standard deviation of a_t
- many properties of this model similar to " $a \geq 0$ " model
 - but tail behavior is different

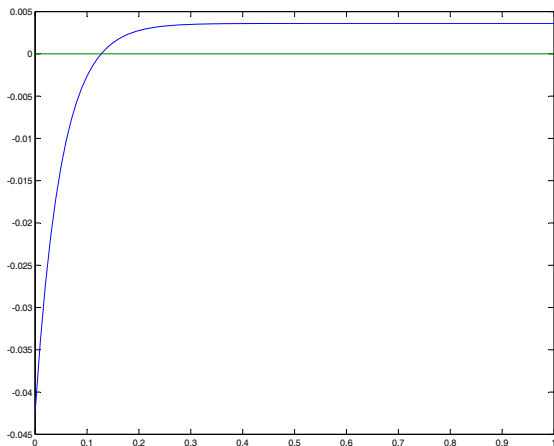
FOC

$$\frac{c_t^{-\nu}}{1+r} + \frac{\partial P(a_t)}{\partial a_t} = \beta \mathbf{E}_t [c_{t+1}^{-\nu}]$$

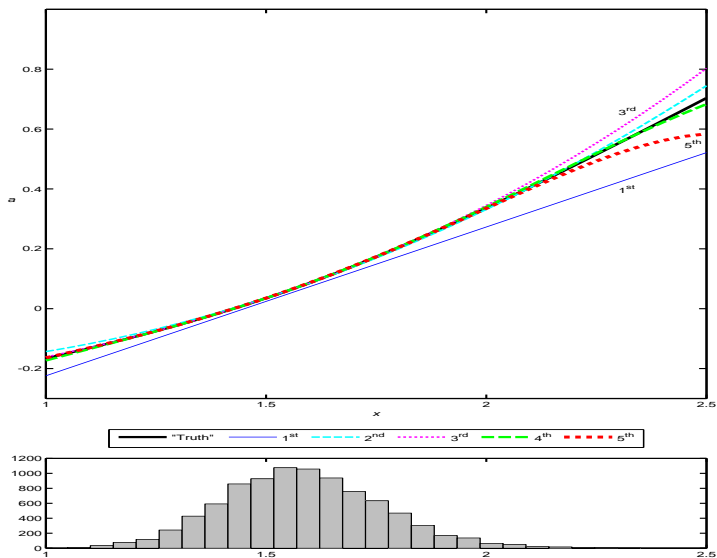
Penalty term in FOC; $\eta_0=10$



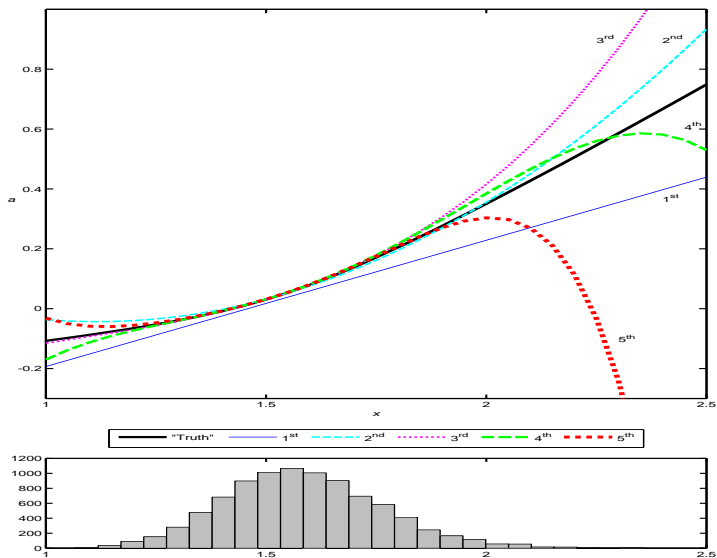
Penalty term in FOC; $\eta_0=20$



Perturbation solutions when $\eta_0 = 10$



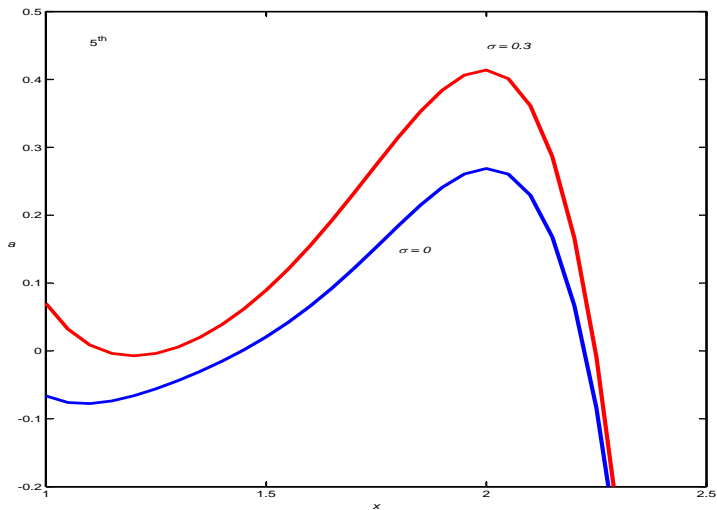
Perturbation solutions when $\eta_0 = 20$



