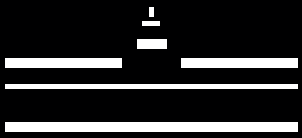


Westfälische
Wilhelms-Universität
Münster

Local identification of nonlinear and non-Gaussian DSGE models

Willi Mutschler

Westfälische
Wilhelms-Universität
Münster

Local identification of nonlinear and non-Gaussian DSGE models

Inauguraldissertation
zur Erlangung des akademischen Grades eines
Doktors der Wirtschaftswissenschaften durch die
Wirtschaftswissenschaftliche Fakultät der
Westfälischen Wilhelms-Universität Münster

vorgelegt von
Willi Mutschler

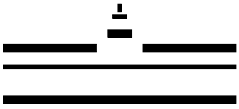
– Münster, 2015 –

Dekanin	Prof. Dr. Theresia Theurl
Erstgutachter	Prof. Dr. Mark Trede
Zweitgutachter	Prof. Dr. Bernd Kempa
Mündliche Prüfung	19.11.2015

Willi Mutschler

Local identification of nonlinear and non-Gaussian DSGE models

MV WISSENSCHAFT



Wissenschaftliche Schriften der WWU Münster

Reihe IV

Band 10

Willi Mutschler

Local identification of nonlinear and non-Gaussian DSGE models

MV WISSENSCHAFT

Wissenschaftliche Schriften der WWU Münster
herausgegeben von der Universitäts- und Landesbibliothek Münster
http://www.ulb.uni-muenster.de

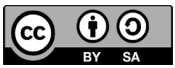
Bibliografische Information der Deutschen Nationalbibliothek:
Die Deutsche Nationalbibliothek verzeichnet diese Publikation in der Deutschen Nationalbibliografie; detaillierte bibliografische Daten sind im Internet über http://dnb.d-nb.de abrufbar.

Dieses Buch steht gleichzeitig in einer elektronischen Version über den Publikations- und Archivierungsserver der WWU Münster zur Verfügung.
http://www.ulb.uni-muenster.de/wissenschaftliche-schriften

Willi Mutschler
„Local identification of nonlinear and non-Gaussian DSGE models“
Wissenschaftliche Schriften der WWU Münster, Reihe IV, Band 10

© 2016 der vorliegenden Ausgabe:
Die Reihe „Wissenschaftliche Schriften der WWU Münster“ erscheint im Verlagshaus Monsenstein und Vannerdat OHG Münster
www.mv-wissenschaft.com

Dieses Werk ist unter der Creative-Commons-Lizenz vom Typ 'CC BY-SA 3.0 DE' lizenziert: http://creativecommons.org/licenses/by-sa/3.0/de/
Von dieser Lizenz ausgenommen sind Abbildungen, welche sich nicht im Besitz des Autors/ der Autorin oder der ULB Münster befinden.



ISBN 978-3-8405-0135-7 (Druckausgabe)
URN urn:nbn:de:hbz:6-97219489383 (elektronische Version)

direkt zur Online-Version:

© 2016 Willi Mutschler
Alle Rechte vorbehalten

Satz: Willi Mutschler
Umschlag: MV-Verlag
Druck und Bindung: MV-Verlag

Für meine Familie

Contents

List of Figures

List of Tables

Chapter 1

Introduction

Many different methods of solving and estimating DSGE models have been developed and used in order to obtain a detailed analysis and thorough estimation of dynamic macroeconomic relationships. Recently, the question of identifying DSGE models has proven to be of major importance, especially since the identification of a model precedes estimation and inference. Identification problems arise if distinct parameter values do not lead to distinct probability distributions of data. Even with an infinite sample it is not possible to pin down some (sets of) parameters, no matter what estimation procedure one uses. In a full-information setting this often evokes a badly shaped likelihood function, which modern Bayesian estimation can conveniently circumvent by using tight priors. However, in the case of prior dependence the comparison of prior and posterior for non-identified parameters can be misleading (Poirier, 1998). Moreover, calibrating unidentified parameters can lead to wrong conclusions, since other parameters might depend on the calibrated ones (Canova and Sala, 2009). Lack of identification leads thus to wrong conclusions from calibration, estimation and inference, whereas the source of identification influences empirical findings (Ríos-Rull et al., 2012).

Identification has a long history in econometrics. Research into the parameter identification problem dates back to the beginnings of the Cowles Comission, see e.g. Monographs No. 10 and No. 14 (Koopmans, 1950; Koopmans and Hood, 1953). In fact, the problem of identification is often associated with the birth of econometrics owing to major contributions from Frisch (1934, Ch. 11), Haavelmo (1944, Ch. V) and Koopmans (1949). Since then it has gained importance in almost all fields of econometrics.[1] The first known approach to analyze the identification problem focused on "*the extent to which statistically determined demand curves may be identified with theoretical demand curves*" (Working, 1925, p. 526). The basic problem is that the demand as well as the supply curve are subject to shifts and it is in general not possible to determine the true demand curve from observational evidence.

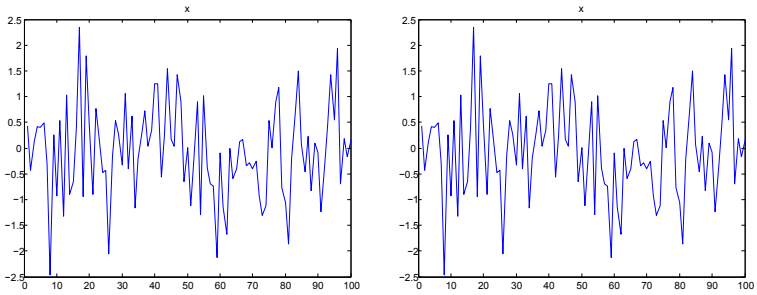
> *From this, it would seem that, whether we obtain a demand curve or a supply curve, by fitting a curve to a series of points which represent the quantities of an article sold at various prices, depends upon the fundamental nature of the supply and demand conditions. It implies the need of some term in addition to that of elasticity in order to describe the nature of supply and demand. The term "variability" may be used for this purpose. (Working, 1927, p. 224)*

Thus, a consistent estimation of the demand curve requires either (i) the inclusion of truly exogenous variables or (ii) the relative shift of the supply curve to be greater than the shift of the demand curve. Wright (1928) gives a rigorous analysis of the first condition and basically formulates methods known later as *indirect inference* and *instrumental variables estimation*.[2] Moreover, the second point emphasizes the need for additional restrictions in order to identify the true structure of the demand curve from observational data.

[1] For a good textbook overview of the early contributions see Fisher (1966), Hsiao (1983) and the references therein.

[2] See also Stock and Trebbi (2003) for a historical classification.

Figure 1.1: ARMA(1,1) example



$\theta_0 = (0.4, 0.4, 1)'$ $\theta_1 = (0, 0, 1)'$

Identifiability is a model property and it can be analyzed without actually estimating the model. I will demonstrate this by means of a simple ARMA(1,1) model

$$x_t = \phi_1 x_{t-1} + \varepsilon_t - \phi_2 \varepsilon_{t-1}$$

with $\varepsilon_t \sim N(0, \sigma^2)$ and parameter vector $\theta = (\phi_1, \phi_2, \sigma)'$. Figure 1.1 simulates data for two different parameterizations with $\theta_0 = (0.4, 0.4, 1)'$ and $\theta_1 = (0, 0, 1)'$. Note that in both cases $\phi_1 = \phi_2$. Obviously, both parametrizations generate the same data (as long as the shocks ε_t and starting value are the same). The underlying structure, i.e. whether θ_0 or θ_1 are the true parameters, cannot be identified from observed data no matter which estimation method one uses. Since x_t is normally distributed and covariance-stationary, all time series properties are completely characterized by its unconditional first two moments (or equivalently by the mean and power spectrum). Let $\gamma_h := E(x_t x_{t-h})$ denote the autocovariance function, that is

$$\gamma_0 = \frac{(1 + \phi_2^2 - 2\phi_1\phi_2)\sigma^2}{1 - \phi_1^2}, \gamma_1 = \frac{(\phi_1 - \phi_2)(1 - \phi_1\phi_2)\sigma^2}{1 - \phi_1^2}, \gamma_h = \phi_1 \gamma_{h-1}$$

For $\phi_1 = \phi_2$ this simplifies to $\gamma_0 = \sigma^2, \gamma_k = 0, k \geq 1$, which is identical to the white-noise process given by θ_1. In other words, θ_0 and θ_1 are observational equivalent structures if and only if $\phi_1 = \phi_2$. More formally, consider the mapping from θ to $\Gamma := (\gamma_0, \gamma_1, \gamma_2, \dots)$. This mapping is unique if and only if the Jacobian of Γ, denoted with $\mathcal{D}\Gamma$, has full column rank when evaluated at θ_0 or θ_1. Specifically, for $\phi_1 = \phi_2 = \phi$ $\mathcal{D}\Gamma$ is equal to

$$\mathcal{D}\Gamma = \begin{pmatrix} -2\phi\sigma^2 & 0 & -2\sigma(\phi^2 - 1) \\ \sigma^2 & -\sigma^2 & 0 \\ \phi\sigma^2 & -\phi\sigma^2 & 0 \\ \phi^2\sigma^2 & -\phi^2\sigma^2 & 0 \\ \phi^3\sigma^2 & -\phi^3\sigma^2 & 0 \\ \vdots & \vdots & \vdots \end{pmatrix}$$

Obviously, there are only two linearly independent rows, the rank is equal to two. Analyzing the nullspace one can then pinpoint the problematic parameters, i.e. the set (θ_1, θ_2) is indistinguishable, whereas σ is partially identifiable. This example demonstrates the basic idea behind the rank conditions in the classical literature. Intuitively, one asks, *if and how* theoretical moments change, if the vector of deep parameters changes. To this end, one defines mappings and derives a rank condition for the corresponding Jacobian.

Accordingly, identification criteria in the DSGE literature are concerned with two mappings: one from the deep parameters to the reduced-form parameters, i.e. the uniqueness of the solution, and one from the solution to observable data, i.e. the uniqueness of the probability distribution. An adaption of classical concepts is not as straightforward, since the reduced-form parameters are nonlinear functions of the deep parameters. Moreover, there might be complicated cross-equation restrictions, which can often only be evaluated numerically, a point emphasized by Komunjer and Ng (2011). Recently, several formal methods have been proposed to check lo-

cal identification in linearized DSGE models via (i) observational equivalent first and second moments (Iskrev, 2010a), (ii) observational equivalent spectral densities (Qu and Tkachenko, 2012), (iii) implications from control theory for observational equivalent minimal systems (Komunjer and Ng, 2011) and (iv) Bayesian indicators (Koop et al., 2013). The first three approaches derive rank conditions similar to the classical literature on identification, whereas the fourth approach uses insights from Bayesian asymptotic theory. The rank criteria all stem essentially from Rothenberg (1971), who proposes identifiability conditions based on injectivity of functions. The mappings to consider are the unconditional mean, autocovariogram and corresponding spectral density. Therefore, the fundamental idea is to determine whether these mappings are unique for the deep parameters. Basic mathematical results for systems of equations can then be applied. This set of criteria is the most basic and the closest to ideas from the early work on identification in systems of linear equations, since it is based on the uniqueness of a solution (Koopmans and Reiersøl, 1950; Fisher, 1966; Hannan, 1976). Consequently, rank and order conditions are derived, and it is also possible to pinpoint the (sets of) parameters that are indistinguishable from one another. Even though all methods seem similar, there has been – to my knowledge – no study of the advantages and drawbacks of implementing the different methods. Furthermore, whereas there is a growing literature on the estimation of nonlinear DSGE models (Andreasen, 2011, 2013; Fernández-Villaverde and Rubio-Ramírez, 2007; Herbst and Schorfheide, 2014; Ivashchenko, 2014; Kollmann, 2015), all identification methods focus on the linearized DSGE model with Gaussian innovations. Accordingly, the proposed criteria are based on first and second moments only. If, however, one relaxes the assumption of linearity or Gaussianity, it is natural to analyze whether it is possible to exploit information from higher-order moments for the calibration, estimation and iden-

tification of parameters. Researchers in mathematics, statistics and signal processing have developed tools, called higher-order statistics (HOS), to solve detection, estimation and identification problems when the noise source is non-Gaussian or one is faced with nonlinearities. The basic tools of HOS are cumulants, which are defined as the coefficients in the Taylor expansion of the log moment generating function in the time domain; and polyspectra, which are defined as Fourier transformations of the cumulants in the frequency domain. Applications in the macroeconometric literature are rather sparse, in the DSGE literature they are – so far – nonexistent.

In this thesis, I therefore relax the assumptions of linearity and Gaussianity and show how to check rank criteria for nonlinear DSGE models solved by higher-order approximations of the policy functions and by considering higher-order statistics for non-Gaussian innovations. However, there is a caveat, since higher-order approximations may yield explosive or non-stationary processes. Therefore, I use the pruning scheme proposed by Kim et al. (2008) and operationalized by Andreasen et al. (2014). They show that the pruned state-space is a linear, stationary and ergodic system, but with non-Gaussian innovations. Similar to the identification of demand curves, this may yield additional restrictions on unconditional moments and polyspectra of the observables that can be used to identify (sets of) parameters which are not identifiable in the linearized DSGE model with Gaussian innovations.

The thesis is structured as follows. Chapter 2 sets up the general DSGE framework, discusses linear and nonlinear solution methods and derives the (pruned) state-space system. In the derivations, I limit myself to Taylor approximations and pruning up to second-order, since an extension beyond second-order is – apart from tedious notation – straightforward. To make the exposition of the thesis illustrative, Chapter 3 outlines two example models, namely those of Kim (2003) and An and Schorfheide (2007), which are

known to have a lack of identification in their linearized Gaussian versions. I derive both models theoretically and show how to squeeze them into the general DSGE framework considered in this thesis. In the following chapters, extensive use of both models is made to demonstrate the theoretical results and procedures. In Chapter 4, I derive closed-form expressions for unconditional moments, cumulants and polyspectra of order higher than two for linear and nonlinear (pruned) DSGE models. Third-order cumulants and the bispectrum capture nonlinearities (or non-Gaussianity) for a skewed process, whereas the fourth-order cumulants and the trispectrum can be used in the case of a non-Gaussian symmetric probability distribution. Therefore, I limit myself to these higher-order statistics in the theoretical derivations. The procedures are then demonstrated by means of the An and Schorfheide (2007) model. To this end, I compare the theoretical expressions for skewness and excess kurtosis with their empirical counterparts within a Monte-Carlo framework, since these are typical measures an applied researcher would like to match in a calibration exercise. On the other hand, auto- and cross-skewness as well as kurtosis may contain valuable information in an estimation exercise. Both, the Gaussian as well as Student's t-distribution are considered as the underlying stochastic process. Useful matrix tools and computational aspects are also discussed. In the following Chapter 5, I then show how to analytically calculate the Jacobians of unconditional second-, third- and fourth-order moments, cumulants and corresponding polyspectra with respect to the deep parameters of the DSGE model. In this manner, I am able to derive formal identification criteria based on theoretical higher-order moments for linearized DSGE models in Chapter 7 and for nonlinear DSGE models in Chapter 8. But first, the underlying assumptions and definitions of local identifiability are stated formally in Chapter 6. Chapter 7 then reviews and compares the formal rank criteria for linearized DSGE models as well

as the Bayesian indicator. Furthermore, I extend ideas from Iskrev (2010a) and Qu and Tkachenko (2012) to establish rank criteria for higher-order statistics, both in the time and in the frequency domains. This approach is then adapted to nonlinear DSGE models in Chapter 8. All methods are demonstrated by means of the two example models. In particular, I show that the parameters governing the investment adjustment costs in Kim (2003) are not separately identifiable in the linearized DSGE model. However, they can be identified from the mean and second moments or power spectrum as well as from higher-order statistics of the pruned state-space given a second-order approximation. A similar finding holds for the An and Schorfheide (2007) model. In the linearized model, the elasticity of demand, price stickiness and steady state of government spending are not identifiable. Moreover, using an output-gap specification for the monetary policy rule makes the parameters of the Taylor rule jointly unidentified. However, in case of an output-growth rule these parameters can be identified. A second-order approximation yields even more restrictions. All parameters of the model can be identified from the mean and second moments or power spectrum, as well as from higher-order statistics, no matter which specification one uses for the Taylor rule. Lastly, Chapter 9 summarizes the contributions of the thesis. A nonlinear approach enriches model dynamics, estimability and identifiability.

Chapter 2

DSGE framework

2.1 General model and solution

Let E_t be the expectation operator conditional on information available at time t, then

$$E_t f\left(x_{t+1}, u_{t+1}, y_{t+1}, x_t, u_t, y_t|\theta\right) = 0 \tag{2.1}$$

$$x_{t+1} = h(x_t, u_{t+1}, \sigma|\theta) \tag{2.2}$$

$$y_{t+1} = g(x_t, u_{t+1}, \sigma|\theta) \tag{2.3}$$

is called the general DSGE model with deep parameters θ, states x_t, controls y_t, stochastic innovations u_t, and perturbation parameter σ, which can be converted into a nonlinear first-order system of expectational difference equations f. For the sake of notation, it is assumed that all control variables are observable. The vector of innovations u_t has $E(u_t) = 0$ and finite covariance matrix $E(u_t u_t') =: \Sigma = \sigma^2 \eta\eta'$. Thus, σ is set to be dependent on the standard deviation of one of the shocks, while scaling all other variances and cross-correlations through η accordingly. Furthermore, u_t is nth-order white noise with finite higher-order moments, where n

depends on the order of approximation.[3] Apart from the existence of moments and temporal independence no distributional assumptions are needed.[4] See Chapter 3 on how to squeeze the example models into this framework.[5]

The solution of such rational expectation models is characterized by *policy functions*, g and h, that solve (at least approximately) the system of equations f. Following Schmitt-Grohé and Uribe (2004) I use perturbation techniques to solve the model around the non-stochastic steady state given by $\bar{x} = h(\bar{x}, 0, 0|\theta)$, $\bar{y} = g(\bar{x}, 0, 0|\theta)$, $\bar{u} = 0$ and $\bar{f} = f(\bar{x}, \bar{u}, \bar{y}|\theta) = 0$. Moreover, I exploit ideas of Gomme and Klein (2011) to approximate the policy functions, using the Magnus and Neudecker (1999) definition of the Hessian. Denote the Jacobian of f evaluated at the steady state as

$$\mathcal{D}\bar{f} := \left(\frac{\partial \bar{f}}{\partial x'_{t+1}}, \frac{\partial \bar{f}}{\partial u'_{t+1}}, \frac{\partial \bar{f}}{\partial y'_{t+1}}, \frac{\partial \bar{f}}{\partial x'_t}, \frac{\partial \bar{f}}{\partial u'_t}, \frac{\partial \bar{f}}{\partial y'_t} \right), \tag{2.4}$$

[3] Because the proposed identification criteria in Chapters 7 and 8 are based on the first four moments of observables, the stochastic innovations are required to have at least finite fourth moments for a first-order approximation, finite eighth moments for a second-order approximation and finite twelfth moments for a third-order approximation. In other words, u_t is at least an fourth-, eighth- or twelfth-order white noise process, which implies y_t being stationary of order four, see Priestley (1983, p. 105) for a definition of stationary up to order n.

[4] The MATLAB code can handle both the Gaussian, as well as Student's t-distribution provided the moments exist.

[5] This is basically a mixture of the DYNARE framework (innovations enter nonlinearly, no distinction between states and controls) and the framework of Schmitt-Grohé and Uribe (2004) (innovations enter linearly, distinction between states and controls). It can be shown that both frameworks are equivalent, given an extended state vector, see the technical appendix in Andreasen et al. (2014). In the same fashion, it is possible to add measurement equations and measurement errors by simply extending the model equations, state and control variables accordingly. A selection matrix can be premultiplied to consider only a subset of controls as observables, see the An and Schorfheide (2007) model in Chapter 3.2 for an example.

then $\mathcal{H}\bar{f} := \mathcal{D}vec([\mathcal{D}\bar{f}]')$ is defined as the Magnus-Neudecker Hessian of f, evaluated at the nonstochastic steady state. This definition simplifies the computations, as well as the analytical derivatives, since no tensor notation is needed and basic matrix algebra can be used.[6]

The approximations of the policy functions are a straightforward application of Taylor series expansions in the state variables:

$$\begin{aligned}\hat{x}_{t+1} &= h_x\hat{x}_t + h_u u_{t+1} \\ &\quad + \frac{1}{2}\left[H_{xx}\left(\hat{x}_t \otimes \hat{x}_t\right) + H_{uu}\left(u_{t+1} \otimes u_{t+1}\right)\right] \\ &\quad + \frac{1}{2}\left[H_{ux}\left(u_{t+1} \otimes \hat{x}_t\right) + H_{xu}\left(\hat{x}_t \otimes u_{t+1}\right) + h_{\sigma\sigma}\sigma^2\right]\end{aligned} \tag{2.5}$$

$$\begin{aligned}\hat{y}_{t+1} &= g_x\hat{x}_t + g_u u_{t+1} \\ &\quad + \frac{1}{2}\left[G_{xx}\left(\hat{x}_t \otimes \hat{x}_t\right) + G_{uu}\left(u_{t+1} \otimes u_{t+1}\right)\right] \\ &\quad + \frac{1}{2}\left[G_{ux}\left(u_{t+1} \otimes \hat{x}_t\right) + G_{xu}\left(\hat{x}_t \otimes u_{t+1}\right) + g_{\sigma\sigma}\sigma^2\right]\end{aligned} \tag{2.6}$$

$\hat{x}_t = x_t - \bar{x}$ and $\hat{y}_t = y_t - \bar{y}$ denote deviations from steady state. h_x, h_u, g_x and g_u denote the solution matrices of the first-order approximation. H_{xx} is a $n_x \times n_x^2$ matrix containing all second-order terms for the ith state variable in the ith row, whereas G_{xx} is a $n_y \times n_x^2$ matrix containing all second-order terms for the ith control variable in the ith row. H_{xu}, H_{ux}, G_{xu} and G_{ux} are accordingly shaped for the cross terms of states and shocks, and H_{uu} and G_{uu} contain the second-order terms for the product of shocks. Lastly, $h_{\sigma\sigma}$ and $g_{\sigma\sigma}$ are the Hessians of h and g with respect to the perturbation parameter σ. Note that all matrices are evaluated at the nonstochastic steady state.