Perturbation and higher uncertainty

- oscillations more problematic when $\sigma \uparrow$
 - (more likely to get into problematic part)
- but higher-order perturbation solution adjust when $\sigma \uparrow$
 - (problematic part may move away from steady state)

Fifth-order perturbation and uncertainty



Simulating

- 2nd & 3rd explode
- 4th & 5th are inaccurate

Pruning - procedure

All steady states are set equal to 0 to simplify notation

Pruning - procedure

1. Split up perturbation solution into two parts

$$p_{N,\text{pert}}(a_{t-1}, \theta_t) =$$

linear part $\gamma_{N,k} a_{t-1} + \gamma_{N,\theta} \theta_t$

nonlinear part $+ \tilde{p}_{N,\text{pert}}(a_{t-1}, \theta_t)$

Pruning - procedure

2. Simulate a_t^* using

$$a_t^* = \gamma_{N,k} a_{t-1}^* + \gamma_{N,\theta} \theta_t$$

3. Simulate $a_{\text{prune},t}$ using

$$\begin{aligned} & a_{\text{prune},t} \\ = & \gamma_{N,k} a_{\text{prune},t-1} + \gamma_{N,\theta} \theta_t + \tilde{p}_{N,\text{pert}}(a_{t-1}^*, \theta_t) \end{aligned}$$

Pruning - procedure

$$a_{\text{prune},t} = \gamma_{N,k} a_{\text{prune},t-1} + \gamma_{N,\theta} \theta_t + \tilde{p}_{N,\text{pert}}(a_{t-1}^*, \theta_t)$$

- $a_{\text{prune},t}$ is not a function of just the state variables
 - $a_{\text{prune},t-1}$ and θ_t
- $a_{\text{prune},t}$ also depends on $a_{t-1}^* \implies$

$a_{\text{prune},t}$ is a *correspondence* of state variables

Perturbation principle

- **Objective of perturbation:** If $h(x)$ is such that

$$f(h(x)) = 0 \quad \forall x$$

then we want to solve for

$$h_{\text{approx}}(x) = h(\bar{x}) + \left. \frac{\partial h(x)}{\partial x} \right|_{x=\bar{x}} (x - \bar{x}) + \left. \frac{\partial^2 h(x)}{\partial x^2} \right|_{x=\bar{x}} \frac{(x - \bar{x})^2}{2!} + \dots + \left. \frac{\partial^n h(x)}{\partial x^n} \right|_{x=\bar{x}} \frac{(x - \bar{x})^n}{n!}$$

- Pruning does not generate a function of the form

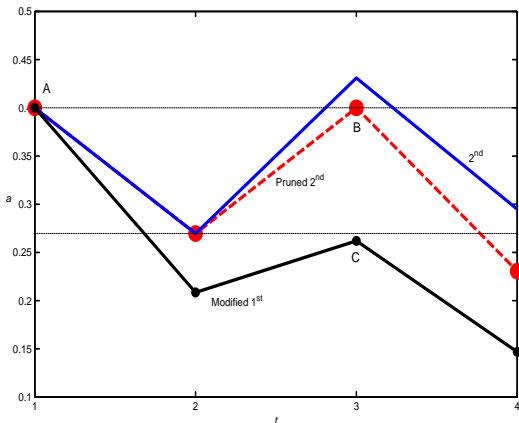
$$h(x)$$

- As a function of x you get a correspondence

Why don't you get a policy function?

Additional state variables introduced by pruning procedure
 $\implies h_{\text{prune}}$ is not a function of x

Why don't you get a policy function?

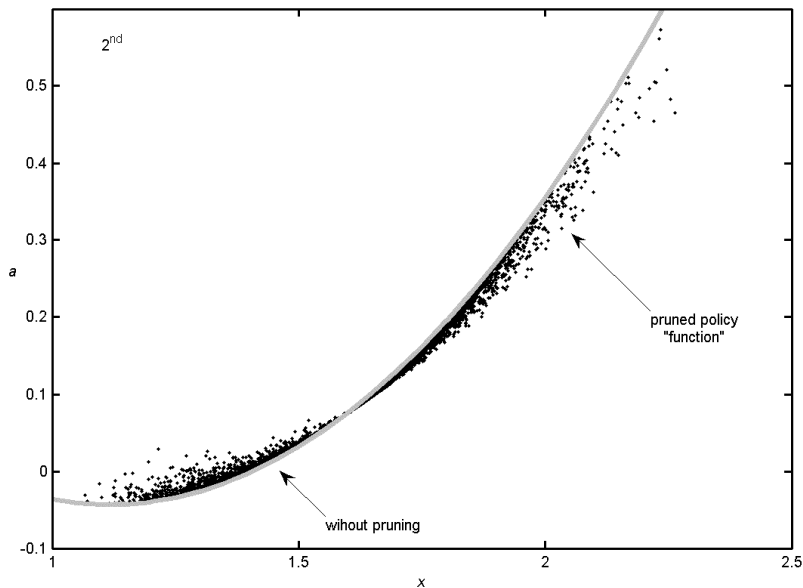


Pruning - graphs

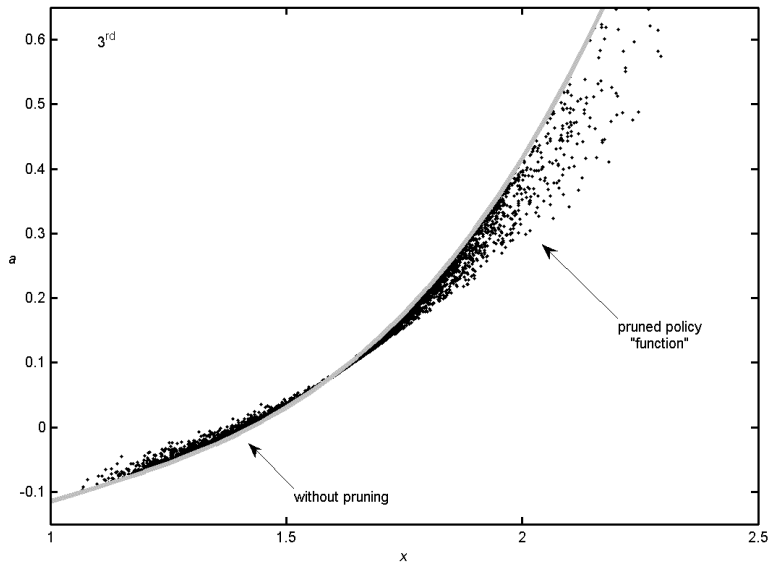
Our model only has one state variable, $x_t = a_{t-1} + \theta_t$

- Generate $\{a_{\text{prune},t}\}_{t=1}^T$
- plot $a_{\text{prune},t}$ as function of $x_{\text{prune},t} = a_{\text{prune},t} + \theta_t$