[6] For recent literature in favor of this definition, see Magnus (2010) and Pollock (2013). See Appendix A for additional material which clarifies the concept and notation.

There are several methods and algorithms for calculating the first-order solution matrices, since these are the coefficients of a first-order linearization or log-linearization of the model. I follow Klein (2000) to obtain the first-order solution using the generalized Schur decomposition.[7] The second-order solution matrices can be calculated by inserting the policy functions (2.2) and (2.3) into the model equations (2.1) and noting that the expression is known at the nonstochastic steady state. Therefore, all derivatives of f must be 0 when evaluated at the nonstochastic steady state. Differentiating f twice using the chain rule of Magnus and Neudecker (1999, p. 110), evaluating the Jacobian $\mathcal{D}\bar{f}$ and Hessian $\mathcal{H}\bar{f}$ of f at the nonstochastic steady state, and setting it to zero yields (after some algebraic calculations, see Gomme and Klein (2011)):

$$\begin{bmatrix} vec(g_{vv}) \\ vec(h_{vv}) \end{bmatrix} = -Q^{-1} vec(R), \qquad \begin{pmatrix} h_{\sigma\sigma} \\ g_{\sigma\sigma} \end{pmatrix} = -S^{-1}U. \tag{2.7}$$

g_{vv} and h_{vv} are the Magnus-Neudecker Hessians of g and h with respect to the vector $v_{t|t+1} = (x_t', u_{t+1}')'$, $g_{\sigma\sigma}$ and $h_{\sigma\sigma}$ are the corresponding terms with respect to the perturbation parameter. Note that all terms are evaluated at the nonstochastic steady state. See Appendix B for the exact expressions of the auxiliary matrices Q, R, S and U. Furthermore, h_{vv} contains all second-order terms required for the transition equation of states and g_{vv} for the transition equation of controls. I separate these using index matrices to set up equations (2.5) and (2.6), see Appendix C for an example that clarifies the procedure. A third-order approximation using Magnus-Neudecker Hessians is given in Binning (2013). For the sake

[7] Anderson (2008) provides a comparison of algorithms, which are basically all equivalent and differ only (slightly) in computational burden. Furthermore, all check the Blanchard and Kahn (1980) conditions that are necessary in order to have a stable saddle-path solution, i.e. a unique mapping between state and control variables.

of argument, it is sufficient to note that there are closed-form solutions that I will differentiate in Chapter 5 with respect to the deep parameters.

2.2 Pruned state-space system

Various simulation studies show that Taylor approximations of an order higher than one may generate explosive time paths, even though the first-order approximation is stable. This is due to artificial fixed points of the approximation. Kim et al. (2008) provide an example using the simple univariate model: $y_t = \rho y_{t-1} + \alpha y_{t-1}^2 + \varepsilon_t$ with $|\rho| < 1$ and $\alpha > 0$. This model has two steady states: a unique and locally stable one at $\bar{y} = 0$ and a second one at $\bar{y} = (1-\rho)/\alpha$ due to the higher-order approximation. If a shock sets y above the second steady state it will tend to diverge.

> *This is likely to be a generic problem with quadratic expansions – they will have extra steady states not present in the original model, and some of these steady states are likely to mark transitions to unstable behavior (Kim et al., 2008, p. 3408).*

Thus, the model may be neither stationary nor imply an ergodic probability distribution, yet these two assumptions are essential for identification and estimation. Thus, Kim et al. (2008) propose the pruning scheme, in which one omits terms from the policy functions that have higher-order effects than the approximation order.[8] For instance, given a second-order approximation, the state vector is decomposed into first-order ($\hat{x}_t^f$) and second-order ($\hat{x}_t^s$) effects ($\hat{x}_t = \hat{x}_t^f + \hat{x}_t^s$). The law of motions for these variables then only preserve

[8] This may seem ad hoc, but pruning can also be founded theoretically as a Taylor expansion in the perturbation parameter (Johnston et al., 2014; Lombardo and Uhlig, 2014) or on an infinite moving average representation (Lan and Meyer-Gohde, 2013). Schmitt-Grohé and Uribe (2004) also implicitly use pruning in their code to compute unconditional moments.

effects up to the second-order[9]:

$$\hat{x}^f_{t+1} = h_x \hat{x}^f_t + h_u u_{t+1}, \tag{2.8}$$

$$\begin{aligned}\hat{x}^s_{t+1} = {} & h_x \hat{x}^s_t + \frac{1}{2} h_{\sigma\sigma} \sigma^2 \\ & + \frac{1}{2}\left[H_{xx}\left(\hat{x}^f_t \otimes \hat{x}^f_t\right) + H_{uu}\left(u_{t+1} \otimes u_{t+1}\right)\right], \\ & + \frac{1}{2}\left[H_{xu}\left(\hat{x}^f_t \otimes u_{t+1}\right) + H_{ux}\left(u_{t+1} \otimes \hat{x}^f_t\right)\right]\end{aligned} \tag{2.9}$$

$$\begin{aligned}\hat{x}^f_{t+1} \otimes \hat{x}^f_{t+1} = {} & (h_x \otimes h_x)\left(\hat{x}^f_t \otimes \hat{x}^f_t\right) \\ & + (h_u \otimes h_u)\left(u_{t+1} \otimes u_{t+1}\right) \\ & + (h_x \otimes h_u)\left(\hat{x}^f_t \otimes u_{t+1}\right), \\ & + (h_u \otimes h_x)\left(u_{t+1} \otimes \hat{x}^f_t\right)\end{aligned} \tag{2.10}$$

$$\begin{aligned}\hat{y}_{t+1} = {} & g_x(\hat{x}^f_t + \hat{x}^s_t) + g_u u_{t+1} + \frac{1}{2} g_{\sigma\sigma} \sigma^2 \\ & + \frac{1}{2}\left[G_{xx}\left(\hat{x}^f_t \otimes \hat{x}^f_t\right) + G_{uu}\left(u_{t+1} \otimes u_{t+1}\right)\right]. \\ & + \frac{1}{2}\left[G_{xu}\left(\hat{x}^f_t \otimes u_{t+1}\right) + G_{ux}\left(u_{t+1} \otimes \hat{x}^f_t\right)\right]\end{aligned} \tag{2.11}$$

Thus, terms containing $\hat{x}^f_t \otimes \hat{x}^s_t$ and $\hat{x}^s_t \otimes \hat{x}^s_t$ are omitted, since they reflect third-order and fourth-order effects which are higher than the approximation order. Also, there are no second-order effects in u_{t+1}. It is convenient to extend the state vector to $z_t := [(\hat{x}^f_t)', (\hat{x}^s_t)', (\hat{x}^f_t \otimes \hat{x}^f_t)']'$: equations (2.8), (2.9), (2.10) and (2.11) can then be rewritten as a linear system of equations called the pruned state-space:

$$\begin{aligned} z_{t+1} &= c + A z_t + B \xi_{t+1} \\ \hat{y}_{t+1} &= d + C z_t + D \xi_{t+1} \end{aligned}$$

[9] See the technical appendix of Andreasen et al. (2014) for details.

$$\text{with } \xi_{t+1} := \begin{bmatrix} u_{t+1} \\ u_{t+1} \otimes u_{t+1} - vec(\Sigma) \\ u_{t+1} \otimes \hat{x}_t^f \\ \hat{x}_t^f \otimes u_{t+1} \end{bmatrix} \text{ and system matrices}$$

$$A := \begin{bmatrix} h_x & 0 & 0 \\ 0 & h_x & \frac{1}{2}H_{xx} \\ 0 & 0 & h_x \otimes h_x \end{bmatrix}, B := \begin{bmatrix} h_u & 0 & 0 & 0 \\ 0 & \frac{1}{2}H_{uu} & \frac{1}{2}H_{ux} & \frac{1}{2}H_{xu} \\ 0 & h_u \otimes h_u & h_u \otimes h_x & h_x \otimes h_u \end{bmatrix},$$

$$C := \begin{bmatrix} g_x & g_x & \frac{1}{2}G_{xx} \end{bmatrix}, D := \begin{bmatrix} g_u & \frac{1}{2}G_{uu} & \frac{1}{2}G_{ux} & \frac{1}{2}G_{xu} \end{bmatrix},$$

$$c := \begin{bmatrix} 0 \\ \frac{1}{2}h_{\sigma\sigma}\sigma^2 + \frac{1}{2}H_{uu}vec(\Sigma) \\ (h_u \otimes h_u)vec(\Sigma) \end{bmatrix}, d := \begin{bmatrix} \frac{1}{2}g_{\sigma\sigma}\sigma^2 + \frac{1}{2}G_{uu}vec(\Sigma) \end{bmatrix}.$$

Thus, conceptually, the solution of the pruned nonlinear DSGE model is a state-space system with a linear law of motion in z_t that is very similar to the canonical ABCD representation of a log-linearized DSGE model; hence, many concepts carry over. First, it is trivial to show that ξ_t is zero mean white noise with finite moments, since it is a function of $\hat{x}_t^f, u_{t+1}$ and $u_{t+1} \otimes u_{t+1}$. Regarding the computation of product moments up to the fourth order of ξ_t, see also Appendix D.[10] There, it is shown that even if the underlying shock process u_t is Gaussian, ξ_t is not normally distributed, since its higher-order cumulants are not equal to zero, therefore leaving scope for higher-order moments to contain additional information. Furthermore, if u_t has finite fourth moments, the pruned

[10] The product moments can be partitioned into several submatrices, which can be computed symbolically element-by-element, and contain many duplicate entries. For instance, note that $E[\xi_t \otimes \xi_t \otimes \xi_t]$ is of dimension n_ξ^3, but the number of distinct elements is $n_\xi(n_\xi + 1)(n_\xi + 2)/6$, because $\xi_{i,t}\xi_{j,t}\xi_{k,t} = \xi_{j,t}\xi_{i,t}\xi_{k,t} = \xi_{i,t}\xi_{k,t}\xi_{j,t}$ and so forth. One can use special matrix algebra analogous to the duplication matrix, called triplication and quadruplication matrix (Meijer, 2005), to ease the computations for higher-order product-moments of ξ_t.

state-space system then has finite second moments (see Andreasen et al. (2014) for closed-form expressions). Below, it is shown that if u_t has finite eighth moments, the pruned state-space system then has finite fourth moments. Note that apart from the existence of moments and temporal independence no distributional assumptions on u_t have to be imposed. Moreover, it can be shown that if the first-order approximation is stable, i.e. all Eigenvalues of h_x have modulus less than one, the pruned state-space is then also stable. In other words, all higher-order terms are unique and all Eigenvalues of A have modulus less than one. In fact, this approach works for any order of approximation, therefore:

Proposition 1 (Pruned state-space)**.** *Given an extended state vector z_t and an extended vector of innovations ξ_t, the pruned solution of a DSGE model can be rewritten as a linear time-invariant state-space system:*

$$z_{t+1} = c + Az_t + B\xi_{t+1} \tag{2.12}$$
$$y_{t+1} = \bar{y} + d + Cz_t + D\xi_{t+1} \tag{2.13}$$

Proof. See Andreasen et al. (2014) for the general algorithm. Mutschler (2015a) provides the corresponding vectors and matrices up to a third-order approximation. □

For the derivation of moments, cumulants and polyspectra, it is advisable to work with zero mean variables to simplify notation and the expressions in the following chapters. Accordingly, the mean of the extended state vector is equal to:

$$\mu_z := E(z_t) = (I_{n_z} - A)^{-1}c \tag{2.14}$$

with $n_z = 2n_x + n_x^2$. Intuitively, the mean of the second-order approximation consists of two effects: The first-order effect ($E(\hat{x}_t^f) = E(x_t^f) - \bar{x} = 0$) simply states certainty equivalence, i.e. the mean of x_t is equal to the steady state in a first-order approximation. Using

a second-order approximation, the mean is adjusted for risk, given a constant $\frac{1}{2}h_{\sigma\sigma}\sigma^2$ and the variance of the states Σ_x:

$$vec(\Sigma_x) = (I_{n_x^2} - h_x)^{-1}(h_u \otimes h_u)vec(\Sigma). \tag{2.15}$$

Since there is a linear relationship between y_t and z_{t-1} in (2.13), the mean of y_t is equal to

$$\mu_y := E(y_t) = \bar{y} + C\mu_z + d. \tag{2.16}$$

In summary, the pruned state-space representation is a stable system and has well-defined statistical properties. In particular, an approximation to higher orders yields non-Gaussian innovations. Intuitively, this yields additional restrictions on moments and polyspectra, which may tighten the identifiability of model parameters. Also, since higher-order cumulants and polyspectra measure the departure from Gaussianity, additional information may also be gained by considering higher-order statistics. In Chapters 7 and 8, I show how to incorporate these additional restrictions into formal identifiability criteria.

Chapter 3

Example models

3.1 The Kim (2003) model

The Kim (2003) model builds upon the canonical neoclassical growth model (see for example Schmitt-Grohé and Uribe (2004)), however, extending it for two kinds of investment adjustment costs. First, intertemporal adjustment costs in the fashion of Lucas and Prescott (1971) are introduced into the capital accumulation equation governed by a parameter ϕ, which involve a nonlinear substitution between capital k_t and investment i_t:

$$k_t = \left[\delta \left(\frac{i_t}{\delta} \right)^{1-\phi} + (1-\delta) \left(k_{t-1} \right)^{1-\phi} \right]^{\frac{1}{1-\phi}} \tag{3.1}$$

with δ denoting the depreciation rate. Note that $\phi = 0$ implies the usual linear capital accumulation specification.

Second, multisectoral adjustment costs in the fashion of Sims (1989) and Valles (1997) enter the national budget constraint given a parameter θ, which are captured by a nonlinear transformation

between consumption c_t and investment i_t:

$$a_{t-1}k_{t-1}^{\alpha} = \left[(1-s)\left(\frac{c_t}{1-s}\right)^{1+\theta} + s\left(\frac{i_t}{s}\right)^{1+\theta}\right]^{\frac{1}{1+\theta}} \tag{3.2}$$

with a_t denoting the level of technology. The composite parameter $s = \frac{\beta\delta\alpha}{1-\beta+\delta\beta}$ consists of the depreciation rate δ, the discount factor β and the share of capital in production α. Note that for $\theta = 0$ the transformation is linear.

The representative agent maximizes

$$E_0 \sum_{t=0}^{\infty} \beta^t \ln c_t$$

subject to the national income identity and the capital accumulation equation. The corresponding Euler equation is

$$\lambda_t(1+\theta)\left(\frac{i_t}{s}\right)^{\theta}\left(\frac{i_t}{\delta k_t}\right)^{\phi} = \beta E_t\lambda_{t+1}\left[\alpha(1+\theta)a_t^{1+\theta}k_t^{\alpha(1+\theta)-1} + (1-\delta)(1+\theta)\left(\frac{E_t i_{t+1}}{\delta k_t}\right)^{\phi}\left(\frac{E_t i_{t+1}}{s}\right)^{\theta}\right] \tag{3.3}$$

with auxiliary variable $\lambda_t = \frac{(1-s)^{\theta}}{(1+\theta)c_t^{1+\theta}}$. Note that for $\phi = \theta = 0$ this simplifies to the canonical Euler equation.

To close the model, technology evolves according to

$$log(a_t) = \rho_a log(a_{t-1}) + \varepsilon_{a,t} \tag{3.4}$$

with ρ_a measuring persistence and $\varepsilon_{a,t} \sim iid(0, \sigma_a^2)$. Lastly, an auxiliary equation for the expectation of future technology shocks is added to the model equations:

$$E_t\varepsilon_{a,t+1} = 0 \tag{3.5}$$

In summary, the model equations f are given by (3.1), (3.2), (3.3), (3.4) and (3.5). There are two exogenous (k_t and a_t) and no endogenous states. The controls are c_t and i_t and are both assumed to be observable. There is one shock on technology $\varepsilon_{a,t}$ with standard deviation σ_a, which I set equal to the perturbation parameter. Thus, the definitions and ordering of variables is given by

$$u_t = \varepsilon_{a,t}, \quad x_t = (k_t, a_t)', \quad y_t = (c_t, i_t)', \quad \sigma = \sigma_A, \quad \eta = (0 \;\; 1)'.$$

The steady state of the model is given by

$$\varepsilon_a = 0,\ a = 1,\ k = \left(\frac{\delta}{sa}\right)^{\frac{1}{\alpha-1}},\ i = \delta k,$$

$$c = (1-s)\left[\frac{(\alpha k^{\alpha})^{1+\theta} - s\left(\frac{i}{s}\right)^{1+\theta}}{1-s}\right]^{\frac{1}{1+\theta}}.$$

I will consider identification of the parameter vector θ at local point θ_0 and prior specification given in Table 3.1. The code contains three model specifications:

1. The shock on technology is Gaussian.
2. The shock on technology is Gaussian and Gaussian measurement errors are added to the model ($E_t u_{c,t+1} = E_t u_{i,t+1} = 0$).
3. The shock on technology is t-distributed with $df = 10$ degrees of freedom and the prior for df is uniform with lower bound 8 and upper bound 20.

Table 3.1: Parameters, priors and bounds for Kim (2003)

Parameters		*Prior specification*			*Bounds*	
θ	θ_0	Density	Para (1)	Para (2)	Lower	Upper
α	0.60	Gamma	0.60	0.30	1e-5	1
β	0.99	Uniform	0.95	0.9999	0.9	0.99999
δ	0.0125	Uniform	0.01	0.02	0.01	0.02
θ	1	Normal	1.00	0.50	-5	5
ρ_a	0.7	Beta	0.50	0.20	1e-5	0.99999
ϕ	2	Normal	2.00	0.50	-5	5
σ_a	0.5	InvGamma	0.50	4.00	1e-8	5

Notes: Para (1) and (2) list the means and the standard deviations for Beta, Gamma, and Normal distributions; the upper and lower bound of the support for the Uniform distribution; s and v for the Inverse Gamma distribution, where $\wp_{IG}(\sigma|v,s) \propto \sigma^{-v-1}e^{-vs^2/2\sigma^2}$. The effective prior is truncated at the boundary of the determinacy region.

3.2 The An and Schorfheide (2007) model

The An and Schorfheide (2007) model is a prototypical DSGE model, representative of the class of models used in the analysis of monetary policy. The model economy consists of a representative household, a continuum of intermediate goods producing firms, a final good producing firm, a fiscal authority and a monetary authority. A domestic bond market is assumed where the household has access to bonds issued by the government. By assumption, a firm $j \in [0,1]$ of the intermediate goods sector produces exactly one intermediate good Y_t^j. Then all intermediate goods are used in the production of the final good Y_t. There is perfect competition in the final good sector, while the producers of intermediate goods are in monopolistic competition to each other and set their price subject to quadratic price adjustment costs in the fashion of Rotemberg (1982). In the following, I will outline the optimality conditions of all agents, the law of motions of the exogenous processes and the mapping into the gen-

eral framework of Chapter 2. The section is concluded by a short comparison of the model to the widely-cited Smets and Wouters (2007) model.