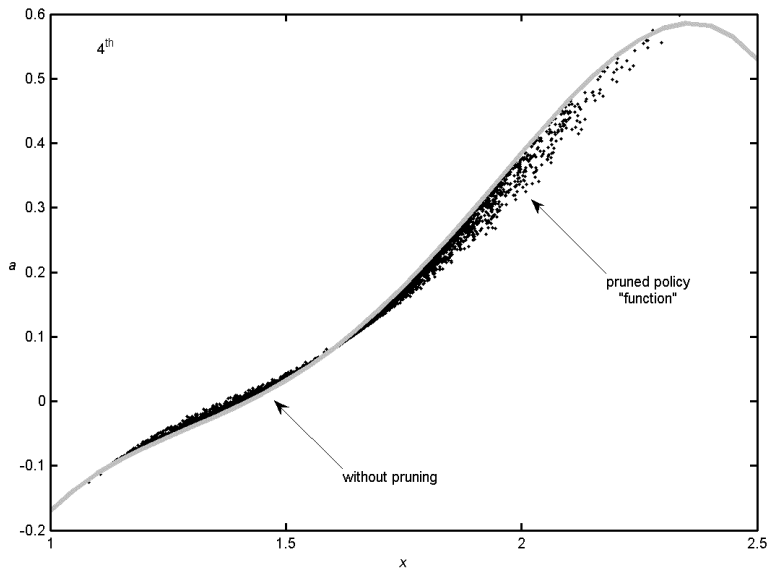
Pruning - second-order



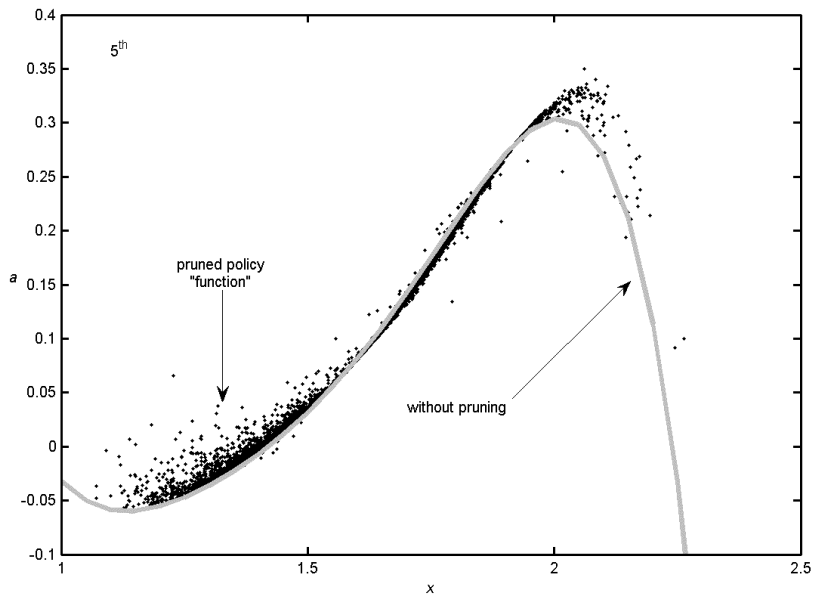
Pruning - third-order



Pruning - fourth-order



Pruning - fifth-order



Improvements

- simple improvements
- improvements based on alternative perturbation solutions

Measuring data

Data:

length of observed data set T_{nobs} :

observed data $y^{T_{\text{nobs}}} = \{y_{t,\text{data}}\}_{t=1}^{T_{\text{nobs}}} :$

moment of interest $M\left(y_i^{T_{\text{nobs}}}\right)$

Original Kydland and Prescott approach:

Model:

data generated in i^{th} replication $y_i^{T_{\text{nobs}}} = \{y_{t,i}\}_{t=1}^{T_{\text{nobs}}}$:

mean of moment of interest $\bar{M}_I = \frac{\sum_{i=1}^I M(y_i^{T_{\text{nobs}}})}{I}$

st. dev. of moment of interest $\frac{\sum_{i=1}^I (M(y_i^{T_{\text{nobs}}}) - \bar{M}_I)}{I}$

Most common approach

Model:

data generated in 1 replication $y_i^{T_{\text{large}}} = \{y_{t,i}\}_{t=1}^{T_{\text{large}}}$:

mean of moment of interest $M\left(y_i^{T_{\text{large}}}\right)$

st. dev. of moment of interest 0

Differences

- In general:

$$\lim_{T_{\text{large}} \rightarrow \infty} M\left(y_i^{T_{\text{large}}}\right) \neq \lim_{I \rightarrow \infty} \bar{M}_I$$

except for first-order moments

- KP approach deals with fact that small sample results may differ

Back to explosive perturbation solutions

- (perturbation) approximations explode \implies use KP instead of the T_{large} approach
- But sharply diverging behavior still possible
 - Solution: simply exclude those replications
 - Drawbacks:
 - need a criterion to exclude
 - need initial conditions

Exclusion criterion

- \bar{M}_I^{1st} : moment according to first-order perturbation solution
- Exclude i^{th} sample if

$$M\left(y_i^{T_{\text{noobs}}}\right) > \Lambda \bar{M}_I^{1st}$$

- We experimented with $\Lambda = 2, 3$

Initial conditions

- Ideally: initial conditions drawn from ergodic distribution
- One can approximate this using first-order solution (which is stable)

Understanding perturbation

Let

$$\begin{aligned}h(k) &= \text{truth} \\g(k; \gamma) &= \text{approximation}\end{aligned}$$

- Find coefficients γ such that

$$\left. \frac{\partial g^n(k; \gamma)}{\partial k^n} \right|_{x=\bar{x}} = \left. \frac{\partial h^n(k)}{\partial k^n} \right|_{x=\bar{x}} \quad \text{for } n = 0, 1, \dots, N$$

Understanding perturbation's flexibility

❶ You are not restricted to use polynomials

❷ Values of

$$\left. \frac{\partial g^n(k; \gamma)}{\partial k^n} \right|_{x=\bar{x}} \quad \text{for } n > N$$

are *not* restricted to be anything

Exploiting higher-order degrees of freedom

- Suppose you are given

$$h(\bar{k}), \frac{\partial h(\bar{k})}{\partial k}, \frac{\partial h^2(\bar{k})}{\partial k}$$

and consider

$$g(k; \eta) = \eta_0 + \eta_1(k - \bar{k}) + \eta_2(k - \bar{k})^2 + \eta_3(k - \bar{k})^3$$

- Standard perturbation

$$\eta_3 = 0$$

- But this is arbitrary
- Derivatives have no information on this
- You could use this additional degree of freedom to implement another desired property

Exploit functional form flexibility

- Suppose you are given

$$\left. \frac{\partial h^n(k)}{\partial k^n} \right|_{x=\bar{x}} \quad \text{for } n = 0, 1, \dots, N$$

- You would like to use

$$g(k; \eta) = \eta_0 g_0(k) + \eta_1 g_1(k) + \dots + \eta_N g_N(k)$$

- Solve for the values of a from the following $N + 1$ equations

$$\left. \frac{\partial h^n(k)}{\partial k^n} \right|_{k=\bar{k}} = [\eta_0, \eta_1, \dots, \eta_N] \begin{bmatrix} \left. \frac{\partial g_0^n(k)}{\partial k^n} \right|_{k=\bar{k}} \\ \vdots \\ \left. \frac{\partial g_N^n(k)}{\partial k^n} \right|_{k=\bar{k}} \end{bmatrix}$$

Simple example

$$1/x$$

- Fourth-order Taylor series expansion

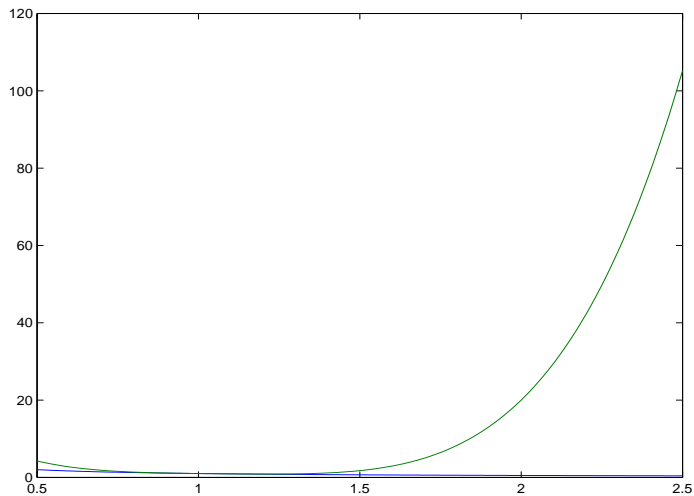
$$1/x \approx 1 - (x - 1) + 2(x - 1)^2 - 6(x - 1)^3 + 24(x - 1)^4$$

- Alternative

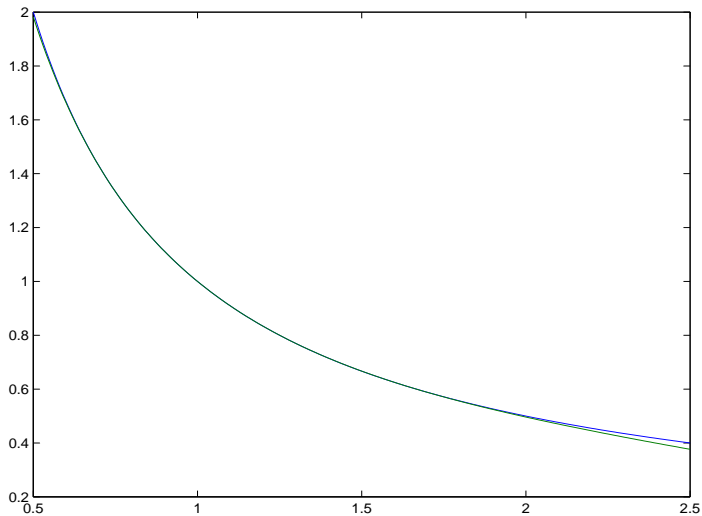
$$1/x \approx \eta_0 e^{-2x} + \eta_1 e^{-2x} x + \eta_2 e^{-2x} x^2 + \eta_3 e^{-2x} x^3 + \eta_4 e^{-2x} x^4$$

- note that this is not a transformation

Standard Taylor expansion



Alternative Taylor expansion



Generate stable perturbation solutions

- ① Use alternative basis functions
 - trivial modification for 2nd-order perturbation
- ② Use a *perturbation-consistent* weighted combination

Alternative basis functions

- Original model:

$$F(k_{-1}, k, k_{+1}) \equiv 0$$

$$F(k_{-1}, h(k_{-1}), h(h(k_{-1}))) \equiv 0$$

- From (say) Dynare you get

$$g(k; \eta) = \eta_0 + \eta_1(k - \bar{k}) + \eta_2(k - \bar{k})^2$$

Alternative basis functions

- Instead of $g(k; \eta)$ use $\tilde{g}(k; \eta)$

$$\tilde{g}(k; \tilde{\eta}) = \tilde{\eta}_0 + \tilde{\eta}_1 (k - \bar{k}) + \tilde{\eta}_2 (k - \bar{k})^2 \exp\left(- (k - \bar{k})^2\right)$$