Household sector

The representative household optimizes present as well as expected future utility

$$E_t \sum_{s=0}^{\infty} \beta^s U_{t+s}. \tag{3.6}$$

with $\beta < 1$ denoting the discount factor. The contemporaneous utility function

$$U_t = \frac{(C_t/A_t)^{1-\tau} - 1}{1-\tau} + \ln\left(\frac{M_t}{P_t}\right) - H_{t+s}$$

has three arguments: consumption C_t, hours worked H_t and real money holdings M_t/P_t. The marginal utility of additional consumption and money is positive, whereas more labor reduces utility. The utility function is additively separable such that the linearly approximate Euler equation can be readily interpreted as a New-Keynesian IS curve (see Driscool (2000) or Ireland (2004) for a discussion). τ measures the relative risk aversion of the household.[11] Furthermore, consumption is considered with respect to a habit stock represented by the level of technology A_t. This improves the reaction of output and consumption to various shocks in so far as the impulse-responses exhibit a *hump-shaped* pattern. The dynamic behavior of the model in the short term is both qualitatively and statistically more realistic (see Fuhrer (2000) or McCallum and Nelson (1999) for further discussion). Productivity A_t is the driving force of the economy and

[11] $1/\tau = d[\ln(C_{t+1}/C_t)]/d[\ln(-dC_{t+1}/dC_t)]$ is the intertemporal elasticity of substitution.

evolves according to a unit root process, $\ln A_t - \ln A_{t-1} = \ln \gamma + \ln z_t$. On average, aggregate productivity grows with $\ln \gamma$, while z_t are innovations to this rate:

$$\ln z_t = \rho_z \ln z_{t-1} + \epsilon_{z,t} \tag{3.7}$$

with $\epsilon_{z,t} \sim iid(0, \sigma_z^2)$. In each period the household's assets are given by money holdings of the previous period M_{t-1} and the returns of the risk-free government bond B_{t-1}. Formally, these are one-period securities which pay gross interest R_{t-1} in the following period. The household takes the real wage W_t as given and supplies perfectly elastic labor service to the intermediate goods firms. In return, she receives real labor income in the amount of $W_t H_t$ and, additionally, aggregated residual profits D_t from the firms, since it is assumed that the firms are owned by the household. Income and wealth are used to finance (i) consumption C_t given its price P_t, (ii) new domestic bonds B_t and (iii) new money holdings M_t. Lastly, the household pays a lump-sum tax to the government. In total, this defines the (real) budget constraint

$$C_t + \frac{M_t}{P_t} + \frac{B_t}{P_t} + \frac{T_t}{P_t} = W_t H_t + D_t + \frac{M_{t-1}}{P_t} + \frac{R_{t-1} B_{t-1}}{P_t}. \tag{3.8}$$

Maximizing the objective function (3.6) subject to the budget constraint (3.8) yields

$$\underbrace{\left(\frac{C_t}{A_t}\right)^{-\tau} \frac{1}{A_t}}_{U_t^c} = \beta E_t \left[\frac{R_t}{\pi_{t+1}} \underbrace{\left(\frac{C_{t+1}}{A_{t+1}}\right)^{-\tau} \frac{1}{A_{t+1}}}_{U_{t+1}^c} \right], \tag{3.9}$$

$$W_t = \left(\frac{C_t}{A_t}\right)^{\tau} A_t, \tag{3.10}$$

$$\left(\frac{M_t}{P_t}\right)^{-1} = \frac{(C_t/A_t)^{-\tau}}{A_t} \left(\frac{R_t - 1}{P_t R_t}\right). \tag{3.11}$$

Equation (3.9) is the Euler equation of intertemporal optimality. It reflects the trade-off between consumption and savings. If the household saves a (marginal) unit of consumption, she can consume R_t/π_{t+1} units in the following period. The marginal utility of consuming today is equal to U_t^c, whereas consuming tomorrow has expected marginal utility of $E_t\left[U_{t+1}^c\right]$. Discounting expected marginal utility with β, the household must be indifferent between both choices in the optimum. Equations (3.10) and (3.11) reflect intratemporal optimality. On the one hand, the optimal choice for the labor supply: the real wage must be equal to the marginal rate of substitution between labor and consumption. On the other hand, a New-Keynesian LM curve: real money demand depends positively on consumption (relative to habit) and negatively on the interest rate. Due to preferences being additively separable in the utility function, money holding will not enter in any of the other structural equations. Equation (3.11) then becomes completely recursive to the rest of the system of equations and can be omitted.

Final good sector

A representative firm aggregates all intermediate goods into a final consumption good using a Dixit and Stiglitz (1977) aggregator:

$$Y_t = \left[\int_0^1 (Y_t^j)^{1-v}\, dj\right]^{\frac{1}{1-v}}. \tag{3.12}$$

Let P_t^j denote the price of intermediate good Y_t^j, then the demand function for good Y_t^j can be derived from the cost minimization as

$$Y_t^j = \left(\frac{P_t^j}{P_t}\right)^{\frac{-1}{v}} Y_t. \tag{3.13}$$

Obviously, $\frac{1}{v}$ is the elasticity of demand for each intermediate good Y_t^j. Inserting this into (3.12) yields

$$P_t = \left[\int_0^1 (P_t^j)^{\frac{v-1}{v}} \, dj \right]^{\frac{v}{v-1}}. \tag{3.14}$$

P_t is a composite index of all intermediate goods prices and can thus be interpreted as the price for final good Y_t.

Intermediate goods sector

Each firm has market power for its produced good Y_t^j and maximizes expected profits given a linear production function:

$$Y_t^j = A_t N_t^j. \tag{3.15}$$

A_t is the common exogenous parameter of productivity and N_t^j is the input of hours worked compensated by the real wage W_t. Furthermore, to introduce nominal price rigidities into the model, the firms face quadratic price adjustment costs as in Rotemberg (1982):

$$AC_t^j = \frac{\phi}{2} \left(\frac{P_t^j}{P_{t-1}^j} - \pi \right)^2 Y_t^j.$$

$\pi \geq 1$ denotes the rate of inflation in equilibrium, which is targeted by the monetary authority. If the firm links its price to inflation, there are no adjustment costs. $\phi \geq 0$ is, therefore, a measure of nominal price rigidity in the economy. It is taken into consideration that the firms are owned by the household. Accordingly, $\beta^s Q_{t+s|t}$ is the present value of a unit of consumption in period $t+s$ or, respectively, the marginal utility of an additional unit of profit; therefore

$$Q_{t+s|t} \equiv \frac{U_{t+s}^C}{U_t^C} = \left(\frac{C_{t+s}}{C_t} \right)^{-\tau} \left(\frac{A_t}{A_{t+s}} \right)^{1-\tau}. \tag{3.16}$$

and in particular $Q_{t|t} = 1$. U_{t+s}^C is the value of an additional unit of profits in period $t+s$, while U_t^C is the marginal utility of an additional unit of consumption. In other words, U_t^C are the costs of transforming a unit of consumption into additional profits. The factor $Q_{t+s|t}$ defines this as a ratio, which is known as the (marginal) Tobin-Q (Tobin, 1969). Each firm chooses the required amount of labor N_t^j and the price P_t^j to maximize the present value of expected future profits. It takes into consideration the production function (3.15), the supply of labor given in equation (3.10), and the demand for its good given in equation (3.13). In summary, real profit D_{t+s}^j of an intermediate goods producer is given by:

$$D_{t+s}^j = \underbrace{\beta^s Q_{t+s|t}}_{\text{discount factor}} \left(\underbrace{\frac{P_{t+s}^j}{P_{t+s}} Y_{t+s}^j}_{\text{revenues}} - \underbrace{W_{t+s} N_{t+s}^j}_{\text{labor costs}} - \underbrace{\frac{\phi}{2} \left(\frac{P_{t+s}^j}{P_{t+s-1}^j} - \pi \right)^2 Y_{t+s}^j}_{\text{price adjustment costs}} \right) \tag{3.17}$$

The condition of optimality is then equal to

$$\begin{aligned} &\left[\left(1 - \frac{1}{v}\right) Y_t^j \frac{1}{P_t} + \frac{1}{v} \left(\frac{C_t}{A_t} \right)^\tau Y_t^j \left(\frac{P_t^j}{P_t} \right)^{-1} \frac{1}{P_t} \right. \\ &\left. -\phi \left(\frac{P_t^j}{P_{t-1}^j} - \pi \right) \frac{1}{P_{t-1}^j} Y_t^j + \frac{\phi}{2v} \left(\frac{P_t^j}{P_{t-1}^j} - \pi \right)^2 Y_t^j \left(\frac{P_t^j}{P_t} \right)^{-1} \frac{1}{P_t} \right] \\ &+ \phi \beta E_t Q_{t+1|t} \left[\left(\frac{P_{t+1}^j}{P_t^j} - \pi \right) \frac{P_{t+1}^j}{(P_t^j)^2} Y_{t+1}^j \right] = 0. \end{aligned} \tag{3.18}$$

Assuming flexible prices ($\phi = 0$) this simplifies to:

$$P_t^j = \frac{1}{1-v} P_t \left(\frac{C_t}{A_t}\right)^{\tau} \overset{(3.10)}{=} \frac{1}{1-v} P_t \frac{W_t}{A_t}. \tag{3.19}$$

Without adjustment costs, each firm sets its price P_t^j as a mark-up $1/(1-v)$ over marginal costs $W_t P_t / A_t$. Introducing adjustment costs, however, the price will be set as a mark-up on an average of future expected marginal costs.

Government sector

The government consists of two institutions, a monetary and a fiscal authority. The monetary authority sets the nominal interest rate R_t using an interest rate feedback rule in the fashion of Taylor (1993):

$$R_t = R_t^{*\,1-\rho_R} R_{t-1}^{\rho_R} e^{\epsilon_{R,t}} \tag{3.20}$$

Two specifications for R_t^* are considered:

$$R_t^* = \begin{cases} r\pi^* \left(\frac{\pi_t}{\pi^*}\right)^{\psi_1} \left(\frac{Y_t}{Y_t^*}\right)^{\psi_2} & \text{(output-gap rule)} \\ r\pi^* \left(\frac{\pi_t}{\pi^*}\right)^{\psi_1} \left(\frac{Y_t}{\gamma Y_{t-1}}\right)^{\psi_2} & \text{(output-growth rule)} \end{cases}$$

The first specification assumes that the monetary authority responds to inflation deviations as well as deviations of output Y_t from potential output Y_t^*, which is the output level in absence of nominal rigidities ($\phi = 0$). In the second specification, the central bank does not react to the output gap, but to deviations of output growth from its equilibrium steady state γ. ψ_1 and ψ_2 are the corresponding weights, while ρ_R measures the persistence of the interest rate. $\epsilon_{R,t} \sim iid(0, \sigma_R^2)$ is a monetary policy shock.

The revenue side of fiscal policy is characterized by lump-sum taxes, issuance of new government bonds and seignorage on money

holdings. These are used to finance public expenditures $P_t G_t$. The government budget is therefore equal to

$$\underbrace{P_t G_t}_{\text{expenditures}} = \underbrace{T_t}_{\text{taxes}} + \underbrace{B_t - R_{t-1} B_{t-1}}_{\text{new bonds}} + \underbrace{M_t - M_{t-1}}_{\text{seignorage}}.$$

Real government spending are assumed to evolve stochastically as a ratio $\zeta_t \in [0; 1]$ of output, $G_t = \zeta_t Y_t$, such that

$$\frac{Y_t}{Y_t - G_t} = \frac{1}{1 - \zeta_t} =: g_t. \tag{3.21}$$

By assumption, the log of g_t follows an AR(1)-process:

$$\ln(g_t) = (1 - \rho_g) \ln(g) + \rho_g \ln(g_{t-1}) + \epsilon_{g,t} \tag{3.22}$$

with $\epsilon_{g,t} \sim iid(0, \sigma_g^2)$ and ρ_g measuring persistence.

Measurement equations

Lastly, the model needs to be linked to data. It is assumed that quarterly data is available for the (i) quarter-on-quarter growth rate of per capita GDP (YGR_t), (ii) annualized inflation rate ($INFL_t$) and (iii) annualized nominal interest rate (INT_t). These observable variables are linked to the model equations:

$$\begin{aligned} YGR_t &= \gamma^{(Q)} + 100(\hat{y}_t - \hat{y}_{t-1} + \hat{z}_t), \\ INFL_t &= \pi^{(A)} + 400\hat{\pi}_t, \\ INT_t &= \pi^{(A)} + r^{(A)} + 4\gamma^{(Q)} + 400\hat{R}_t. \end{aligned}$$

The parameters $\gamma^{(Q)}$, $r^{(A)}$ and $\pi^{(A)}$ are accordingly linked to γ, β and π:

$$\gamma = e^{\frac{\gamma^{(Q)}}{100}}, \quad \beta = e^{-\frac{r^{(A)}}{400}}, \quad \pi = e^{\frac{\pi^{(A)}}{400}} \tag{3.23}$$

Additionally, it is assumed that the target rate for inflation corresponds to the inflation rate in equilibrium: $\pi = \pi^*$.

Symmetric equilibrium

All intermediate goods producing firms make identical choices with respect to prices and labor inputs. Hence, one can omit the index j, i.e. $Y_t^j = Y_t$, $N_t^j = N_t$, $P_t^j = P_t$ and $\pi_t = P_t/P_{t-1}$. Furthermore, the market clearing conditions have to hold for all periods. That is, $H_t = N_t, B_t = 0$ and $M_t - M_{t-1} = 0$. Given these conditions, equation (3.17) can be simplified to $D_t = Y_t - W_t H_t - \frac{\phi}{2}(\pi_t - \pi)^2 Y_t$. The budget of the fiscal authority in equilibrium is equal to $G_t = \frac{T_t}{P_t}$, while the budget (3.8) of the household is then given by

$$Y_t = C_t + G_t + \frac{\phi}{2}(\pi_t - \pi)^2 Y_t. \tag{3.24}$$

Potential output Y_t^* is obtained under the assumption of flexible prices, i.e. setting $\phi = 0$. Inserting (3.19) and (3.21) in (3.24) yields:

$$Y_t^* = (1-v)^{\frac{1}{\tau}} A_t g_t. \tag{3.25}$$

The optimality condition (3.18) of the intermediate goods producers simplifies to

$$\begin{aligned} 1 = \frac{1}{v}\left[1 - \left(\frac{C_t}{A_t}\right)^{\tau}\right] + \phi(\pi_t - \pi)\left[\left(1 - \frac{1}{2v}\right)\pi_t + \frac{\pi}{2v}\right] \\ - \phi\beta E_t\left[\left(\frac{\frac{C_{t+1}}{A_{t+1}}}{\frac{C_t}{A_t}}\right)^{-\tau} \frac{\frac{Y_{t+1}}{A_{t+1}}}{\frac{Y_t}{A_t}}(\pi_{t+1} - \pi)\pi_{t+1}\right]. \end{aligned} \tag{3.26}$$

Structural form

Equations (3.7), (3.9), (3.20), (3.22), (3.24) and (3.26) characterize the optimal dynamics of the four endogenous variables Y_t, C_t, π_t, R_t and the two exogenous processes g_t and z_t. The functional form implies that Y_t and C_t follow a unit root process due to the unit root

in A_t. Since the following analysis focuses on stationary variables, all variables are detrended relative to aggregate productivity, i.e. $y_t = \frac{Y_t}{A_t}$ and $c_t = \frac{C_t}{A_t}$. Now, in the absence of shocks, the economy converges to a constant steady state which is given by:

$$\gamma \overset{(3.7)}{=} \frac{A_{t+1}}{A_t},\ r \overset{(3.9)}{=} \frac{\gamma}{\beta},\ c \overset{(3.19)}{=} (1-v)^{\frac{1}{\tau}},\ R \overset{(3.20)}{=} r\pi,\ y \overset{(3.25)}{=} c \cdot g. \tag{3.27}$$

Let $\hat{c}_t = \ln(c_t/c),\ \hat{y}_t = \ln(y_t/y),\ \hat{g}_t = \ln(g_t/g),\ \hat{\pi}_t = \ln(\pi_t/\pi),\ \hat{R}_t = \ln(R_t/R)$ and $\hat{z}_t = \ln(z_t/1)$ denote log deviations of a variable from steady state, then the structural equations can be reformulated to:

$$\begin{aligned} 0 = {} & \frac{1-v}{v\phi\pi^2}\left(e^{\tau\hat{c}_t}-1\right) - \left(e^{\hat{\pi}_t}-1\right)\left[\left(1-\frac{1}{2v}\right)e^{\hat{\pi}_t} + \frac{1}{2v}\right] \\ & + \beta E_t\left(e^{\hat{\pi}_{t+1}}-1\right)e^{-\tau\hat{c}_{t+1}+\tau\hat{c}_t+\hat{y}_{t+1}-\hat{y}_t+\hat{\pi}_{t+1}} \end{aligned} \tag{3.28}$$

$$\begin{aligned} \hat{R}_t = {} & \rho_R\hat{R}_{t-1} + (1-\rho_R)\psi_1\hat{\pi}_t \\ & + (1-\rho_R)\psi_2 \begin{cases} (\hat{y}_t - \hat{g}_t) + \epsilon_{R,t} \\ (\hat{y}_t - \hat{y}_{t-1} + \hat{z}_t) + \epsilon_{R,t} \end{cases} \end{aligned} \tag{3.29}$$

$$1 = E_t\left[e^{-\tau\hat{c}_{t+1}+\tau\hat{c}_t+\hat{R}_t-\rho_z\hat{z}_t-\hat{\pi}_{t+1}}\right] \tag{3.30}$$

$$e^{\hat{c}_t-\hat{y}_t} = e^{-\hat{g}_t} - \frac{\phi\pi^2 g}{2}\left(e^{\hat{\pi}_t}-1\right)^2 \tag{3.31}$$

$$\hat{g}_t = \rho_g\hat{g}_{t-1} + \epsilon_{g,t} \tag{3.32}$$

$$\hat{z}_t = \rho_z\hat{z}_{t-1} + \epsilon_{z,t} \tag{3.33}$$

$$YGR_t = \gamma^{(Q)} + 100(\hat{y}_t - \hat{y}_{t-1} + \hat{z}_t), \tag{3.34}$$

$$INFL_t = \pi^{(A)} + 400\hat{\pi}_t, \tag{3.35}$$

$$INT_t = \pi^{(A)} + r^{(A)} + 4\gamma^{(Q)} + 400\hat{R}_t \tag{3.36}$$

$$E_t\epsilon_{R,t+1} = 0 \tag{3.37}$$

$$E_t\epsilon_{g,t+1} = 0 \tag{3.38}$$

$$E_t\epsilon_{z,t+1} = 0. \tag{3.39}$$

Use is made of the fact that for any variable $x_t = xe^{\hat{x}_t}$.

Mapping into DSGE framework

In summary, equations (3.28) to (3.39) represent a system of twelve nonlinear expectational equations. There are three exogenous states $\hat{R}_t$, $\hat{g}_t$ and $\hat{z}_t$, and one endogenous state variable $\hat{y}_t$. The controls are $\hat{c}_t$, $\hat{\pi}_t$, $YGR_t, INFL_t$ and INT_t. Note that only $YGR_t, INFL_t$ and INT_t are assumed to be observable. There are three stochastic innovations: a monetary $\epsilon_{R,t}$, a fiscal $\epsilon_{g,t}$ and a technological shock $\epsilon_{z,t}$ with standard deviations σ_R, σ_g and σ_z. Furthermore, I set the perturbation parameter equal to the standard deviation of the shock on technology. Thus, the definition and ordering of variables is given by $u_t = (\epsilon_{R,t}, \epsilon_{g,t}, \epsilon_{z,t})'$, $x_t = (\hat{y}_t, \hat{R}_t, \hat{g}_t, \hat{z}_t)'$, $y_t = (\hat{c}_t, \hat{\pi}_t, YGR_t, INFL_t, INT_t)'$, $\sigma = \sigma_z$ and

$$\eta = \begin{bmatrix} \sigma_R/\sigma_z & 0 & 0 \\ 0 & \sigma_g/\sigma_z & 0 \\ 0 & 0 & 1 \end{bmatrix}, \quad S_{obs} = \begin{bmatrix} 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 1 \end{bmatrix}$$

with observables equal to $S_{obs} \cdot y_t$. The steady state of this model is given by $\hat{y} = \hat{R} = \hat{g} = \hat{z} = \hat{c} = \hat{\pi} = 0$, $YGR = \gamma^{(Q)}$, $INFL = \pi^{(A)}$, $INT = \pi^{(A)} + r^{(A)} + 4\gamma^{(Q)}$. I will consider identification of the parameter vector θ at the local point θ_0 and prior specification given in Table 3.2. The code contains three model specifications:

1. All structural shocks are Gaussian.
2. All structural shocks are Gaussian and Gaussian measurement errors are added to the model ($E_t\epsilon_{YGR,t+1} = 0$, $E_t\epsilon_{INFL,t+1} = 0$ and $E_t\epsilon_{INT,t+1} = 0$).
3. All structural shocks are t-distributed with $df = 10$ degrees of freedom and the prior for df is uniform with lower bound 8 and upper bound 20.

Table 3.2: Parameters, priors and bounds for An/Schorfheide (2007)

Parameters		*Prior specification*			*Bounds*	
θ	θ_0	Density	Para (1)	Para (2)	Lower	Upper
τ	2.00	Gamma	2.00	0.50	1e-5	10
ϕ	50	Gamma	50	20	1e-5	100
ψ_1	1.50	Gamma	1.50	0.25	1e-5	10
ψ_2	0.125	Gamma	0.50	0.25	1e-5	10
ρ_R	0.75	Beta	0.50	0.20	1e-5	0.99999
ρ_g	0.95	Beta	0.80	0.10	1e-5	0.99999
ρ_z	0.90	Beta	0.66	0.15	1e-5	0.99999
$r^{(A)}$	1.00	Gamma	0.80	0.50	1e-5	10
$\pi^{(A)}$	3.20	Gamma	4.00	2.00	1e-5	20
γ_Q	0.55	Normal	0.40	0.20	-5	5
$100\sigma_R$	0.2	InvGamma	0.30	4.00	1e-8	5
$100\sigma_g$	0.6	InvGamma	0.40	4.00	1e-8	5
$100\sigma_z$	0.3	InvGamma	0.40	4.00	1e-8	5
ν	0.10	Beta	0.10	0.05	1e-5	0.99999
c/y	0.85	Beta	0.85	0.10	1e-5	0.99999

Notes: Para (1) and (2) list the means and the standard deviations for Beta, Gamma, and Normal distributions; the upper and lower bound of the support for the Uniform distribution; s and v for the Inverse Gamma distribution, where $\wp_{IG}(\sigma|v,s) \propto \sigma^{-v-1}e^{-vs^2/2\sigma^2}$. The effective prior is truncated at the boundary of the determinacy region.

Comparison to Smets and Wouters (2007)

The An and Schorfheide (2007) model is a simplified version of the widely-cited Smets and Wouters (2007) model. Since current state-of-the-art DSGE models like the NAWM model (Christoffel et al., 2008), the EAGLE model (Gomes et al., 2012) or the GIMF model (Laxton et al., 2010) build upon Smets and Wouters (2007), it is worthwhile to outline the differences to An and Schorfheide (2007).