- Globally stable for $|\tilde{\eta}_1| < 1$

Alternative basis functions

- Implementing perturbation principle: solve $\tilde{\eta}$ from

$$\tilde{g}(\bar{k}; \tilde{\eta}) = h(\bar{k})$$

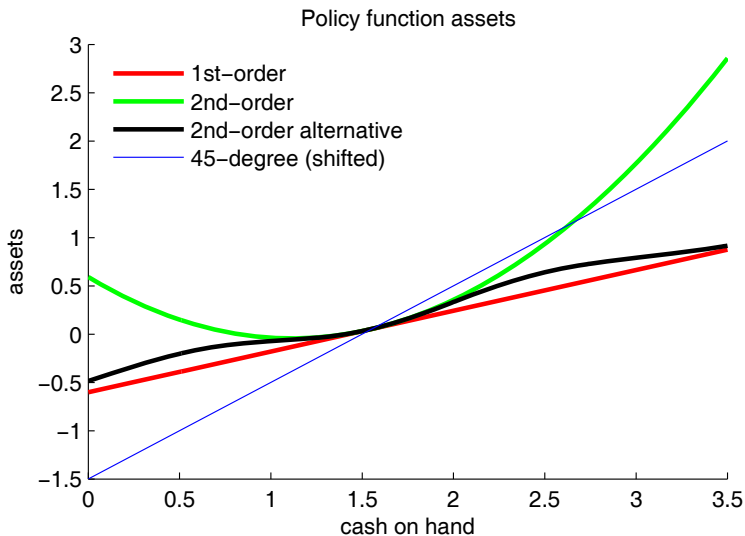
$$\frac{\partial \tilde{g}(\bar{k}; \tilde{\eta})}{\partial \bar{k}} = \frac{\partial h(\bar{k})}{\partial \bar{k}}$$

$$\frac{\partial^2 \tilde{g}(\bar{k}; \tilde{\eta})}{\partial \bar{k}^2} = \frac{\partial^2 h(\bar{k})}{\partial \bar{k}^2}$$

- Amazing but true:

$$\eta = \tilde{\eta}$$

Alternative basis functions



Alternative basis functions

- How to remain closer to underlying second-order perturbation?
- Use

$$\exp\left(-\alpha (k - \bar{k})^2\right)$$

and choose low value of α

Perturbation consistent weighting

- Original model:

$$F(k_{-1}, k, k_{+1}) \equiv 0$$

- add new variable y and new equation

$$k = \begin{matrix} y & \times & \exp\{-\alpha(k_{-1} - \bar{k})^2\} \\ \eta_{1^{st},0} + \eta_{1^{st},1}k_{-1} & \times & 1 - \exp\{-\alpha(k_{-1} - \bar{k})^2\} \end{matrix}$$

- α controls speed of convergence towards stable part

Perturbation consistent weighting

- Solve for perturbation solutions of $h_k(k_{-1})$ and $h_y(k_{-1})$
- Do *not* use $h_k(k_{-1})$, but use

$$k = \tilde{h}_k(k_{-1}) = \eta_{1^{\text{st}},0} + \eta_{1^{\text{st}},1}k_{-1} \times \frac{h_y(k_{-1}) \times \exp\{-\alpha(k_{-1} - \bar{k})^2\}}{1 - \exp\{-\alpha(k_{-1} - \bar{k})^2\}}$$

Perturbation consistent weighting

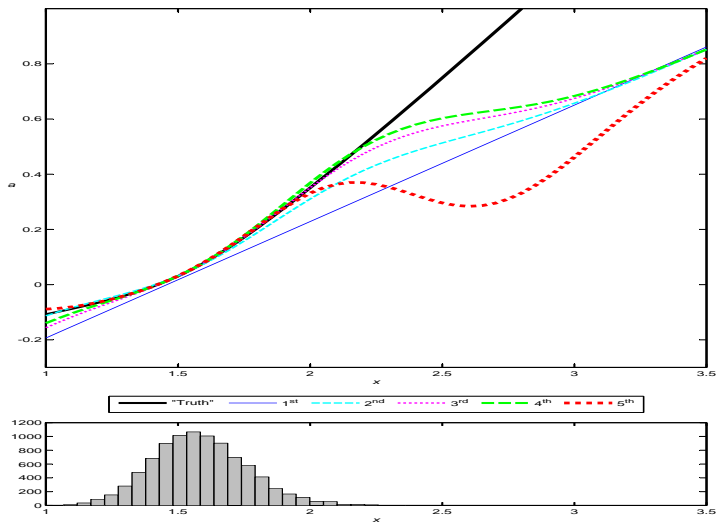
- Approximation is a *function* not a correspondence
- Derivatives of $h_y(k_{-1})$ correspond to true derivatives at $\bar{k} \implies$
- Derivatives of $\tilde{h}_k(k_{-1})$ correspond to true derivatives at \bar{k}
- and $k = \tilde{h}_k(k_{-1})$ is globally stable

Note the difference with

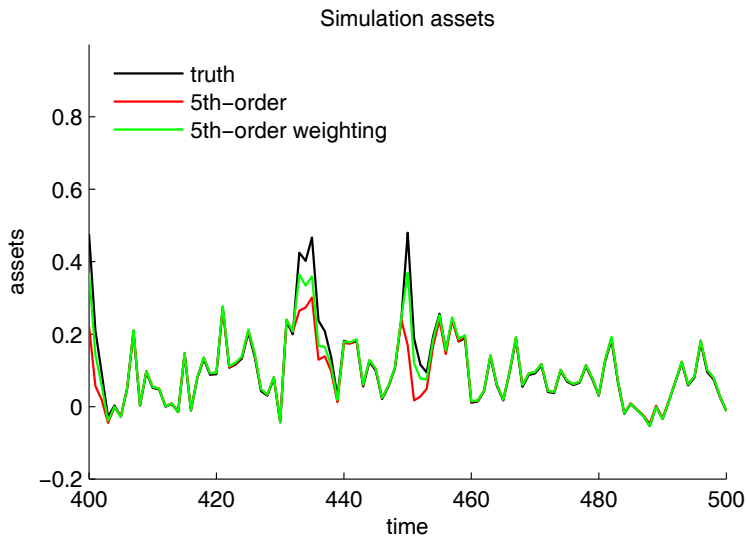
$$k = \hat{h}_k(k_{-1}) = \frac{p_{k^{\text{th}}}(k_{-1})}{\eta_{1^{\text{st}},0} + \eta_{1^{\text{st}},1}k_{-1}} \times \frac{\exp\{-\alpha(k_{-1} - \bar{k})^2\}}{1 - \exp\{-\alpha(k_{-1} - \bar{k})^2\}}$$

- Derivatives of $\hat{h}_k(k_{-1})$ are *not* correct derivatives of $h(k_{-1})$

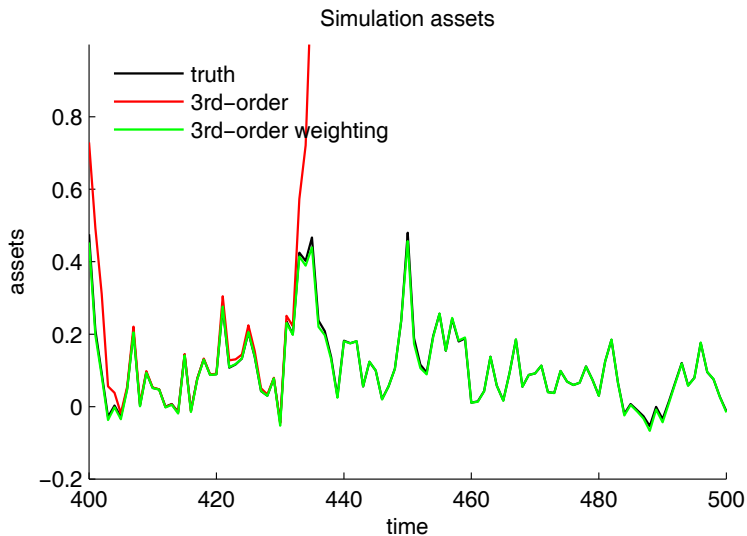
Perturbation consistent weighting



Perturbation consistent weighting



Perturbation consistent weighting

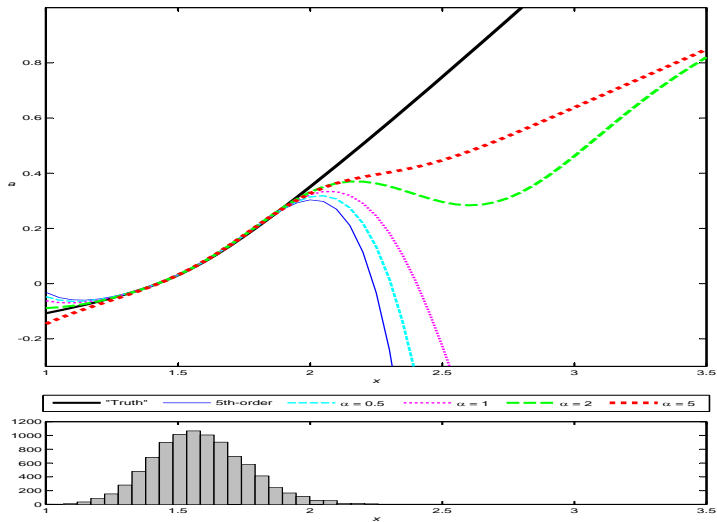


How to choose α ?

How to choose α ?

- Not that difficult if you can plot the policy function
- Make estimated guess
 - e.g., 3 standard deviations away from \bar{s} , weight on first-order should be 0.28
- Try different values for α and use accuracy test (e.g. dynamic Euler equation test)

Perturbation consistent weighting



Multi-dimensional problems

- Let s be the $N \times 1$ vector of state variables
- Solve first-order solution: $k = a_{1^{st},0} + a'_{1^{st},1}s$
- Calculate Ω , the variance covariance matrix of s_t

How to choose alpha?