The largest differences apply to the labor and capital market. In Smets and Wouters (2007)'s model, households perform as mo-

nopolists for labor service and set wages according to their decision problem. Furthermore, Smets and Wouters (2007) account for wage rigidities by using a Calvo (1983) rule, where only a certain proportion of households is able to adjust their wages at each period. Therefore, a continuum of households is used instead of a representative household as in the An and Schorfheide (2007) model. In addition to governmental bonds, Smets and Wouters (2007) implement the households ability to accumulate capital. For that matter, households choose between saving in bonds or investing in their capital stock, which is rented out to the intermediate goods producing firms. The households receive a return on their investment and the firm's output increases according to their production function. The final good sector is almost identical in both models despite replacing the Dixit and Stiglitz (1977) aggregator by the more general aggregator developed in Kimball (1995). The intermediate sector differs in two details. On the one hand, the firm's production technology accounts for the capital stock as well as the labor input. A Cobb-Douglas production function is used lessened by a fix cost. On the other hand, Smets and Wouters (2007) model price rigidities by a Calvo (1983) rule similar to the wage instead of quadratic price adjustment costs. Furthermore, Smets and Wouters (2007) embed various different shocks in the model equations allowing for more degrees of freedom in their Bayesian estimation exercise. In comparison, Smets and Wouters (2007) is more elaborate but also significantly more complex. An and Schorfheide (2007)'s model encompasses many features of modern DSGE models none the less. In particular, it focuses on the analysis of monetary policy, which is modelled as a feedback rule in the fashion of Taylor (1993).

Chapter 4

Higher-order statistics for DSGE models

Since a Gaussian process is completely characterized by its first two moments, most linear DSGE models focus on Gaussian innovations for simplicity.[12] If, however, one relaxes the assumption of linearity or uses non-Gaussian innovations, it is natural to analyze whether it is possible to exploit information from higher-order moments for the calibration, estimation and identification of parameters. Researchers in mathematics, statistics and signal processing have developed tools, called higher-order statistics (HOS), to solve detection, estimation and identification problems when the noise source is non-Gaussian or one is faced with nonlinearities; however,

[12] Two notable exceptions are Curdia et al. (2014) and Chib and Ramamurthy (2014) who estimate two standard linear DSGE models with Student's t-distributed errors and conclude that these models outperform their Gaussian counterparts. The code can handle this case as well, since a t-distributed random variable can be represented as the product of two independent random variables, an inverse Gamma distributed variable and a Gaussian one. See Appendix D for details.

applications in the macroeconometric literature are rather sparse.[13] The basic tools of HOS are cumulants, which are defined as the coefficients in the Taylor expansion of the log moment generating function in the time domain; and polyspectra, which are defined as Fourier transformations of the cumulants in the frequency domain.

In this chapter, I derive closed-form expressions for unconditional third- and fourth-order moments, cumulants and corresponding polyspectra for non-Gaussian and/or nonlinear DSGE models. I limit the exposition to fourth-order statistics, since third-order cumulants and the bispectrum capture nonlinearities (or non-Gaussianity) for a skewed process, whereas the fourth-order cumulants and the trispectrum can be used in the case of a non-Gaussian symmetric probability distribution. Regarding the approximation of nonlinear DSGE models I focus on the pruning scheme of section 2.2, since the pruned state-space (PSS from now on) is a linear, stationary and ergodic state-space system. In the PSS, however, Gaussian innovations do not imply Gaussian likelihood, leaving scope for higher-order statistics to capture information from nonlinearities and non-Gaussianity. In the following exposition I focus on Taylor approximations and pruning up to second-order, since an extension beyond second-order is – apart from tedious notation – straightforward. Accordingly, the procedures are demonstrated by means of the An and Schorfheide (2007) model, see Chapter 3.2 for the model description. Using both the Gaussian as well as Student's t distribution as the underlying shock process, I compare the theoretical results with simulated higher-order moments. I focus particularly on skewness and excess kurtosis in the simulations, since these are

[13] For introductory literature and tutorials on HOS, see the textbooks of Brillinger (2001), Nikias and Petropulu (1993), Priestley (1983) and the references therein. Most theoretical and applied econometric literature is either concerned with tests for normality (e.g. Bao (2013); Rusticelli et al. (2008)) or method of moments estimation (e.g. Dagenais and Dagenais (1997); Erickson and Whited (2002)).

typical measures an applied researcher would like to match in a calibration exercise. On the other hand auto- and cross-skewness as well as kurtosis may contain valuable information in an estimation exercise.[14]

4.1 Moments, cumulants and polyspectra

Reconsider the model framework of Chapter 2. Proposition 1 expresses the solution of a DSGE model as the linear time-invariant state-space system. Focusing on mean zero variables the PSS is given by

$$\widetilde{z}_{t+1} = A\widetilde{z}_t + B\xi_{t+1} \tag{4.1}$$
$$\widetilde{y}_{t+1} = C\widetilde{z}_t + D\xi_{t+1} \tag{4.2}$$

with states z_t, controls y_t and stochastic innovations ξ_t. A tilde denotes deviations from the unconditional mean, e.g. $\widetilde{y}_t := y_t - E(y_t)$. For the sake of notation, all control variables are assumed to be observable. The vector of innovations ξ_t has $E(\xi_t) = 0$ and finite covariance matrix $E(\xi_t\xi_t') =: \Sigma_\xi$. Furthermore, ξ_t is kth-order white noise with finite higher-order moments, which implies y_t is a kth-order stationary process.[15]

Formally, the kth-order ($k = 2, 3, 4$) cumulants of the kth-order stationary, mean zero vector process $\widetilde{z}_t$ ($t_1, t_2, t_3 \geq 0$) are given by

[14] Mutschler (2015a) demonstrates the procedures by means of the Smets and Wouters (2007) model for a first-order approximation, the An and Schorfheide (2007) model for a second-order approximation and the neoclassical growth model for a third-order approximation.

[15] This is basically an extension of the usual covariance stationarity assumption. See also Priestley (1983, p. 105) for a formal definition of stationary up to order k.

the n_z^k vectors $\mathcal{C}_{k,z}$ as

$$\begin{aligned}
\mathcal{C}_{2,z}(t_1) &:= E[\widetilde{z}_0 \otimes \widetilde{z}_{t_1}], \\
\mathcal{C}_{3,z}(t_1, t_2) &:= E[\widetilde{z}_0 \otimes \widetilde{z}_{t_1} \otimes \widetilde{z}_{t_2}], \\
\mathcal{C}_{4,z}(t_1, t_2, t_3) &:= E[\widetilde{z}_0 \otimes \widetilde{z}_{t_1} \otimes \widetilde{z}_{t_2} \otimes \widetilde{z}_{t_3}] - P'_{n_z} \left(\mathcal{C}_{2,z}(t_2) \otimes \mathcal{C}_{2,z}(t_3 - t_1)\right) \\
&\quad - \mathcal{C}_{2,z}(t_1) \otimes \mathcal{C}_{2,z}(t_2 - t_3) - P_{n_z} \left(\mathcal{C}_{2,z}(t_3) \otimes \mathcal{C}_{2,z}(t_1 - t_2)\right),
\end{aligned}$$

where $P_{n_z} = I_{n_z} \otimes U_{n_z^2 \times n_z}$ and $U_{n_z^2 \times n_z}$ is a $(n_z^3 \times n_z^3)$ permutation matrix with unity entries in elements $[(i-1)n_z + j, (j-1)n_z^2], i = 1, \ldots, n_z^2$ and $j = 1, \ldots, n_z$, and zeros else. Here, I adopt the notation of Swami and Mendel (1990) and store all product-moments of a mean zero vector-valued process in a vector using Kronecker products.[16] There is an intimate relationship between moments and cumulants: If two probability distributions have identical moments, they will have identical cumulants as well. In particular, the second cumulant is equal to the autocovariance matrix and the third cumulant to the autocoskewness matrix. The fourth-order cumulant, however, is the fourth-order product-moment (autocokurtosis matrix) less permutations of second-order moments. In general, for cumulants above three, lower-order cumulants need to be known. Nevertheless, using cumulants is preferable for several reasons. For instance, all cumulants of a Gaussian process of order three and above are zero, whereas the same applies only to odd product-moments. Furthermore, the cumulant of two statistically independent random processes equals the sum of the cumulants of the individual processes (which is not true for higher-order moments). And lastly, cumulants of a white noise sequence, such as ξ_t, are Kronecker delta functions, so that their polyspectra are flat (Mendel, 1991).[17]

[16] For example, the second moments of $\widetilde{z}_t$ can either be stored in a $n_z \times n_z$ matrix $E(\widetilde{z}_t \cdot \widetilde{z}'_t) =: \Sigma_z$ or in the $n_z^2 \times 1$ vector $E(\widetilde{z}_t \otimes \widetilde{z}_t) = vec(\Sigma_z)$; this notion naturally carries over to higher orders.

[17] For a mathematical discussion of using cumulants instead of moments in terms of ergodicity and proper functions, see Brillinger (1965).

Assuming that $\mathcal{C}_{k,z}(t_1, \ldots, t_{k-1})$ is absolutely summable, the kth-order polyspectrum $\mathcal{S}_{k,z}$ is defined as the (k-1)-dimensional Fourier transform of the kth-order cumulant

$$\mathcal{S}_{k,z}(\omega_1, .., \omega_{k-1}) := (2\pi)^{1-k} \sum_{t_1=-\infty}^{\infty} .. \sum_{t_{k-1}=-\infty}^{\infty} \mathcal{C}_{k,z}(t_1, .., t_{k-1}) \cdot e^{-i \sum_{j=1}^{k-1} \omega_j t_j},$$

with $\omega_j \in [-\pi; \pi]$ and imaginary i (see Swami et al. (1994) for further details). The second-, third- and fourth-order polyspectra are called the power spectrum, bispectrum and trispectrum, respectively. The power spectrum corresponds to the well-studied spectral density, which is a decomposition of the autocorrelation structure of the underlying process (Wiener-Khinchin theorem). The bispectrum can be viewed as a decomposition of the third moments (auto- and cross-skewness) over frequency and is useful for considering systems with asymmetric nonlinearities. In studying symmetric nonlinearities, the trispectrum is a more powerful tool, as it represents a decomposition of (auto- and cross-) kurtosis over frequency. Furthermore, both the bi- and trispectrum will be equal to zero for a Gaussian process, such that departures from Gaussianity will be reflected in these higher-order polyspectra.

Standard results from VAR(1) systems and insights from higher-order statistics can be used regarding the computation of unconditional cumulants and polyspectra. The kth-order cumulants of ξ_t are

$$\mathcal{C}_{k,\xi}(t_1, \ldots, t_{k-1}) = \begin{cases} \Gamma_{k,\xi} & \text{if } t_1 = \cdots = t_{k-1} = 0, \\ 0 & \text{otherwise,} \end{cases}$$

and corresponding polyspectra

$$\mathcal{S}_{k,\xi}(\omega_1, \ldots, \omega_{k-1}) = (2\pi)^{1-k} \Gamma_{k,\xi}$$

are flat.[18]

[18] For the computation of $\Gamma_{k,\xi}$ see also Appendix D.

Letting $[\otimes_{j=1}^{k} X(j)] = X(1) \otimes X(2) \otimes \cdots \otimes X(k)$ for objects $X(j)$, Swami and Mendel (1990) show that the cumulants of the state vector $\widetilde{z}_t$

$$\mathcal{C}_{k,z}(t_1, \ldots, t_{k-1}) = [\otimes_{j=0}^{k-1} A^{t_j}] \cdot \mathcal{C}_{k,z}(0, \ldots, 0)$$

are given in terms of their zero-lag cumulants

$$\mathcal{C}_{k,z}(0, \ldots, 0) = (I_{n_z^k} - [\otimes_{j=1}^{k} A])^{-1} \cdot [\otimes_{j=1}^{k} B] \cdot \Gamma_{k,\xi}$$

which can be computed efficiently using iterative algorithms for generalized Sylvester equations (see Appendix E). Furthermore, there is considerable symmetry (by using appropriate permutation matrices); in particular, all second-order cumulants can be computed from $t_1 > 0$, all third-order cumulants from $t_1 \geq t_2 > 0$ and all fourth-order cumulants from $t_1 \geq t_2 \geq t_3 > 0$. Since there is a linear relationship between $\widetilde{y}_t$ and $\widetilde{z}_{t-1}$ in (4.2), expressions for the kth-order cumulants of observables are obtained in closed-form. That is, for $t_j > 0$

$$\mathcal{C}_{k,y}(0, .., 0) = [\otimes_{j=1}^{k} C]\mathcal{C}_{k,z}(0, .., 0) + [\otimes_{j=1}^{k} D]\Gamma_{k,\xi}, \tag{4.3}$$

$$\mathcal{C}_{k,y}(t_1, .., t_{k-1}) = [\otimes_{j=1}^{k} C]\mathcal{C}_{k,z}(t_1, .., t_{k-1}). \tag{4.4}$$

Stacking all theoretical second, third and fourth cumulants into vectors yields

$$m_2(\theta, T+1) = \left(\mathcal{C}_{2,y}(0)', \quad .., \quad \mathcal{C}_{2,y}(T)'\right)', \tag{4.5}$$

$$m_3(\theta, T+1) = \left(\mathcal{C}_{3,y}(0,0)', \quad .., \quad \mathcal{C}_{3,y}(T,T)'\right)', \tag{4.6}$$

$$m_4(\theta, T+1) = \left(\mathcal{C}_{4,y}(0,0,0)', \quad .., \quad \mathcal{C}_{4,y}(T,T,T)'\right)'. \tag{4.7}$$

Regarding the computation of polyspectra, consider the vector moving average representation (VMA) of $\widetilde{z}_t = \sum_{j=0}^{\infty} A^j B \xi_{t-j}$. Using

equation (4.2) and lag operator L, the VMA for the control variables is given by $\widetilde{y}_t = \sum_{j=0}^{\infty} CA^j B\xi_{t-j-1} + D\xi_t = H_\xi(L^{-1})\xi_t$ with transfer function

$$H_\xi(\mathfrak{z}) = D + C\left(\mathfrak{z}I_{n_z} - A\right)^{-1} B \text{ for } \mathfrak{z} \in \mathbb{C}. \tag{4.8}$$

Setting $\mathfrak{z}_j = e^{-i\omega_j}$, with imaginary i and $\omega_j \in [-\pi;\pi]$, the Fourier transformations of the cumulants of $\widetilde{y}_t$, i.e. the power spectrum $\mathcal{S}_{2,y}$, bispectrum $\mathcal{S}_{3,y}$ and trispectrum $\mathcal{S}_{4,y}$, are obtained as:

$$\mathcal{S}_{2,y} = (2\pi)^{-1}\left[H(\mathfrak{z}_1^{-1}) \otimes H(\mathfrak{z}_1)\right]\Gamma_{2,\xi}, \tag{4.9}$$

$$\mathcal{S}_{3,y} = (2\pi)^{-2}\left[H(\mathfrak{z}_1^{-1}\cdot\mathfrak{z}_2^{-1}) \otimes H(\mathfrak{z}_1) \otimes H(\mathfrak{z}_2)\right]\Gamma_{3,\xi}, \tag{4.10}$$

$$\mathcal{S}_{4,y} = (2\pi)^{-3}\left[H(\mathfrak{z}_1^{-1}\cdot\mathfrak{z}_2^{-1}\cdot\mathfrak{z}_3^{-1}) \otimes H(\mathfrak{z}_1) \otimes H(\mathfrak{z}_2) \otimes H(\mathfrak{z}_3)\right]\Gamma_{4,\xi}. \tag{4.11}$$

Again, there is considerable symmetry easing the computations. To approximate the interval $[-\pi;\pi]$, I divide it into N subintervals to obtain $N+1$ frequency indices with ω_s denoting the sth frequency in the partition. The bispectrum can be computed from $s_1 \leq s_2$ and the trispectrum from $s_1 \leq s_2 \leq s_3$ (s_j=1,...,$N+1$; $j = 1,2,3$), since these determine all other spectra through permutations. The computations of the bispectrum can be accelerated further by noting that the sum $\omega_{s_1} + \omega_{s_2}$ contains many duplicate elements, since $\omega_{s_j} \in [-\pi;\pi]$. Thus, one does not need to do the computations for all $N(N+1)/2$ runs, but rather for a much smaller set. Similarly, there is no need to evaluate all $N(N+1)(N+2)/6$ possible values of $\omega_{s_1} + \omega_{s_2} + \omega_{s_3}$ for the trispectrum but only the unique values. See Chandran and Elgar (1994) for a thorough discussion of principal domains of polyspectra.

4.2 Monte-Carlo analysis

In this section I demonstrate the formulas by a Monte-Carlo analysis using the DSGE model by An and Schorfheide (2007) for a first- and a second-order approximation. It is well known that simulating higher-order moments requires both a large sample size as well as many simulation runs, since one deals with outliers taken to the powers of three and above.[19] Therefore, I simulate 1 000 trajectories of the PSS with 10 000 data points each (after discarding 1 000 points) and using antithetic shocks.[20] I focus on the original parametrization of the model, however, I impose both the Gaussian as well as Student's t distribution with 15 degrees of freedom as the underlying shock process. I then compute the sample variance, skewness and excess kurtosis of the stochastic innovations and observables of each trajectory and average over all Monte-Carlo runs. Lastly, these are compared to their theoretical counterparts using the formulas derived in this chapter.[21] Table 4.1 summarizes the results of a first-order approximation. The empirical variance, skewness and excess kurtosis are very close to their theoretical values no matter which distribution is imposed on the shocks. Table 4.2 summarizes the results for a second-order approximation. Here the discrepancies

[19] Bai and Ng (2005) derive sampling distributions for the coefficients of skewness and kurtosis for serially correlated data. They also assume stationarity up to eighth order and show in a simulation exercise of an AR(1) process that test statistics for skewness have acceptable finite sample size and power, whereas for kurtosis the size distortions are tremendous. See also Bao (2013) on finite sample biases.

[20] In the code one can change all settings regarding the Monte-Carlo framework in a graphical user-interface, i.e. number of trajectories, sample size, burn-in phase and use of antithetic shocks.

[21] Note that the second-order zero-lag cumulant of y_t is equal to the covariance matrix. Skewness can either be computed via standardized product moments or via the ratio of the third zero-lag cumulant and the 1.5th power of the second zero-lag cumulant. Furthermore, excess kurtosis is the fourth zero-lag cumulant normalized by the square of the second-order cumulant.

in the skewness and in particular excess kurtosis are more evident: matching higher-order moments in simulation studies can be hard. Increasing the number of Monte-Carlo runs as well as sample size would on the one hand increase the precision but on the other hand also the computational time. For an applied researcher who uses a trial-and-error approach to match third-order or fourth-order characteristics of an observable variable in a calibration exercise this is unfeasible. Hence, the expressions derived in this chapter are a convenient and fast way to compute higher-order statistics for linear and nonlinear (pruned) DSGE models.

Table 4.1: Higher-order statistics for An/Schorfheide (2007): first-order

	VARIANCE				SKEWNESS				EXCESS KURTOSIS			
	Gaussian		Student-t		Gaussian		Student-t		Gaussian		Student-t	
shocks	(S)	(T)	(S)	(T)	(S)	(T)	(S)	(T)	(S)	(T)	(S)	(T)
ε_R	0.000	0.000	0.000	0.000	0.000	0	0.000	0	-0.002	0	0.539	0.545
ε_g	0.000	0.000	0.000	0.000	-0.000	0	0.000	0	-0.001	0	0.545	0.545
ε_z	0.000	0.000	0.000	0.000	0.000	0	0.000	0	0.003	0	0.539	0.545
obs	(S)	(T)	(S)	(T)	(S)	(T)	(S)	(T)	(S)	(T)	(S)	(T)
YGR	1.207	1.208	1.395	1.393	0.000	0	-0.000	0	-0.002	0	0.341	0.340
$INFL$	7.998	8.003	9.233	9.235	0.000	0	-0.000	0	-0.007	0	0.142	0.152
INT	10.88	10.88	12.55	12.56	-0.000	0	0.000	0	-0.012	0	0.023	0.043

Notes: Simulated (S) and theoretical (T) statistics for observable variables (obs).

Table 4.2: Higher-order statistics for An/Schorfheide (2007): second-order

	VARIANCE				SKEWNESS				EXCESS KURTOSIS			
	Gaussian		Student-t		Gaussian		Student-t		Gaussian		Student-t	
shocks	(S)	(T)	(S)	(T)	(S)	(T)	(S)	(T)	(S)	(T)	(S)	(T)
ε_R	0.000	0.000	0.000	0.000	0.000	0	0.000	0	-0.002	0	0.539	0.545
ε_g	0.000	0.000	0.000	0.000	-0.000	0	0.000	0	-0.001	0	0.545	0.545
ε_z	0.000	0.000	0.000	0.000	0.000	0	0.000	0	0.003	0	0.539	0.545
obs	(S)	(T)	(S)	(T)	(S)	(T)	(S)	(T)	(S)	(T)	(S)	(T)
YGR	1.238	1.238	1.439	1.438	0.285	0.153	0.357	0.218	0.177	0.136	0.660	0.594
$INFL$	8.004	8.010	9.243	9.245	0.101	0.030	0.116	0.041	0.009	0.006	0.166	0.162
INT	10.88	10.89	12.56	12.57	0.080	0.010	0.089	0.014	-0.000	0.001	0.038	0.044

Notes: Simulated (S) and theoretical (T) statistics for observable variables (obs).