Use

$$k = \frac{h_y(x)}{a_{1^{\text{st}},0} + a_{1^{\text{st}},1}s} \times \frac{\exp \left\{ -\frac{\alpha}{N} (s_{-1} - \bar{s})' \Omega^{-1} (s_{-1} - \bar{s}) \right\}}{1 - \exp \left\{ -\frac{\alpha}{N} (s_{-1} - \bar{s})' \Omega^{-1} (s_{-1} - \bar{s}) \right\}}$$

or

$$k = \frac{h_y(x)}{a_{k^{\text{th}},0} + a_{k^{\text{th}},1}s} \times \frac{\exp \left\{ -\frac{\alpha}{N} (s_{-1} - \bar{s}_{k^{\text{th}}})' \Omega^{-1} (s_{-1} - \bar{s}_{k^{\text{th}}}) \right\}}{1 - \exp \left\{ -\frac{\alpha}{N} (s_{-1} - \bar{s}_{k^{\text{th}}})' \Omega^{-1} (s_{-1} - \bar{s}_{k^{\text{th}}}) \right\}}$$

Multidimensional problems

- Try different values for α and use accuracy test
 - e.g. dynamic Euler equation test

Penalty functions

- to *approximate* inequality constraint
- to *describe* feature in actual economy

Overview

- Example
- How to choose parameters
- Different from inequality constraint?
- Blanchard-Kahn conditions
- Functional form
 - try to get them analytic
 - stay in space of perturbation approximation

Example

$$P(a) = \frac{\eta_1}{\eta_0} \exp(-\eta_0 a) - \eta_2 a.$$

Calibrating the penalty function

- η_0 , η_1 , and η_2 can be chosen to match data characteristics
 - η_0 clearly a key parameter

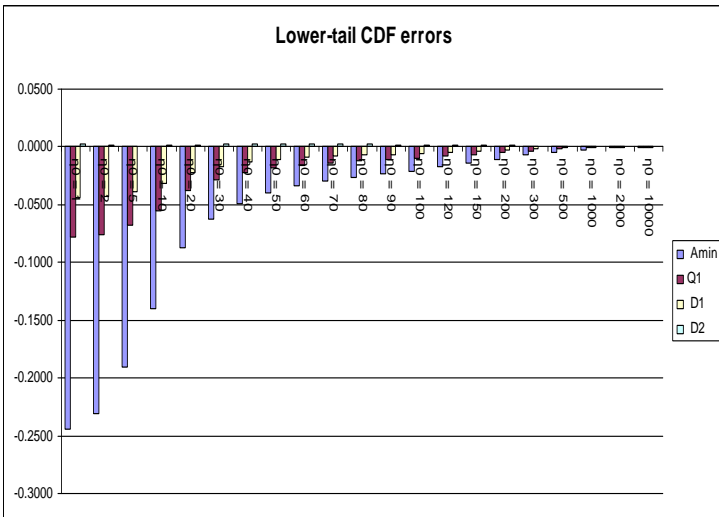
Penalty versus inequality

- different values for curvature parameter, η_0
 - η_1 and η_2 chosen to match mean and standard deviation of a_t
 - \implies these two properties "correct"
 - how different is tail behavior when no numerical errors are made?

Lower tail

We look at

- A_{\min} : minimum value of A attained
- Q1: first quintile
- D1: first decile
- D2: second decile



First-order condition

$$\frac{c_t^{-\nu}}{1+r} + \eta_1 \exp(-\eta_0 a) - \eta_2 = \beta E_t [c_{t+1}^{-\nu}]$$

Suppose there is no penalty function

Eigenvalues

$$\begin{aligned}\lambda_+ &= 1 + r \\ \lambda_- &= \frac{1}{(1+r)\beta}\end{aligned}$$

typical impatience assumption:

$$\beta < \frac{1}{1+r}$$

\implies BK conditions not satisfied

How to satisfy Blanchard-Kahn conditions?

- Put in penalty function
- Will Blanchard-Kahn condition be satisfied?
 - possibly not for high value of η_0
 - penalty term too flat at high η_0 values

How to satisfy Blanchard-Kahn conditions?

- Are local dynamics necessarily unstable for high η_0 ?
- NO
 - with uncertainty:
 - higher-order perturbation change first-order term
- How to implement this with Dynare?

Functional forms used

- Preston and Roca (2007)

$$P(a) = \frac{\eta}{(a - \bar{a})^2}$$

- Kim, Kollmann, and Kim (2010)

$$\eta \left(\ln \frac{a}{a_{SS}} - \frac{a - a_{SS}}{a_{SS}} \right)$$

- Drawback of both:
 - not analytic

Functional forms used

- Den Haan and De Wind (2010)

$$P(a) = \frac{\eta_1}{\eta_0} \exp(-\eta_0 a) - \eta_2 a$$

- Advantage
 - analytic
- Drawback
 - not clear how perturbation solution will behave

Possible fix

- Suppose you use second-order approximation
- Let $P(a)$ be such that
 - $\frac{\partial P(a)}{\partial a} = \eta_0 + \eta_1 a + \eta_2 a^2$
 - problematic behavior far enough away from steady state