Chapter 5

Analytical derivatives

The rank criteria in Chapters 7 and 8 require derivatives of all solution matrices, cumulants and polyspectra with respect to the vector of deep parameters θ. Following ideas from Iskrev (2008) and Schmitt-Grohé and Uribe (2012, Suppl. Mat., Sec. A.3), f as well as the Jacobian of f are functions of θ and of the steady state vector $\overline{xuy}(\theta) := (\overline{x}(\theta)', \overline{u}(\theta)', \overline{y}(\theta)')'$, which is also a function of θ. Thus, implicitly, $f(\overline{xuy}(\theta), \theta) = 0$. Differentiating yields

$$\begin{aligned} \mathtt{d}f := \frac{\partial f(\overline{xuy}(\theta), \theta)}{\partial \theta'} &= \frac{\partial f}{\partial \overline{xuy}'} \frac{\partial \overline{xuy}}{\partial \theta'} + \frac{\partial f}{\partial \theta'} = 0 \\ \Leftrightarrow \frac{\partial \overline{xuy}}{\partial \theta'} &= -\left[\frac{\partial f}{\partial \overline{xuy}'}\right]^{-1} \frac{\partial f}{\partial \theta'}. \end{aligned}$$

This expression can easily be obtained analytically using, for example, MATLAB's symbolic toolbox. The derivative of the Jacobian $\mathcal{D}\overline{f}(\overline{xuy}(\theta), \theta)$ with respect to θ is then given by

$$\mathtt{d}\mathcal{D}f := \frac{\partial vec(\mathcal{D}f(\overline{xuy}(\theta), \theta))}{\partial \theta'} = \frac{\partial vec(\mathcal{D}f)}{\partial \overline{xuy}'} \frac{\partial \overline{xuy}}{\partial \theta'} + \frac{\partial vec(\mathcal{D}f)}{\partial \theta'}.$$

Note that $\mathtt{d}\mathcal{D}f$ can be partitioned

$$\mathcal{D}\bar{f} := \left(\underbrace{\frac{\partial \bar{f}}{\partial x'_{t+1}}, \frac{\partial \bar{f}}{\partial u'_{t+1}}}_{=:f_1}, \underbrace{\frac{\partial \bar{f}}{\partial y'_{t+1}}}_{=:f_2}, \underbrace{\frac{\partial \bar{f}}{\partial x'_{t}}, \frac{\partial \bar{f}}{\partial u'_{t}}}_{=:f_3}, \underbrace{\frac{\partial \bar{f}}{\partial y'_{t}}}_{=:f_4} \right)$$

to obtain $\mathtt{d}f_1$, $\mathtt{d}f_2$, $\mathtt{d}f_3$ and $\mathtt{d}f_4$. This approach can be extended to calculate the analytical derivative of the Magnus-Neudecker Hessian with respect to θ, since $H := \mathcal{H}f(\overline{xuy}(\theta), \theta)$:[22]

$$\mathtt{d}H := \frac{\partial vec(\mathcal{H}f(\overline{xuy}(\theta), \theta))}{\partial \theta'} = \frac{\partial vec(\mathcal{H}f)}{\partial \overline{xuy}'} \frac{\partial \overline{xuy}}{\partial \theta'} + \frac{\partial vec(\mathcal{H}f)}{\partial \theta'}.$$

The accompanying MATLAB code writes all analytical derivatives, using symbolic expressions, into script files for further evaluation. For numerical derivatives, a two-sided central difference method is employed as outlined in Appendix F. Note that from now on I use the following notation: $\mathtt{d}X := \frac{\partial vec(X)}{\partial \theta'}$ for the derivative of a matrix. Furthermore, I repeatedly use the commutation matrix $K_{m,n}$ which transforms the $m \times n$ matrix A, such that $K_{m,n} vec(A) = vec(A')$[23], and the following useful results from matrix differential calculus:

Theorem 1 (Derivative of products)**.** *Let A be a $(m \times n)$ matrix, B a $(n \times o)$ matrix, C a $(o \times p)$ matrix and D a $(p \times q)$ matrix, then the derivative of $vec(ABCD)$ with respect to θ is given by*

$$\begin{aligned} \mathtt{d}(ABCD) = (D'C'B' \otimes I_m)\,\mathtt{d}A + (D'C' \otimes A)\,\mathtt{d}B \\ + (D' \otimes AB)\,\mathtt{d}C + (I_q \otimes ABC)\,\mathtt{d}D \end{aligned}$$

Proof. Magnus and Neudecker (1999, p. 175). □

[22] Another (faster) approach is based on generalized Sylvester equations in the manner of Ratto and Iskrev (2012).

[23] See Magnus and Neudecker (1999, p. 46) for the definition and Magnus and Neudecker (1999, p. 182) for an application regarding derivatives.

Theorem 2 (Derivative of Kronecker products)**.** *Let X be a $(n \times q)$ matrix, Y a $(p \times r)$ matrix and $K_{r,n}$ the commutation matrix of order (r, n), then the derivative of $vec(X \otimes Y)$ with respect to θ is given by*

$$\mathbf{d}(X \otimes Y) = (I_q \otimes K_{r,n} \otimes I_p)\left[(I_{nq} \otimes vec(Y))\,\mathbf{d}X + (vec(X) \otimes I_{pr})\,\mathbf{d}Y\right]$$

Proof. Magnus and Neudecker (1999, p. 185). □

Moreover, I make use of the following algorithm:

Algorithm 1 (Derivative of partitioned matrix)**.** *Let X be a $(m \times n)$ matrix, partitioned such that $X = \begin{bmatrix} X_1 & X_2 \end{bmatrix}$, with X_1 being $(m \times n_1)$ and X_2 being $(m \times n_2)$, $n = n_1 + n_2$.*

1. *Derive $\mathbf{d}X_1$ and $\mathbf{d}X_2$; $\mathbf{d}X_1$ is of dimension $(mn_1 \times n_\theta)$ and $\mathbf{d}X_2$ of dimension $(mn_2 \times n_\theta)$.*

2. *For $i = 1, \ldots, n_\theta$*

   (a) *Denote the ith column of $\mathbf{d}X_1$ and $\mathbf{d}X_2$ as $\mathbf{d}X_1^i$ and $\mathbf{d}X_2^i$ respectively. $\mathbf{d}X_1^i$ is of dimension $(mn_1 \times 1)$ and $\mathbf{d}X_2^i$ of dimension $(mn_2 \times 1)$.*

   (b) *Reshape $\mathbf{d}X_1^i$ into a $(m \times n_1)$ matrix $\underset{(m \times n_1)}{[\mathbf{d}X_1^i]}$ and $\mathbf{d}X_2^i$ into a $(m \times n_2)$ matrix $\underset{(m \times n_2)}{[\mathbf{d}X_2^i]}$.*

   (c) *Store $vec\begin{pmatrix} \underset{(m \times n_1)}{[\mathbf{d}X_1^i]} & \underset{(m \times n_2)}{[\mathbf{d}X_2^i]} \end{pmatrix}$ into the ith column of a matrix $\mathbf{d}X$.*

3. *$\mathbf{d}X$ is the derivative of X with respect to θ and is of dimension $(mn \times n_\theta)$.*

Derivatives of first-order solution matrices Let $n_v = n_x + n_u$, $n = n_v + n_y$, $K_{n,q}$ be the commutation matrix of order (n, q),

$$F = -(h_v' g_v' \otimes I_n)\mathtt{d}f_2 - (h_v' \otimes I_n)\mathtt{d}f_1 - (g_v' \otimes I_n)\mathtt{d}f_4 - \mathtt{d}f_3,$$

$g_v = [g_x \; g_u]$ and $h_v = \begin{pmatrix} h_x & h_u \\ 0 & 0 \\ n_u \times n_x & n_u \times n_u \end{pmatrix}$, then the derivatives of the first-order solution matrices are given by:

$$\begin{bmatrix} \mathtt{d}g_v \\ \mathtt{d}h_v \end{bmatrix} = \Big[(h_v' \otimes f_2) + (I_{n_v} \otimes f_4) \quad (I_{n_v} \otimes f_2 g_v) + (I_{n_v} \otimes f_1)\Big]^{-1} \cdot F,$$
$$\mathtt{d}g_v' = K_{n_y,n_v}\mathtt{d}g_v, \quad \mathtt{d}h_v' = K_{n_v,n_v}\mathtt{d}h_v,$$
$$\mathtt{d}h_v^t = (I_{n_v} \otimes (h_v)^{t-1})\mathtt{d}h_v + (h_v' \otimes I_{n_v})\mathtt{d}(h_v^{t-1}), \qquad t \geq 2.$$

See Schmitt-Grohé and Uribe (2012, Suppl. Mat., Sec. A.3) for the derivation of these results. Since I use indices to keep track of terms belonging to states and shocks in h_v and g_v, it is straightforward to compute $\mathtt{d}h_x, \mathtt{d}h_u, \mathtt{d}g_x$ and $\mathtt{d}g_u$ by simply selecting the corresponding rows of $\mathtt{d}h_v$ and $\mathtt{d}g_v$ accordingly.

Derivatives of second-order solution matrices Differentiating (2.7) with respect to θ requires the analytical derivatives of Q^{-1}, R, S^{-1} and T. See Appendix B for the derivation of these objects. Then, the analytical derivatives of the second-order solution matrices with respect to θ can be summarized as

$$\mathtt{d}\begin{bmatrix} vec(g_{vv}) \\ vec(h_{vv}) \end{bmatrix} = -Q^{-1}\mathtt{d}R - (vec(R)' \otimes I_{nn_v^2})\mathtt{d}(Q^{-1}),$$
$$\mathtt{d}\begin{bmatrix} h_{ss} \\ g_{ss} \end{bmatrix} = -(T' \otimes I_n)\mathtt{d}(S^{-1}) - S^{-1}\mathtt{d}T.$$

The Jacobians of G_{xx}, G_{xu}, G_{ux}, G_{uu}, H_{xx}, H_{xu}, H_{ux} and H_{uu} are simple permutations of the rows in $\mathtt{d}g_{vv}$ and $\mathtt{d}h_{vv}$, see Appendix C for an example.

Derivatives of pruned state-space solution matrices Differentiating A, B, C, D, c and d with respect to θ is a straightforward application of Algorithm 1 for partitioned matrices. This requires the analytical derivatives of first- and second-order solution matrices (see above), as well as of Σ, which is given analytically by the model.

Derivatives of means Differentiating the means of z_t (2.14) and y_t (2.16) with respect to θ requires the analytical derivatives of the pruned state-space solution matrices:

$$\mathtt{d}\mu_z = \left([(I_{n_z} - A)']^{-1} \otimes (I_{n_z} - A)^{-1}\right) \mathtt{d}A + (I_{n_z} - A)^{-1}\mathtt{d}c,$$
$$\mathtt{d}\mu_y = \mathtt{d}\bar{y} + C\mathtt{d}\mu_z + (\mu_z' \otimes I_{n_y})\mathtt{d}C + \mathtt{d}d.$$

Derivatives of cumulants Differentiating the cumulants for y_t of order k, requires $\mathtt{d}\Sigma_\xi$ (given in Appendix B), derivatives of Kronecker products of the solution matrices (due to Theorem 2), and the Jacobians of the zero-lag cumulants, $\mathtt{d}\mathcal{C}_{k,z}(0,\ldots,0)$, for which I make use of generalized Sylvester equations as outlined in Appendix E. The analytical derivatives of the cumulants in equations (4.3) and (4.4) are then a simple application of Theorem 1.

Derivatives of polyspectra Differentiating $\mathcal{S}_{k,y}$ with respect to θ in equations (4.9), (4.10) and (4.11) analytically, I divide the interval $[-\pi;\pi]$ into N subintervals to obtain $N+1$ frequency indices with ω_s denoting the sth frequency in the partition. The following steps can be conducted simultaneously using parallel computation: For each $\mathfrak{z}_j = e^{-i\omega_{s_j}} (j = 1,\ldots,k-1; s_j = 1,\ldots N+1, s_1 \leq \cdots \leq s_{k-1}, \text{imaginary } i)$ I first compute the derivative of $H_\xi(\mathfrak{z}_j)$ and its conjugate transpose, using the expression in Appendix B. The Jacobians $\mathtt{d}\mathcal{S}_{k,y}(\omega_1,\ldots,\omega_{k-1})$ then follow due to Theorem 1.

Chapter 6

Assumptions and definitions of local identifiability

For the identification methods in the following chapters it is important to first state the underlying assumptions and definitions of local identifiability.

Assumption 1 (Determinacy)**.** *Let Θ be the parameter space that yields the determinacy region of the DSGE model.*

Assumption 2 (Stationary up to order four)**.** *The vector y_t (t=1,...,T) is stationary to at least order four.*

The first assumption is standard in the DSGE literature due to the rational expectation hypothesis, see Milani (2012) for a discussion. The second assumption needs some clarification. It requires observables to have finite and constant first, second, third and fourth moments, that only depend on the time difference but not on time itself. This is basically an extension of the usual covariance stationarity assumption. The literature on ARCH(∞) discusses some

practical aspects of fourth-order stationarity (see e.g. Teyssiére and Kirman (2011, Ch. 1) and the references therein). The HOSA toolbox for MATLAB provides guidance on the computational aspects of sample estimates and testing constancy of higher-order moments, cumulants and polyspectra.

Definition 1 (Identifiability from first four moments)**.** *$\theta_0 \in \Theta$ is said to be locally identifiable from the first four moments (or cumulants) of y_t, if there is an open neighborhood of θ_0 in which*

$$\begin{aligned} &\mu_y(\theta_0) = \mu_y(\theta_1), && m_3(\theta_0, T) = m_3(\theta_1, T), \\ &m_2(\theta_0, T) = m_2(\theta_1, T), && m_4(\theta_0, T) = m_4(\theta_1, T), \end{aligned}$$

imply $\theta_0 = \theta_1$ for any $\theta_1 \in \Theta$.

Definition 2 (Identifiability from mean and polyspectra)**.** *$\theta_0 \in \Theta$ is said to be locally identifiable from the mean, power spectrum, bispectrum and trispectrum of y_t, if there is an open neighborhood of θ_0 in which*

$$\begin{aligned} &\mu_y(\theta_0) = \mu_y(\theta_1), && \mathcal{S}_{3,y}(\omega_1, \omega_2; \theta_0) = \mathcal{S}_{3,y}(\omega_1, \omega_2; \theta_1), \\ &\mathcal{S}_{2,y}(\omega_1; \theta_0) = \mathcal{S}_{2,y}(\omega_1; \theta_1), && \mathcal{S}_{4,y}(\omega_1, \omega_2, \omega_3; \theta_0) = \mathcal{S}_{4,y}(\omega_1, \omega_2, \omega_3; \theta_1), \end{aligned}$$

for all $\omega_j \in [-\pi, \pi]$ (j=1,2,3) imply $\theta_0 = \theta_1$ for any $\theta_1 \in \Theta$.

I follow Deistler (1976) and define identifiability as a concept in moments, cumulants and polyspectra, not in probability laws.[24] I call $\theta_0 \in \Theta$ and $\theta_1 \in \Theta$ observationally equivalent (with respect to data $\{y_t\}$), if they generate the same first four moments, cumulants or polyspectra of y_t. Definition 1 states identifiability for a finite number of moment conditions in the time domain (similar to Iskrev (2010a)), whereas Definition 2 corresponds to the complete set of dynamic properties in the frequency domain (similar to Qu and Tkachenko (2012)).

[24] See Deistler and Seifert (1978) for a thorough discussion of identifiability and estimability.

Chapter 7

Identification of linearized DSGE models: a review and departure from Gaussianity

Reconsider the model framework of Chapter 2 and Proposition 1 for a first-order approximation:

$$\hat{x}_{t+1} = h_x \hat{x}_t + h_u u_{t+1}, \tag{7.1}$$

$$\hat{y}_{t+1} = g_x \hat{x}_t + g_u u_{t+1}. \tag{7.2}$$

I will refer to this model framework as the *linearized DSGE model* with states x_t, controls y_t and stochastic innovations u_t. A hat denotes deviations from steady state, e.g. $\hat{x}_{t+1} = x_{t+1} - \bar{x}$. Note that in a first-order approximation there is certainty-equivalence, i.e. the unconditional means are equal to their steady state values. For the sake of notation, all control variables are assumed to be observable. The vector of stochastic innovations u_t has $E(u_t) = 0$

and finite covariance matrix $E(u_t u_t') =: \Sigma$. Furthermore, u_t is at least 4th-order white noise with finite moments, which implies y_t is at least a 4th-order stationary process. $\Gamma_{k,u}$ denotes the kth-order cumulant of u_t and $\mathcal{S}_{k,y}$ the kth-order polyspectrum of y_t.

Several formal methods have been proposed to check identification in linearized DSGE models via (i) observational equivalent first and second moments (Iskrev, 2010a), (ii) observational equivalent spectral densities (Qu and Tkachenko, 2012), (iii) implications from control theory for observational equivalent spectral densities in minimal systems (Komunjer and Ng, 2011) and (iv) Bayesian indicators (Koop et al., 2013). The first three approaches derive rank conditions similar to the classical literature on identification, whereas the fourth approach uses insights from Bayesian asymptotic theory. Even though all methods seem similar, there has been – to my knowledge – no study of the advantages and drawbacks of implementing the different methods. Furthermore, one important limitation is inherent to all methods: they focus on first and second moments (or equivalently the mean and power spectrum) only. This is due to the fact that the majority of papers which use linearized DSGE models assume Gaussian innovations, and in linear Gaussian models the first two moments completely characterize the data. Furthermore, most estimation methods in the econometric DSGE literature – e.g. full-information likelihood methods or limited-information methods like impulse-response matching or GMM – exploit information from the first two moments of data only. However, Curdia et al. (2014) and Chib and Ramamurthy (2014) show that using Student's t-distributed errors in otherwise standard linearized DSGE models outperforms the corresponding Gaussian models. From an identification point of view this implies that higher moments, cumulants and polyspectra can help with the estimation and identification of linearized non-Gaussian DSGE models. To this end, the contribution of this chapter is threefold.

First, it provides a comprehensive review of all methods. Methodically, I first derive all criteria in the same framework and model representation of Chapter 2. Second, using results from Chapter 4, I extend the criteria of Iskrev (2010a) and Qu and Tkachenko (2012) in order to additionally check identification from higher-order moments in linearized DSGE models with non-Gaussian innovations. Third, the comparison and extension are demonstrated by means of two linearized DSGE models that are known to have lack of identification in their Gaussian version.

Computationally, the methods heavily depend on the accuracy of computing derivatives and ranks as well as appropriate optimization and MCMC methods. In particular, while Iskrev (2010a) already uses analytical derivatives, Komunjer and Ng (2011) and Qu and Tkachenko (2012) rely on numerical methods. Therefore, for a rigorous comparison, I exploit ideas and expressions of Chapter 5 to discuss the effect of analytical compared to numerical derivatives for each method based on ranks. For the Bayesian learning rate indicator I address different optimization and MCMC techniques and how these influence the findings. I argue that in theory the methods should yield the same results, however, in practice the criteria can differ due to numerical issues. Therefore, I propose a robust procedure, composed of all methods, to pinpoint (sets of) non-identified as well as weakly-identified parameters. The suggested approach is a useful tool for applied researchers to get robust insight into the identifiability of a linearized DSGE model before taking it to data.

7.1 Rank conditions

All methods based on ranks exploit the dynamic structure of the linearized DSGE model to define mappings and to establish conditions for local injectivity of the mappings. The point of departure, however, is different: Iskrev (2010a)'s approach is based in the time

domain, whereas Qu and Tkachenko (2012) derive conditions in the frequency domain. Komunjer (2012)'s approach can be considered to be in between both approaches, since they establish conditions without actually computing autocovariances or the spectral density. I will now state the assumptions, objectives and proofs of each method in the form of propositions. The corresponding remarks contain additional intuition regarding the derivation and lay out possible extensions. Lastly, I compare the methods theoretically and give some guidance on computational aspects.

7.1.1 Time domain

Iskrev (2010a)'s method checks whether the derivatives of the mean and the predicted autocovariogram of the observables with respect to the vector of identifiable parameters has a rank equal to the number of identifiable parameters. The basic idea is to check whether the mapping from the parameter vector θ to the vector of theoretical first two moments (2.16) and (4.5) is injective. In Proposition 2 I state this formally and extend this idea to higher-order cumulants up to fourth order.

Proposition 2 (Iskrev linearized DSGE)**.** *Consider Assumptions 1 and 2 for the linearized DSGE model given in equations* (7.1) *and* (7.2)*. Let $q \leq T$ and stack the mean* (2.16) *and cumulants* (4.5)*,* (4.6) *and* (4.7) *into a vector*

$$\overline{m}(\theta, q) := \begin{pmatrix} \mu_y' & m_2(\theta, q)' & m_3(\theta, q)' & m_4(\theta, q)' \end{pmatrix}'.$$

Assume that $\overline{m}(\theta, q)$ is a continuously differentiable function of $\theta \in \Theta$. Let $\theta_0 \in \Theta$ be a regular point. Furthermore, assume there is an open neighborhood of θ_0 in which $\overline{m}(\theta_0, q)$ has a constant rank. Then θ is locally identifiable at a point θ_0 from the first

four cumulants (or moments) of y_t, if and only if for $q \leq T$

$$\overline{M}(q) := \frac{\partial \overline{m}(\theta_0, q)}{\partial \theta'}$$

has a full column rank equal to the number of parameters.

Proof. Follows Theorem 2 in Iskrev (2010a) and Theorem 6 in Rothenberg (1971). □

Proposition 2 checks whether the first four moments of the linearized DSGE model are uniquely determined by the deep parameters, given a finite number of lags. It therefore corresponds to Definition 1, i.e. identifiability from the first four cumulants or moments given a finite number of moment conditions. This immediately gives rise to a necessary condition: the number of identifiable parameters does not exceed the dimension of $\overline{m}(\theta, T)$. The criteria can also be used for conditional identification, that is, identification for only a subset of parameters. It is also possible to check whether the parameters are identifiable through (i) the mean and second or (ii) the mean and third or (iii) the mean and fourth moments of observables, separately. I denote the corresponding matrices as $\overline{M}_2(q)$, $\overline{M}_3(q)$ and $\overline{M}_4(q)$. Note that, given the Gaussian distribution for u_t, the proposition reduces to the original Theorem 2 of Iskrev (2010a), since all higher-order cumulants are zero in this case. Iskrev (2010a, Corollary 1) also proposes a different necessary condition, that is, checking injectivity of the mapping from the deep parameters to the solution matrices. To do so, all elements of the steady state, the solution matrices as well as all parameters of the stochastic innovations that depend on θ are stacked into a vector

$$\overline{\tau}(\theta) := \begin{pmatrix} \bar{y}' & vec(h_x)' & vec(h_u)' & vec(g_x)' & vec(g_u)' & \Gamma_{2,u}' & \Gamma_{3,u}' & \Gamma_{4,u}' \end{pmatrix}'.$$

Consider the factorization $\overline{M}(q) = \frac{\partial \overline{m}(\theta,q)}{\partial \overline{\tau}(\theta)'} \frac{\partial \overline{\tau}(\theta)}{\partial \theta'}$. An immediate corollary implies that a point θ_0 is locally identifiable, only if the

rank of $\overline{J} := \frac{\partial \overline{\tau}(\theta_0)}{\partial \theta'}$ at θ_0 is equal to n_θ. This condition is, however, only necessary, because $\overline{\tau}$ may be unidentifiable. Lastly, in the applications I will also exclude the mean restrictions. Therefore, J, $M(q)$, $M_2(q)$, $M_3(q)$ and $M_4(q)$ denote the corresponding Jacobians without the mean restrictions.

7.1.2 Frequency domain

Similar to Iskrev (2010a)'s approach, Qu and Tkachenko (2012) focus on the dynamic structure of the DSGE model, but they work in the frequency domain. Their criteria is therefore based on injectivity of the mapping from θ to the mean (2.16) and to the spectral density (4.9). Proposition 3 states this formally. I extend their ideas and check whether the mean, power spectrum, bispectrum and trispectrum of the observables are uniquely determined by the deep parameters at all frequencies.

Proposition 3 (Qu & Tkachenko linearized DSGE)**.** *Consider Assumptions 1 and 2 for the linearized DSGE model given in equations* (7.1) *and* (7.2)*. Assume that the power spectrum* (4.9)*, bispectrum* (4.10) *and trispectrum* (4.11) *are continuous in* $\omega \in [-\pi;\pi]$ *and continuous and differentiable in* $\theta \in \Theta$*. Let*

$$\overline{G}(\theta) = \mathbb{d}(\mu_y(\theta))' \, \mathbb{d}(\mu_y(\theta)) + \int_{-\pi}^{\pi} \mathbb{d}(\mathcal{S}_{2,y}(\omega_1;\theta))^* \, \mathbb{d}(\mathcal{S}_{2,y}(\omega_1;\theta)) \, d\omega_1$$
$$+ \int_{-\pi}^{\pi}\int_{-\pi}^{\pi} \mathbb{d}(\mathcal{S}_{3,y}(\omega_1,\omega_2;\theta))^* \, \mathbb{d}(\mathcal{S}_{3,y}(\omega_1,\omega_2;\theta)) \, d\omega_1 d\omega_2$$
$$+ \int_{-\pi}^{\pi}\int_{-\pi}^{\pi}\int_{-\pi}^{\pi} \mathbb{d}(\mathcal{S}_{4,y}(\omega_1,\omega_2,\omega_3;\theta))^* \, \mathbb{d}(\mathcal{S}_{4,y}(\omega_1,\omega_2,\omega_3;\theta)) \, d\omega_1 d\omega_2 d\omega_3$$

and $\theta_0 \in \Theta$ *be a regular point. Furthermore, assume there is an open neighborhood of* θ_0 *in which* $\overline{G}(\theta_0)$ *has a constant rank. Then,* θ *is locally identifiable at a point* θ_0 *from the mean, power spectrum, bispectrum and trispectrum of* y_t*, if and only if* $\overline{G}(\theta_0)$ *is nonsingular, i.e. its rank is equal to the number of parameters.*

Proof. Follows Theorem 2 in Qu and Tkachenko (2012) and Theorem 1 in Rothenberg (1971). Note that I use the complex conjugate $*$, since the polyspectra are in general complex matrices. $\mathtt{d}\mathcal{S}_{k,y}^{*}\mathtt{d}\mathcal{S}_{k,y}$ is a Gram matrix, therefore it is Hermitian and positive semidefinite. Furthermore, there is an isomorphism between complex and real matrices such that the $(n_y^k \times 1)$ vector $\mathcal{S}_{k,y}$ can be transformed into a $(2n_y^k \times 2)$ real matrix $\mathcal{S}_{k,y}^{R}$ (see Brillinger (2001, p. 71) and Pintelon and Schoukens (2001, p. 553)). The following equivalence holds: $\mathcal{S}_{k,y}^{*}\mathcal{S}_{k,y} \Leftrightarrow \mathcal{S}_{k,y}^{R'}\mathcal{S}_{k,y}^{R}$. Furthermore $2 \cdot rank(\mathcal{S}_{k,y}^{*}\mathcal{S}_{k,y}) = rank(\mathcal{S}_{k,y}^{R'}\mathcal{S}_{k,y}^{R})$. The same is true if one considers the differential of $\mathcal{S}_{k,y}$ with respect to θ_j. The proof requires $rank(\mathtt{d}\mathcal{S}_{k,y}^{R'}\mathtt{d}\mathcal{S}_{k,y}^{R})$ to be nonsingular, i.e. full rank, for θ_0 to be locally identified. This is equivalent to $rank(\mathtt{d}\mathcal{S}_{k,y}^{*}\mathtt{d}\mathcal{S}_{k,y}) = n_\theta$. □

This proposition corresponds to Definition 2, i.e. identifiability from the complete set of dynamic properties. Qu and Tkachenko (2012) provide several extensions. In particular, identification can be checked from a subset of frequencies only, conditional on other parameters being fixed, or including general constraints on the parameters. Moreover, one can check whether the parameters are identifiable through the mean and individual spectra. I denote the corresponding matrices as $\overline{G}_2(\theta_0)$, $\overline{G}_3(\theta_0)$ and $\overline{G}_4(\theta_0)$. Given the Gaussian distribution for u_t, this proposition reduces to the original Theorem 2 of Qu and Tkachenko (2012), since the bi- and trispectrum are zero in this case. Note that Qu and Tkachenko (2012) use numerical derivatives, whereas I am able to use analytical derivatives due to the results of Chapter 5. For both cases, however, it is still necessary to divide the interval $[-\pi; \pi]$ into sufficient subintervals to numerically approximate the integrals. Lastly, in the applications I will also exclude the mean restrictions. Therefore, $G(\theta_0)$, $G_2(\theta_0)$, $G_3(\theta_0)$ and $G_4(\theta_0)$ denote the corresponding Jacobians without the mean restrictions. The dimension, however, is always $n_\theta \times n_\theta$.

7.1.3 Minimal system

Based upon results from control theory for minimal systems Komunjer and Ng (2011) derive restrictions implied by equivalent spectral densities (or equivalent autocovariances) without actually computing them as in Iskrev (2010a) or Qu and Tkachenko (2012).[25] Formally, consider the *minimal linearized DSGE model*, i.e. the model which dynamics are entirely driven by the exogenous states $x_{2,t}$ and the stochastic innovations u_t:

$$\hat{x}_{2,t} = \widetilde{h}_x \hat{x}_{2,t-1} + \widetilde{h}_u u_t, \tag{7.3}$$

$$\hat{y}_t = \widetilde{g}_x \hat{x}_{2,t-1} + \widetilde{g}_u u_t. \tag{7.4}$$

This minimal system has the smallest possible dimension n_{x_2} of the state vector that is able to capture all dynamics and has the familiar state-space representation. As DSGE models are based upon microfoundations, n_{x_2} is not hard to determine for small and medium-sized DSGE models (e.g. the variables which columns in h_x consist only of zeros are endogenous states).[26]

Proposition 4 (Komunjer & Ng linearized DSGE)**.** *Consider Assumptions 1 and 2 for the minimal linearized DSGE model given in equations* (7.3) *and* (7.4)*. Let* $\theta_0 \in \Theta$ *and* $\theta_1 \in \Theta$ *be regular points. Set up a* $n_{x_2} \times n_{x_2}$ *similarity transformation matrix* $\mathcal{T}$ *and a* $n_u \times n_u$ *full column rank matrix* $\mathcal{U} = L_u(\theta_0)\mathcal{V}L_u(\theta_1)^{-1}$

[25] In fact, they have two conditions depending on the relation between the number of shocks and observables. I will focus on singular and squared systems ($n_u \leq n_y$) and assume fundamental innovations. For instance, in a model with anticipated shocks one has to use the innovation representation of the model and derive a slightly different rank condition by reparametrizing the solution system. Since the innovation representation depends on the existence of a discrete algebraic Ricatti equation, it is not possible to derive analytical derivatives. Moreover, in the commonly used squared case ($n_u = n_y$) both conditions coincide.

[26] For the derivation of this model representation, formal conditions and some practical issues regarding the minimal state vector see Appendix G.

such that $\widetilde{h}_x(\theta_1) = \mathcal{T}\widetilde{h}_x(\theta_0)\mathcal{T}^{-1}$, $\widetilde{h}_u(\theta_1) = \mathcal{T}\widetilde{h}_u(\theta_0)\mathcal{U}$, $\widetilde{g}_x(\theta_1) = \widetilde{g}_x(\theta_0)\mathcal{T}^{-1}$, $\widetilde{g_u}(\theta_1) = \widetilde{g_u}(\theta_0)\mathcal{U}$, *and* $\Sigma(\theta_1) = \mathcal{U}^{-1}\Sigma(\theta_0)\mathcal{U}^{-1'}$ *with* L_u *being the Cholesky decomposition of* $\Sigma = L_u L_u'$ *and* $\mathcal{V}$ *a constant matrix such that* $\mathcal{V}\mathcal{V}' = I$. *Assume that the mapping* $\delta : \Theta \times \mathbb{R}^{n_{x_2}^2} \times \mathbb{R}^{n_u^2} \to \mathbb{R}^{n_\Delta}$ *is continuously differentiable in* $\theta \in \Theta$ *with* $n_\Delta := (n_{x_2} + n_y)(n_{x_2} + n_u) + n_u(n_u + 1)/2$. *Let*

$$
\begin{aligned}
\overline{\Delta}(\theta_0) &:= \begin{pmatrix} \frac{\partial \mu_y(\theta_0)}{\partial \theta'} & 0_{n_y \times n_{x_2}^2} & 0_{n_y \times n_u} \\ \frac{\partial \delta(\theta_0, I_{n_{x_2}}, I_{n_u})}{\partial \theta'} & \frac{\partial \delta(\theta_0, I_{n_{x_2}}, I_{n_u})}{\partial vec(T)'} & \frac{\partial \delta(\theta_0, I_{n_{x_2}}, I_{n_u})}{\partial vec(U)'} \end{pmatrix} \\
&= \begin{pmatrix} \frac{\partial \mu_y(\theta_0)}{\partial \theta'} & 0_{n_y \times n_{x_2}^2} & 0_{n_y \times n_u} \\ \frac{\partial vec(\widetilde{h}_x)}{\partial \theta'} & \widetilde{h}_x' \otimes I_{n_{x_2}} - I_{n_{x_2}} \otimes \widetilde{h}_x & 0_{n_{x_2}^2 \times n_u^2} \\ \frac{\partial vec(\widetilde{h}_u)}{\partial \theta'} & \widetilde{h}_u' \otimes I_{n_{x_2}} & I_{n_u} \otimes \widetilde{h}_u \\ \frac{\partial vec(\widetilde{g}_x)}{\partial \theta'} & -I_{n_{x_2}} \otimes \widetilde{g}_x & 0_{n_y n_{x_2} \times n_u^2} \\ \frac{\partial vec(\widetilde{g}_u)}{\partial \theta'} & 0_{n_y n_u \times n_{x_2}^2} & I_{n_u} \otimes \widetilde{g}_u \\ \frac{\partial vech(\Sigma)}{\partial \theta'} & 0_{(n_u(n_u+1)/2) \times n_{x_2}^2} & -2\Xi_{n_u}[\Sigma \otimes I_{n_u}] \end{pmatrix} \\
&=: \begin{pmatrix} \frac{\partial \mu_y(\theta_0)}{\partial \theta'} & 0_{n_y \times n_{x_2}^2} & 0_{n_y \times n_u} \\ \Delta_\Lambda(\theta_0) & \Delta_T(\theta_0) & \Delta_U(\theta_0) \end{pmatrix}
\end{aligned}
$$

with Ξ_{n_u} *being the left-inverse of the* $n_u^2 + n_u(n_u+1)/2$ *duplication matrix* $\mathcal{G}_{n_u}$ *for* $vech(\Sigma)$. *Assume there is an open neighborhood of* θ_0 *in which* $\overline{\Delta}(\theta_0)$ *has a constant rank. Then,* θ *is locally identifiable at a point* θ_0 *from the first two moments (or the mean and power spectrum) of* y_t, *if and only if* $\overline{\Delta}(\theta_0)$ *is nonsingular.*

Proof. Follows Proposition 3 in Komunjer and Ng (2011). The proof uses well-established results from control theory for observational equivalence (e.g. Hannan (1971)). The key insight behind the proof is that the spectral density can be factorized and due to left invertibility the matrix $\mathcal{V}$ is not a polynomial matrix of unknown degree, but a constant matrix. □

This proposition corresponds to Definition 2, i.e. identifiability from the complete set of dynamic properties. Formally, θ is locally identifiable from the spectral density of y_t at a point θ_0 if and only if $\delta(\theta, T, U)$ is locally injective at $(\theta_0, I_{n_{x_2}}, I_{n_u})$. A sufficient condition is thus, that the matrix of partial derivatives of $\delta(\theta, T, U)$ has full column rank at $(\theta_0, I_{n_{x_2}}, I_{n_u})$. The intuition behind this mathematical result is that the solution matrices have to be sensitive to changes in parameters. Moreover, equivalent spectral densities (4.9) arise if either (i) for a given size of shocks each transfer function (4.8) is potentially obtained from a multitude of quadruples of solution matrices, or (ii) there are many pairs of transfer functions and variances of shocks that jointly generate the same spectral density. Therefore, the objective in Proposition 4 contains three blocks:

1. The rank of $\Delta_\Lambda(\theta_0)$ must equal n_θ for the reduced-form being locally invertible at θ or, stated differently, for the solution matrices being sensitive to changes in parameters.

2. The rank of $\Delta_T(\theta_0)$ must equal $n_{x_2}^2$ so that the identity matrix is the only local similarity transformation. In other words, there exists a unique quadruple generating the z-Transform for the spectral density.

3. The rank of $\Delta_U(\theta_0)$ must equal n_u^2 so that the spectral factorization is locally uniquely determined. This indicates that there exists a unique pair of z-Transform and dynamic structure that generates the spectral density.

Notice that the mean is incorporated as additional equations for solving the same number of unknowns; hence, changing the order but not the rank condition. The same is true for any other (nonlinear) restrictions, as long as they can be expressed in form of a matrix that augments the rows of the objective function. Put together, this yields a necessary order condition $n_\theta + n_{x_2}^2 + n_u^2 \leq n_\Delta + n_y$

as well as a necessary and sufficient rank condition: $rank(\overline{\Delta}(\theta_0)) = n_\theta + n_{x_2}^2 + n_u^2$. Lastly, in the applications I will also exclude the mean restrictions. Therefore, $\Delta(\theta_0)$ denotes the corresponding Jacobian without the mean restrictions.

7.1.4 Discussion

All propositions derive necessary as well as sufficient conditions for local identification of linearized DSGE models, based on ranks of Jacobians. It is well-established that given some regularity conditions there is a one-to-one mapping between the time and frequency domain, in particular, between the autocovariances and spectral density due to the Wiener-Khinchin theorem. Thus, in theory, all criteria should generally yield the same results. However, they require different settings and leave scope for various numerical errors and imprecisions. Komunjer and Ng (2011)'s assumptions are the strictest as they rely on minimality and left-invertibility. Even though as a consequence of Kalman's decomposition theorem the minimal representation always exists (see e.g. Antsaklis and Michel (2007, sec. 6.2.3)), deriving and checking the minimality of the model can be tedious in practice.[27] Iskrev (2010a)'s and Qu and Tkachenko (2012)'s methods are therefore more straightforward as they only assume the existence of the VMA representation of the DSGE model. Moreover, having the expressions of Chapter 4, I am able to extend their criteria to include restrictions from higher-order cumulants and polyspectra. Strictly speaking Komunjer and Ng (2011)'s as well as Qu and Tkachenko (2012)'s criteria are only valid for an infinite sample and frequencies, since both focus on the complete dynamic structure. Iskrev (2010a)'s approach, on the other hand, is also valid for a finite number of restrictions. This

[27] In a model with non-fundamental innovations ($n_y < n_u$) the solution has to be reparameterized to get the innovation representation of the model yielding even more scope for numerical issues.

gives rise to numerical issues, since choosing the lag order T, as well as the number of subintervals N for the frequencies, may change the results. In practice, however, this is not a question of extremely sensitive results, but rather one of speed: the higher T or N, the more time the calculations need.[28] The different interpretations of Iskrev (2010a)'s and Komunjer and Ng (2011)'s criteria can further be used as diagnostics for model building. For instance, J as well as Δ_Λ check the mapping from the structural parameters to the solution matrices. The evaluation yields parameters that do not influence the reduced-form solution and may thus be obsolete. A researcher is hence able to reparameterize the model prior to estimation. Moreover, given a known shock a rank deficient $\Delta_{\Lambda T} := (\Delta_\Lambda \ \Delta_T)$ indicates that two structures (e.g. two different policies) might cause the same impulse response of the model, so one has to be careful interpreting the importance of shocks. In contrast, given a rank deficient $\Delta_{\Lambda U} := (\Delta_\Lambda \ \Delta_U)$ one cannot be sure, whether it is the size of the shock or a similar propagating mechanism, that yields the same dynamic structure of the model. Qu and Tkachenko (2012)'s test does not give such diagnostics, however, their approach can be used for a quasi maximum likelihood estimation in the frequency domain. Moreover, using a recursive Euler method it is possible to depict observationally equivalent parameters using so-called non-identification curves. These can be used to get insight into the size of the local neighborhood of the unidentified parameters.

[28] In most practical cases, T between 10 and 30 will be sufficient, since the higher the lag, the less informative the identification restrictions. Furthermore, I experimented with different values for N and find that the results in the applications hardly change. The reason is, that if $\theta_0 \in \Theta$ is identified using only a subset of frequencies (small N), it is also identified if considering the full spectrum ($N \to \infty$) (the converse is not true). Therefore, I recommend starting with $N = 10\,000$ for the power spectrum, $N = 1\,000$ for the bispectrum, $N = 100$ for the trispectrum and increase N if the results are unsatisfactory.

I compute all Jacobians analytically as outlined in Chapter 5 or numerically as outlined in Appendix F and then rescale the rows by its largest element in absolute value. For calculating the ranks, I use the singular value decomposition and count the nonzero entries on the main diagonal. Obviously, this requires a specification of the tolerance level, for which I use, on the one hand, a range from 1e-3 to 1e-15, and on the other hand, a robust tolerance level that depends on the size of the matrix ($max(size(X)) \times eps(norm(X))$), which is also MATLAB's default value. Strictly speaking, the criteria are a *yes* or *no* condition. Loosely speaking, however, if a parameter is identified for very large tolerance levels, then it is most likely strongly identified. If it is identified only for very low levels, this may indicate weak identification.[29] In the case of rank deficiency, one is able to pinpoint sets of problematic parameters by analyzing the nullspace. This will be a vector of zeros, if a parameter does not affect the objective at hand. Furthermore, the columns that are linearly dependent indicate that these sets of parameters are indistinguishable. While this approach, followed by Iskrev (2010a), is computationally very fast, I found that in some cases, there were redundancies in the subsets, since larger subsets may include smaller ones and are not pinpointed separately. Thus, similar to Ratto and Iskrev (2011) and Qu and Tkachenko (2012, Corollary 4), a more robust method is to consider the power set and check the criteria for all possible subsets of parameters in a recursive fashion. In my experience, this *brute-force* approach yields more reliable results and is computationally just slightly slower, because, if one finds a subset of parameters that are not identified, one can exclude that subset

[29] Note that this is not based on the literature on strength of identification, but provides only a rough indication for subsets of strongly identified parameters. Nevertheless, these can be used as an initial guess for the methods used, for instance, in Koop et al. (2013) and Qu (2014) to detect weak identification.

from higher-order subsets.[30] In this line of thought, note that all methods depend heavily on the solution matrices and suffer from possible numerical instability of the solution algorithm. However, since I use the same framework and algorithms across methods, this effect is negligible.

Lastly, all procedures check only local identification.[31] Thus, it is necessary to ensure that the identification analysis is valid for a sufficient range of parameters. Therefore, in the applications, I check all criteria, given first a specific point (e.g. calibrated parameters or prior mean) and second, given many draws from a prespecified prior domain of θ that yield a determinate solution. In this way, one has a quasi-global flavor of the rank criteria for linearized DSGE models. Also, most consistent estimators require only local identification for their asymptotic properties to hold. Nevertheless, even if all prior draws are identifiable, the model is still by no means globally identified.

7.2 Bayesian identification criteria

It is well known that informative marginal priors are sufficient to get well-defined posteriors even for non-identifiable parameters: "*In passing it might be noted that unidentifiability causes no real difficulty in the Bayesian approach*" (Lindley, 1972, p. 46). The usual

[30] I implemented both procedures for all criteria in the code.

[31] See Komunjer (2012) for issues regarding global identification. She uses results from unconditional moment restriction models (properness and homeomorphism) to establish identification conditions in the fashion of GMM identification conditions. Another attempt at global identification of DSGE models is made by Qu (2014) who considers a frequency domain expression for the Kullback-Leibler distance between two DSGE models and shows that global identification fails if and only if the minimized distance equals zero. However, both approaches are not very operable especially in terms of computational costs compared to local identification.

approach in the Bayesian literature is to compare the prior and posterior distribution of a parameter. If they differ, there is apparent *learning*, i.e. data seems to be informative about a parameter. However, as was shown by Kadane (1974) and Poirier (1998) this can be misleading, since this is only valid if the priors are independent and the parameter space is a product space. In other words, if these conditions are not satisfied, then data-based learning about identifiable parameters can spill over onto unidentified ones. Thus, one has to be careful judging identification from apparent learning, i.e. the posterior being different than the prior, a point also emphasized by Canova and Sala (2009) for DSGE models. Therefore, following Poirier (1998, Proposition 2) and Koop et al. (2013), in addition to the usual comparison of posterior and prior, one can exploit the fact that data is always conditionally uninformative for a subset of non-identified parameters given a subset of identified parameters. Let $\theta = (\theta_1, \theta_2) \in \Theta$, $\wp(\theta_1, \theta_2) = \wp(\theta_1|\theta_2)\wp(\theta_2)$ is the joint prior and $\wp(\theta_1|y)$ is the posterior of θ_1. Suppose θ_1 is not identified, such that the likelihood function depends only on θ_2. Then

$$\begin{aligned}\wp(\theta_1|y) &= \int_{\Theta(\theta_2)} \wp(\theta_1|\theta_2, y)\wp(\theta_2|y)d\theta_2 \\ &= \int_{\Theta(\theta_2)} \wp(\theta_1|\theta_2)\wp(\theta_2|y)d\theta_2 = E_{\theta_2|y}\left[\wp(\theta_1|\theta_2)\right].\end{aligned}$$

This result holds even in the case of prior dependence. In other words, it is better practice to compare the properties of the posterior of an unidentified parameter θ_1 to the prior belief conditional on some set of identified parameters. Unfortunately, in practice one first has to find an identified parameter set. Moreover, this indicator is only valid if the reduced-form is identified, which for DSGE models is in general not true. Therefore, I will not follow up on this approach in the applications, since either way the rank criteria already provide sets of non-identified parameters in a non-data driven fashion.

Koop et al. (2013) propose a more interesting indicator using implications of Bayesian asymptotics. Naturally, identification should become better as more data becomes available. With an infinite sample the role of the prior vanishes and Bayesian asymptotics are identical to the asymptotic distribution theory of maximum likelihood theory. This insight can be used to derive an indicator, that is focused on the rate at which learning, interpreted as increasing posterior precision (defined as the inverse of the variance), occurs.

Proposition 5 (Bayesian learning rate indicator)**.** *Consider Assumptions 1 and 2 for the linearized DSGE model given in equations* (7.1) *and* (7.2)*. Let* $\theta_0 = (\theta_1, \theta_2) \in \Theta$ *be a regular point. Assume there is an open neighborhood of* θ_0 *in which*

$$\overline{\tau}(\theta_0) := \begin{pmatrix} \bar{y}' & vec(h_x)' & vec(h_u)' & vec(g_x)' & vec(g_u)' & \Gamma_{2,u}' \end{pmatrix}'$$

has a constant rank. Assume that the prior distribution of θ *is given by the multivariate normal density and the posterior distribution is approximately normally distributed. Furthermore, suppose* θ_2 *is identified, whereas* θ_1 *is weakly identified such that the rank of* $\frac{\partial \overline{\tau}(\theta_0)}{\partial \theta_1}$ *depends on the sample size* T*. For a fixed* T *it can be full rank, but not for* $T \to \infty$*. Then*

1. *the posterior precision of* θ_1 *divided by the sample size will go to zero.*
2. *the posterior precision of* θ_2 *divided by the sample size will go to a constant.*

Proof. See Proposition 2 of Koop et al. (2013). The idea is that in the case of normal priors and identified parameters the precision increases at the rate of the data size, whereas for not or weakly identified parameters the rate is slower. □

Intuitively, the more data is used, the more precisely one can estimate a parameter. This implies a shrinking variance or increasing

precision. According to Koop et al. (2013, Remark 2) Proposition 5 also holds for non-Gaussian priors, however, an exact expression can only be derived for special cases.[32] Implementing this indicator in practice is straightforward: First, one simulates a large data set (e.g. $T = 10\ 000$) and divides it into growing subsamples. Then an estimator for the precision is calculated. I focus on the Hessian at the posterior mode to compute the precision, since the Hessian is equal to the asymptotic precision. Furthermore, when using gradient-based methods to find the posterior mode, the Hessian is a by-product of the optimization algorithm.[33] This naturally gives a second interpretation of this indicator: it exploits the structure of the information matrix, since the expectation of the Hessian is equal to the information matrix.[34] Therefore, a singular Hessian indicates that the likelihood (or posterior density) is flat in the direction of unidentified parameters. A different computational approach is to run a full-fledged Bayesian MCMC estimation and focus on the variance of the posterior distribution of the parameters. I follow standard practice and use a Random-Walk Metropolis-Hastings sampling algorithm to estimate the model in a Bayesian fashion. The algorithm is outlined in Appendix H. In the empirical applications, I calculate the indicator based on the Hessian method as well as on the MCMC method to compare results. In this respect, I follow closely Caglar et al. (2012) who use the Bayesian learning rate indicator on the model of Smets and Wouters (2007).

[32] For identification issues regarding different priors see also Onatski and Williams (2010).

[33] Andreasen (2010) discusses the use of non-gradient optimization methods (CMA-ES and Simulated Annealing) for DSGE models and finds that these often perform better. However, the algorithms have to be reformulated to get the correct Hessian as the by-product.

[34] See also Iskrev (2010b) for a data-independent interpretation of the information matrix.

7.3 Applications

The different methods are now demonstrated by means of two models that are known to have lack of identification: (i) the Kim (2003) model and (ii) the An and Schorfheide (2007) model.

7.3.1 The Kim (2003) model

Reconsider the model of Section 3.1. In the original paper, Kim (2003) log-linearizes this model and shows analytically that there is observational equivalence between intertemporal and multisectoral investment adjustment costs:

> *[W]hen a model already has a free parameter for intertemporal adjustment costs, adding another parameter for multisectoral adjustment costs does not enrich the model dynamics (Kim, 2003, p. 534).*

Thus, in the linearized model, the set (θ, ϕ) is observationally equivalent, since both parameters enter as a ratio $\frac{\phi+\theta}{1+\theta}$ into the linearized solution. Therefore all ranks should be short by one. Table 7.1 confirms this analytical result throughout all identification tests for the prior mean given in Table 3.1. Analyzing the nullspace of rank deficient criteria yields unanimously the result that the combination (θ, ϕ) is observational equivalent in the linearized model. This result is robust across tolerance levels, as well as across the choice of derivatives. Nevertheless there are discrepancies due to numerical issues. For analytical derivatives this holds for tolerance levels as low as 1e-13 for Iskrev (2010a)'s and Komunjer and Ng (2011)'s approach and even lower for Qu and Tkachenko (2012)'s approach. Regarding numerical derivatives, there is a dependence on the differentiation step: the smaller the step, the more likely one erroneously concludes an identified model. This is not surprising, since the numerical error of the solution algorithm will be relative large compared to a very small step size, and therefore loosing precision during the rank

Table 7.1: Rank analysis: linearized Kim (2003) model

	Iskrev			*Qu/Tkachenko*	
Tol	$\overline{J}$	M_2	$\overline{M_2}$	G_2	$\overline{G_2}$
1e-03	6 (6,6,6)	6 (6,6,6)	6 (6,6,6)	4 (4,4,4)	4 (4,4,4)
1e-05	6 (7,6,6)	6 (6,6,6)	6 (6,6,7)	6 (6,6,6)	6 (6,6,6)
1e-07	6 (7,7,7)	6 (6,6,7)	6 (6,6,7)	6 (6,6,6)	6 (6,6,6)
1e-09	6 (7,7,7)	6 (6,6,7)	6 (6,7,7)	6 (6,6,6)	6 (6,6,7)
1e-11	6 (7,7,7)	6 (7,7,7)	6 (7,7,7)	6 (6,6,7)	6 (6,6,7)
1e-13	7 (7,7,7)	6 (7,7,7)	6 (7,7,7)	6 (6,6,7)	6 (6,6,7)
1e-15	7 (7,7,7)	6 (7,7,7)	7 (7,7,7)	6 (6,6,7)	6 (6,6,7)
rob	7 (7,7,7)	6 (7,7,7)	6 (7,7,7)	6 (6,6,7)	6 (6,6,7)
full	7	7	7	7	7

	Komunjer/Ng				
Tol	Δ_Λ	$\Delta_{\Lambda T}$	$\Delta_{\Lambda U}$	Δ	$\overline{\Delta}$
1e-03	6 (6,6,6)	9 (9, 9, 9)	7 (7,7,7)	9 (9, 9, 9)	11 (11,11,11)
1e-05	6 (7,6,6)	10 (10,10,11)	7 (7,7,8)	10 (10,10,11)	11 (11,11,12)
1e-07	6 (7,7,7)	10 (10,10,11)	7 (7,7,8)	11 (11,11,12)	11 (11,11,12)
1e-09	6 (7,7,7)	10 (10,10,11)	7 (7,8,8)	11 (11,11,12)	11 (11,11,12)
1e-11	6 (7,7,7)	10 (11,11,11)	7 (8,8,8)	11 (12,12,12)	11 (12,12,12)
1e-13	7 (7,7,7)	10 (11,11,11)	8 (8,8,8)	11 (12,12,12)	11 (12,12,12)
1e-15	7 (7,7,7)	11 (11,11,11)	8 (8,8,8)	12 (12,12,12)	12 (12,12,12)
rob	7 (7,7,7)	10 (11,11,11)	8 (8,8,8)	11 (12,12,12)	12 (12,12,12)
full	7	11	8	12	12

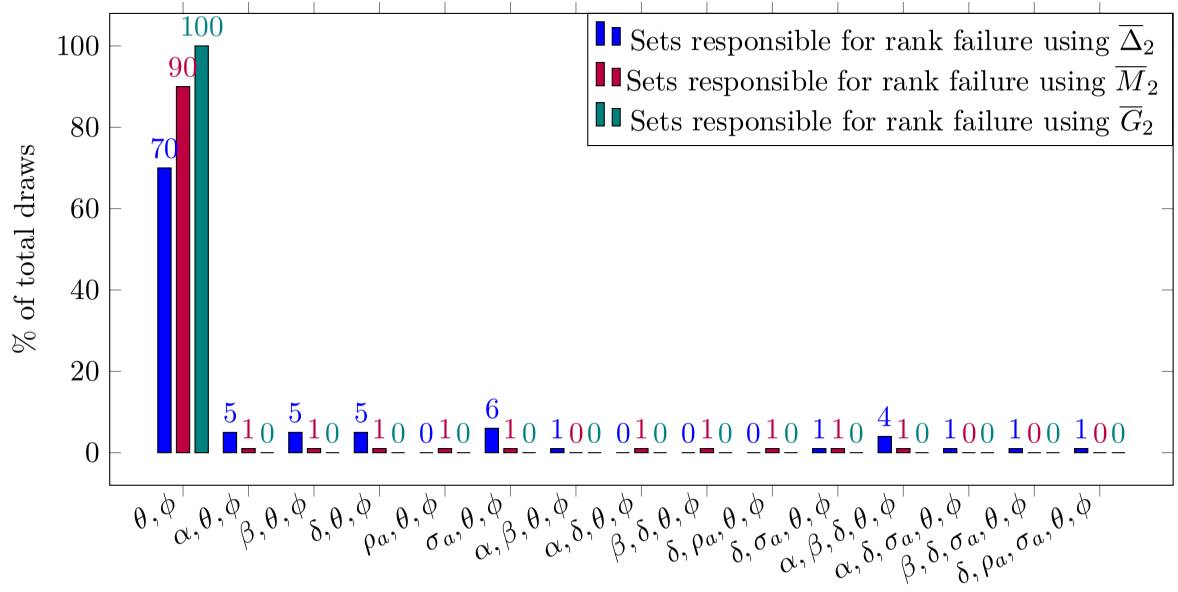
Notes: Ranks with analytical derivatives for different tolerance levels Tol (rob is robust level), lags in autocovariogram $T = 30$, subintervals $N = 10\ 000$. In parenthesis are the corresponding ranks computed with numerical derivatives given differentiation steps 1e-3, 1e-7 and 1e-11, respectively.

computations. Even though Qu and Tkachenko (2012)'s approach is in this regard most reliable, I strongly recommend using analytical derivatives, since this yields numerically better results for all methods. In Figure 7.1, I repeat the exercise for a 100 random draws from the prior domain. All criteria indicate that the set (θ, ϕ) are not identifiable in a first-order approximation.

Note that I focus on the original propositions of Iskrev (2010a), Komunjer and Ng (2011) and Qu and Tkachenko (2012), i.e. using theoretical information from the first two moments of data only. Naturally, using higher-order statistics with Gaussian innovations does not change results, since higher-order cumulants and polyspectra are zero under the assumption of normality. On the other hand, using the t-distribution as the underlying shock process yields the same ranks as in Table 7.1. This is not surprising, since the rank deficiency in this model is due to the first-order approximation. In Chapter 8, I show that an approximation to the second-order yields additional restrictions to identify θ and ϕ separately.

Lastly, Table 7.2 gives insight into the strength of identification according to Proposition 5. I fix one parameter in the problematic set, namely ϕ, in order to estimate posterior precisions for growing datasets in an identified model. To this end, I generate one artificial dataset of 10 000 observations and then estimate the model with Bayesian methods using the first $T = 20$, 50, 100, 1 000 and 10 000 observations. The posterior precision should increase at a rate of T for identified parameters, whereas for weakly identified parameters it increases at a slower rate. In other words, average precision of strongly identified parameters should tend to a constant, whereas for weakly identified parameters it is heading towards zero. The simulation reveals that ϕ is rather weakly identifiable, since the rate at which the precision is updated is slower than the sample size. This is particularly true for the priors specified in Table 3.1. The other parameters seem to be strongly identified.

Figure 7.1: Non-identified sets: linearized Kim (2003)



Notes: Identification results 100 draws from the prior domain using analytical derivatives with robust tolerance level, $T = 30$ and $N = 10\,000$. Sets by brute-force method.

Table 7.2: Average posterior precision: linearized Kim (2003) model

	GAUSSIAN PRIORS						PRIORS TABLE 3.1					
T	α	β	δ	θ	ρ_a	σ_a	α	β	δ	θ	ρ_a	σ_a
	HESSIAN METHOD						HESSIAN METHOD					
20	16.71	28.96	$8{\cdot}10^{13}$	4.992	7.970	6.692	6.080	10.77	$8{\cdot}10^{13}$	0.194	3.731	2.506
50	11.33	26.02	$3{\cdot}10^{13}$	2.004	4.309	3.008	8.295	17.91	$3{\cdot}10^{13}$	0.085	2.904	2.788
100	8.802	29.97	$1{\cdot}10^{13}$	1.005	2.779	1.973	11.71	29.64	$1{\cdot}10^{13}$	0.045	2.316	3.808
1 000	0.295	0.387	146.2	0.107	2.091	0.470	0.103	0.138	54.78	0.009	1.943	0.445
10 000	0.054	0.077	29.13	0.015	1.871	0.302	0.033	0.048	17.96	0.005	1.785	0.311
	MCMC METHOD						MCMC METHOD					
20	21.86	40.12	$3{\cdot}10^{11}$	5.152	8.099	6.860	5.576	22.58	$5{\cdot}10^{11}$	0.190	4.007	0.365
50	11.41	25.50	$1{\cdot}10^{11}$	1.987	4.386	3.071	3.108	12.91	$6{\cdot}10^{10}$	0.090	2.822	0.489
100	8.399	29.14	$8{\cdot}10^{9}$	0.988	2.836	1.824	6.913	22.59	$1{\cdot}10^{10}$	0.046	2.140	1.737
1 000	0.614	0.858	307.7	0.106	2.105	0.363	0.557	0.858	302.9	0.008	1.945	0.293
10 000	0.088	0.128	48.94	0.015	1.999	0.278	0.079	0.116	44.86	0.004	1.955	0.263

Notes: ϕ is fixed at true value. The posterior mode and the Hessian matrix, evaluated at the mode, are computed by a Covariance Matrix Adaptation Evolution Strategy algorithm. Hessian method corresponds to computing the precision using the Hessian-based estimate for the covariance matrix. The MCMC method corresponds to using the variances from the draws of the marginal posterior distributions. A Random-Walk Metropolis-Hastings (RWMH) sampling algorithm is used (as outlined in Appendix H), based on three Markov chains with 20 000 draws, 10 000 draws being discarded as burn-in draws. The mode and Hessian evaluated at the mode are used to determine the initial proposal density for the RWMH algorithm. The scale parameter of the jumping distribution's covariance matrix is set to 1, such that the acceptance ratios lie in between 20%-35%. Gaussian priors correspond to using truncated independent normal distributions with mean set to the true value and standard deviation equal to 0.1. Truncation corresponds to the bounds specified in Table 3.1.

7.3.2 The An and Schorfheide (2007) model

This model is a prototypical DSGE model often cited in the literature concerning a lack of identification. The authors show that (in the version outlined in Chapter 3.2) the set of parameters (ν, ϕ) and the steady state ratio $1/g = c/y$ do not enter the linearized solution. Komunjer and Ng (2011), Ratto and Iskrev (2011) and Qu and Tkachenko (2012) also show that the coefficients entering the Taylor-rule $(\psi_1, \psi_2, \rho_R, \sigma_R)$ are not separately identifiable in the linearized DSGE model. However, they all use a slightly different version of the model (log-linearized model, simplified measurement equations, different parametrization); hence, their results are not directly comparable. Therefore, I will focus on the original specification in its nonlinearized form (equations (3.28) to (3.39)) and check the identification criteria across different tolerance levels, expecting rank deficiencies by three. Furthermore, I will distinguish two models: $\mathcal{TR}_1$ uses the output-gap rule and $\mathcal{TR}_2$ the output-growth rule in the specification of monetary policy in equation (3.29).

Table 7.3 shows the results for model $\mathcal{TR}_1$ given analytical and numerical derivatives with different differentiation steps. First, turning to analytical derivatives, all criteria yield correctly rank deficiency by three. Looking into the nullspace confirms that c/y, the set (ν, ϕ) and the Taylor rule coefficients $(\psi_1, \psi_2, \rho_R, \sigma_R)$ are not identifiable. The thresholds are 1e-15 for Iskrev's and Komunjer and Ng's criteria and 1e-17 for Qu and Tkachenko's test. For very small, i.e. strict tolerance levels Iskrev's and Komunjer and Ng's methods fail to detect the Taylor-rule coefficients. This issue becomes even more severe, when calculating the derivatives numerically. Now there is a trade-off between setting the differentiation step too large (e.g. 1e-3), and thus possibly calculating the derivatives imprecisely, or too small (e.g. 1e-11), such that the numerical error from the solution algorithm possibly outweighs the differentiation error. Both results in false rank calculation, as can be seen

Table 7.3: Rank analysis: linearized An/Schorfheide (2007) model $\mathcal{TR}_1$

	Iskrev			*Qu/Tkachenko*	
Tol	$\overline{J}$	M	$\overline{M}$	G	$\overline{G}$
1e-03	13 (13,13,13)	9 (9,9,9)	12 (12,12,12)	7 (7,7,7)	10 (10,10,10)
1e-05	13 (14,13,13)	10 (10,10,12)	12 (12,12,13)	9 (9,9,9)	12 (12,12,12)
1e-07	13 (14,14,14)	10 (11,10,13)	12 (13,12,14)	10 (10,10,10)	12 (12,12,12)
1e-09	13 (14,14,14)	10 (12,12,13)	12 (14,13,14)	10 (10,10,11)	12 (12,12,12)
1e-11	13 (14,14,14)	10 (12,13,13)	12 (14,14,14)	10 (10,10,12)	12 (12,12,13)
1e-13	13 (14,14,14)	10 (13,13,13)	12 (14,14,14)	10 (10,10,13)	12 (12,12,14)
1e-15	13 (14,14,14)	11 (13,13,13)	13 (14,14,14)	10 (12,10,13)	12 (14,12,14)
robust	13 (14,14,14)	10 (12,13,13)	12 (14,14,14)	10 (11,10,13)	12 (12,12,14)
full	15	15	15	15	15
			Komunjer/Ng		
Tol	Δ_Λ	$\Delta_{\Lambda T}$	$\Delta_{\Lambda U}$	Δ	$\overline{\Delta}$
1e-03	10 (10,10,10)	19 (21,22,21)	19 (22,22,22)	27 (27,27,27)	30 (30,30,30)
1e-05	11 (11,11,11)	20 (22,22,22)	20 (22,22,22)	28 (28,28,29)	30 (30,30,31)
1e-07	11 (13,11,11)	20 (22,22,22)	20 (22,22,22)	28 (29,28,31)	30 (31,30,32)
1e-09	11 (13,13,13)	20 (22,22,22)	20 (22,22,22)	28 (29,29,31)	30 (32,31,32)
1e-11	11 (13,13,13)	20 (22,22,22)	20 (22,22,22)	28 (30,31,31)	30 (32,32,32)
1e-13	11 (13,13,13)	20 (22,22,22)	20 (22,22,22)	28 (31,31,31)	30 (32,32,32)
1e-15	11 (13,13,13)	20 (22,22,22)	20 (22,22,22)	29 (32,31,31)	31 (32,32,32)
robust	11 (13,13,13)	20 (22,22,22)	20 (22,22,22)	28 (31,31,31)	30 (32,32,32)
full	15	24	24	33	33

Notes: Ranks with analytical derivatives for different tolerance levels Tol, lags in autocovariogram $T = 30$, subintervals $N = 10\ 000$. In parenthesis are the corresponding ranks computed with numerical derivatives given differentiation steps 1e-3, 1e-7 and 1e-11, respectively.

in Table 7.3, where all tests fail to detect lack of identification of the Taylor-rule coefficients even for mild tolerance levels. Using a feasible trade-off for the numerical differentiation step (e.g. 1e-7), the threshold of correctly determining rank-deficiency by three are now 1e-7 for Iskrev's and Komunjer and Ng's criteria. Again Qu and Tkachenko's test is most reliable when using numerical derivatives. So compared to the analytical case, one has to loosen the tolerance level for the rank calculations or otherwise one gets wrong results. Therefore, I am strongly in favor of analytical procedures and advise using them whenever feasible; the results are unanimously more reliable even for stricter tolerance levels. Furthermore, $\overline{J}$ and Δ_Λ can be used for diagnostic issues of the model. $\overline{J}$ is rank-deficient by two, i.e. clearly indicating that c/y and (ν, ϕ) do not enter neither the mean nor the first-order solution matrices. Δ_Λ is short by 4, i.e. $\gamma^{(Q)}$, $\pi^{(A)}$, c/y and (ϕ, ν) do not enter the solution matrices. Thus, I conclude that $\pi^{(A)}$ and $\gamma^{(Q)}$ are only identified via the mean using the first-order approximation, which is in accordance to the measurement equations (3.34), (3.35) and (3.36).

Table 7.4 shows the corresponding ranks for model $\mathcal{TR}_2$. Changing the specification of the Taylor rule yields rank shortage by two across methods. Looking into the nullspace confirms that now only c/y and the set (ν, ϕ) are not identifiable. The same result holds when I repeat the analysis for 100 random draws from the prior domain. As can be seen in Figure 7.2 for model $\mathcal{TR}_1$, the Taylor rule coefficients enter the problematic sets, whereas Figure 7.3 indicates that for model $\mathcal{TR}_2$ the Taylor rule parameters are identified. Non-identification of the Taylor rule crucially depends on its specification, a feature that will likely prevail in other models.

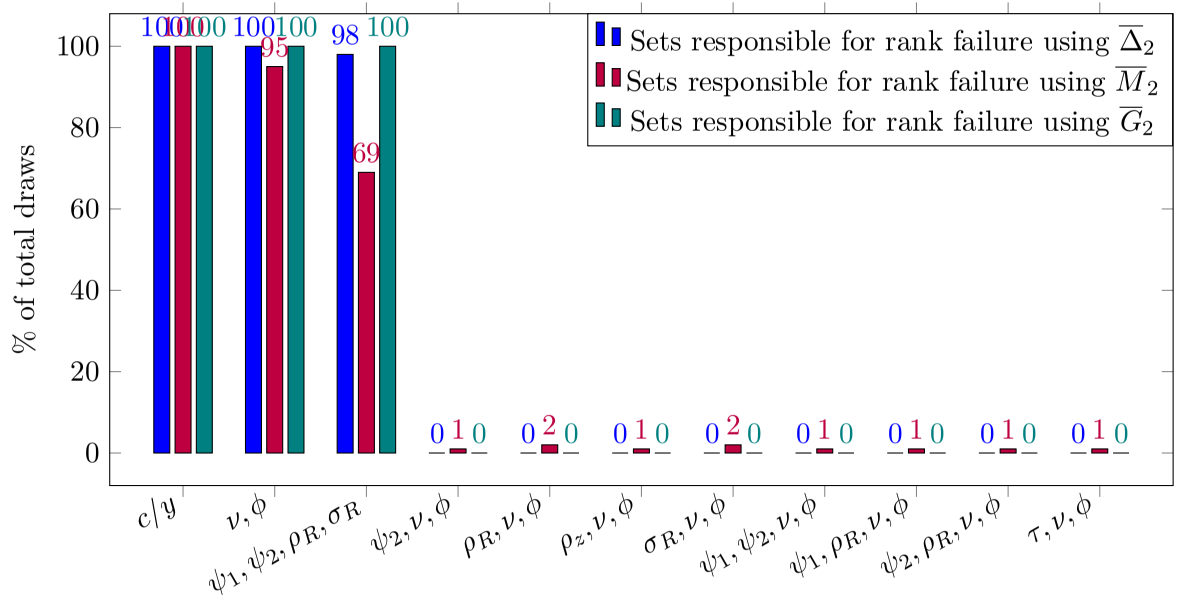
Note that I focus on the original propositions of Iskrev (2010a), Komunjer and Ng (2011) and Qu and Tkachenko (2012), i.e. using theoretical information from the first two moments of data only. Naturally, using higher-order statistics with Gaussian innovations

Table 7.4: Rank analysis: linearized An/Schorfheide (2007) model $\mathcal{TR}_2$

	Iskrev			*Qu/Tkachenko*	
Tol	$\overline{J}$	M	$\overline{M}$	G	$\overline{G}$
1e-03	13 (13,13,13)	10 (10,10,10)	13 (13,13,13)	7 (7,7,7)	10 (10,10,9)
1e-05	13 (14,13,13)	11 (11,11,12)	13 (13,13,14)	10 (10,10,10)	13 (13,13,13)
1e-07	13 (14,14,14)	11 (11,11,13)	13 (13,13,14)	11 (11,11,11)	13 (13,13,14)
1e-09	13 (14,14,14)	11 (12,12,13)	13 (14,14,14)	11 (11,11,12)	13 (13,13,14)
1e-11	13 (14,14,14)	11 (12,13,13)	13 (14,14,14)	11 (11,11,12)	13 (13,13,14)
1e-13	13 (14,14,14)	11 (13,13,13)	13 (14,14,14)	11 (11,11,13)	13 (13,13,14)
1e-15	13 (14,14,14)	12 (13,13,13)	14 (14,14,14)	11 (12,11,13)	13 (14,13,14)
robust	13 (14,14,14)	11 (13,13,13)	13 (14,14,14)	11 (11,11,13)	13 (13,13,14)
full	15	15	15	15	15
			Komunjer/Ng		
Tol	Δ_Λ	$\Delta_{\Lambda T}$	$\Delta_{\Lambda U}$	Δ	$\overline{\Delta}$
1e-03	10 (10,10,10)	19 (19,19,19)	19 (20,20,20)	28 (28,28,28)	31 (31,31,31)
1e-05	11 (11,11,11)	20 (20,20,22)	20 (21,21,22)	29 (29,29,30)	31 (31,31,32)
1e-07	11 (13,11,11)	20 (20,20,22)	20 (21,21,22)	29 (29,29,31)	31 (31,31,32)
1e-09	11 (13,13,13)	20 (21,22,22)	20 (21,22,22)	29 (29,31,31)	31 (32,32,32)
1e-11	11 (13,13,13)	20 (21,22,22)	20 (22,22,22)	29 (30,31,31)	31 (32,32,32)
1e-13	11 (13,13,13)	20 (22,22,22)	20 (22,22,22)	29 (31,31,31)	31 (32,32,32)
1e-15	11 (13,13,13)	20 (22,22,22)	20 (22,22,22)	30 (32,31,31)	32 (32,32,32)
robust	11 (13,13,13)	20 (22,22,22)	20 (22,22,22)	29 (31,31,31)	31 (32,32,32)
full	15	24	24	33	33

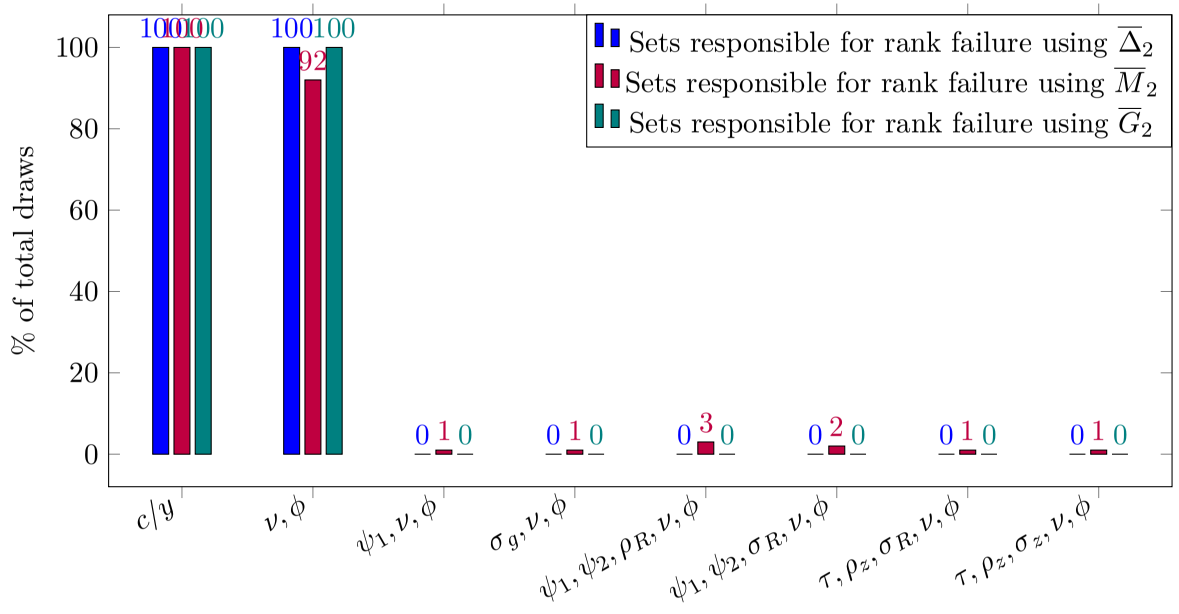
Notes: Ranks with analytical derivatives for different tolerance levels Tol, lags in autocovariogram $T = 30$, subintervals $N = 10\ 000$. In parenthesis are the corresponding ranks computed with numerical derivatives given differentiation steps 1e-3, 1e-7 and 1e-11, respectively.

Figure 7.2: Non-identified sets: linearized An/Schorfheide (2007) model $\mathcal{TR}_1$



Notes: Identification results for 100 draws from the prior domain using analytical derivatives with robust tolerance level, $T = 30$ and $N = 10\ 000$. Sets by brute-force method. $\mathcal{TR}_1$ corresponds to the output-gap specification of the Taylor-rule.

Figure 7.3: Non-identified sets: linearized An/Schorfheide (2007) model $\mathcal{TR}_2$



Notes: Identification results for 100 draws from the prior domain using analytical derivatives with robust tolerance level, $T = 30$ and $N = 10\,000$. Sets by brute-force method. $\mathcal{TR}_2$ corresponds to the output-growth specification of the Taylor-rule.

does not change results, since higher-order cumulants and polyspectra are zero under the assumption of normality. On the other hand, using the t-distribution as the underlying shock process yields the same ranks as in Table 7.1. An and Schorfheide, however, already argue that "*the nonlinear approach is able to extract more information on the structural parameters from the data*" (An and Schorfheide, 2007, p. 164). In Chapter 8, I confirm these alluring results formally. In particular, an approximation to the second-order yields additional restrictions to identify all parameters of the model no matter which specification of monetary policy is used.

Lastly, Tables 7.5 and 7.6 give insight into the strength of identification according to Proposition 5. First, I fix one parameter in each of the three problematic sets of model $\mathcal{TR}_1$, namely c/y, ϕ and ρ_R, in order to estimate posterior precisions for larger and larger datasets in an identified model. To this end, I generate one artificial dataset of 10 000 observations and then estimate the model with Bayesian methods using the first $T = 20$, 50, 100, 1 000 and 10 000 of the artificially generated observations. The posterior precision should increase at a rate of T for identified parameters, whereas for weakly identified parameters it increases at a slower rate. In other words, average precision of strongly identified parameters should tend to a constant, whereas for weakly identified parameters it is heading towards zero. Both the estimation based on the Hessian as well as on the MCMC algorithm reveal that τ is weakly identified. Moreover, the MCMC method also indicates weak identification of ψ_1 and $r^{(A)}$. This is true for either using independent normal distributions or the ones specified in Table 3.2 as priors.

Table 7.5: Average posterior precision via Hessian: linearized An/Schorfheide (2007) model

	PARAMETERS											
T	τ	ψ_1	ψ_2	ρ_g	ρ_z	$r^{(A)}$	$p^{(A)}$	γ_Q	$100\sigma_R$	$100\sigma_g$	$100\sigma_z$	ν
	HESSIAN METHOD - GAUSSIAN PRIORS											
20	5.069	5.085	8.579	3.803	98.85	5.077	5.475	5.905	36.40	9.614	17.40	368.0
50	2.035	2.041	4.120	4.878	46.96	2.108	2.862	3.099	36.25	7.700	10.68	307.9
100	1.030	1.028	3.049	3.334	56.86	1.145	3.563	2.154	43.00	7.132	11.71	264.8
1 000	0.155	0.198	5.488	7.962	73.15	0.170	0.706	0.805	47.64	5.974	14.07	100.0
10 000	0.034	0.152	5.042	5.713	28.69	0.192	0.872	0.635	46.59	4.865	9.360	29.69
	HESSIAN METHOD - PRIORS TABLE 3.2											
20	0.299	1.122	11.15	3.353	78.77	0.142	0.189	1.679	30.36	10.92	18.31	186.3
50	0.189	0.391	2.324	6.183	33.88	0.067	0.417	0.757	31.62	7.148	10.61	121.2
100	0.142	0.192	1.628	3.035	56.61	0.060	0.368	0.471	38.01	7.128	13.22	93.32
1 000	0.058	0.136	4.037	8.117	43.12	0.023	0.094	0.153	45.58	5.996	11.43	44.64
10 000	0.045	0.150	5.284	7.638	45.43	0.117	0.557	0.771	47.51	5.493	12.21	39.44

Notes: ϕ, ρ_R and c/y are fixed at true values. The posterior mode and the Hessian matrix, evaluated at the mode, are computed by a Nelder-Mead simplex optimization routine. Hessian method corresponds to computing the precision using the Hessian-based estimate for the covariance matrix. Gaussian priors correspond to using truncated independent normal distributions with mean set to the true value and standard deviation equal to 0.1. Truncation corresponds to the bounds specified in Table 3.2.

Table 7.6: Average posterior precision via MCMC: linearized An/Schorfheide (2007) model

	PARAMETERS											
T	τ	ψ_1	ψ_2	ρ_g	ρ_z	$r^{(A)}$	$p^{(A)}$	γ_Q	$100\sigma_R$	$100\sigma_g$	$100\sigma_z$	ν
	MCMC METHOD - GAUSSIAN PRIORS											
20	5.281	5.026	10.84	10.69	113.6	5.378	5.310	6.130	30.75	8.704	19.30	364.5
50	2.027	2.075	5.153	10.68	51.34	2.182	3.210	3.059	33.28	6.870	11.07	329.1
100	1.007	1.060	3.099	3.709	50.30	1.111	3.045	1.994	39.19	6.362	10.34	283.1
1 000	0.141	0.101	1.399	7.542	65.27	0.147	0.631	0.737	43.20	5.552	12.93	84.72
10 000	0.058	0.012	0.294	8.165	74.32	0.032	0.131	0.212	46.40	5.346	15.91	43.80
	MCMC METHOD - PRIORS TABLE 3.2											
20	0.250	1.221	4.425	5.477	47.99	0.166	0.250	1.845	20.41	5.340	10.70	163.6
50	0.136	0.353	1.859	8.585	31.02	0.069	0.437	0.783	22.08	5.600	8.856	103.7
100	0.116	0.203	1.586	3.295	47.24	0.057	0.348	0.463	33.38	6.128	11.47	84.10
1 000	0.058	0.025	0.602	8.343	60.43	0.022	0.097	0.168	42.43	5.653	11.20	40.01
1 0000	0.048	0.004	0.106	7.785	72.07	0.016	0.066	0.113	44.16	5.322	15.49	38.36

Notes: ϕ, ρ_R and c/y are fixed at true values. The posterior mode and the Hessian matrix, evaluated at the mode, are computed by a Nelder-Mead simplex optimization routine. MCMC method corresponds to using the variances from the draws of the marginal posterior distributions. Specifically, a Random-Walk Metropolis-Hastings (RWMH) sampling algorithm is used (as outlined in Appendix H), based on three Markov chains with 20000 draws, 10000 draws being discarded as burn-in draws. The mode and Hessian evaluated at the mode are used to determine the initial proposal density for the RWMH algorithm. The scale parameter of the jumping distribution's covariance matrix is set to 0.6, such that the acceptance ratios lie in between 20%-35%. Gaussian priors correspond to using truncated independent normal distributions with mean set to the true value and standard deviation equal to 0.1. Truncation corresponds to the bounds specified in Table 3.2.

7.4 Comparison

All criteria based on ranks yield similar results and one is able to correctly uncover the sources of non-identification. Nevertheless, the choice of tolerance level for rank computations and the use of analytical vs. numerical derivatives is a delicate one, changing results slightly. Similarly, the Bayesian learning rate indicator depends crucially on the ability of the optimizer to find the mode or the convergence of the MCMC estimation. Having non-identifiable parameters in the algorithms amplifies this problem. Therefore, I recommend to first check the criteria based on ranks to find unidentifiable parameters and exclude these from the Bayesian analysis. In other words, the Bayesian learning rate indicator is best used to check for weak identification rather than for non-identification. In summary, a comparative approach is worthwhile to get robust insight into the model dynamics and dependencies.

Chapter 8

Identification of DSGE models: the effect of higher-order approximation and pruning[35]

Whereas there is a growing literature on the estimation of nonlinear DSGE models (Andreasen, 2011, 2013; Fernández-Villaverde and Rubio-Ramírez, 2007; Herbst and Schorfheide, 2014; Ivashchenko, 2014; Kollmann, 2015), all identification methods of Chapter 7 focus on the linear approximation of the DSGE model to the first order. In this chapter, I show how to check rank criteria for a local identification of nonlinear DSGE models, given higher-order approximations and pruning. Reconsider the model framework of Chapter

[35] The ideas and results of this chapter have been published in Mutschler (2015b).

2. Proposition 1 shows that the pruned state-space (PSS) is a linear, stationary and ergodic state-space system, but with non-Gaussian innovations (even if the underlying shock process u_t is Gaussian). From an identification point of view, this may yield additional restrictions on unconditional moments and polyspectra of the observables that can be used to identify (sets of) parameters which are not identifiable in the linearized DSGE model.

Exploiting these insights, the contribution of this chapter is twofold. First, I extend ideas from Iskrev (2010a) and Qu and Tkachenko (2012) to establish rank criteria for higher-order approximations, both in the time and in the frequency domains. Chapter 5 provides the analytical derivatives of all solution matrices, cumulants and polyspectra of the PSS with respect to the deep parameters of the model. In this manner, I am able to check identification, given theoretical higher-order statistics of observables. I limit myself to fourth-order statistics, since third-order cumulants and the bispectrum capture nonlinearities (or non-Gaussianity) for a skewed process, whereas the fourth-order cumulants and the trispectrum can be used in the case of a non-Gaussian symmetric probability distribution. Throughout the exposition, the focus is on a second-order approximation, since extending ideas and propositions is – apart from notation and computational implementation – conceptually straightforward for higher-order approximations. Second, to demonstrate the propositions, all methods are applied to the example models. In particular, I show that the parameters governing the investment adjustment costs in Kim (2003), as well as all parameters in An and Schorfheide (2007), can be identified from the mean and second moments or power spectrum, as well as from higher-order statistics of the PSS given a second-order approximation.

This chapter is also related to Morris (2014), who likewise derives rank criteria for the pruned state-space (PSS) system, yet in the manner of Komunjer and Ng (2011). The key differences be-

tween my work and Morris (2014) can be summarized with respect to three aspects. First, the perspective and system representation is different, since Komunjer and Ng (2011)'s approach assumes a minimal system. For the PSS, this requires the innovations representation. My approach neither relies on a specific model framework nor on a minimal system; thus, I do not need to reparametrize the solution system. Second, the criteria derived in this chapter also include unconditional third and fourth moments in the time domain and the bi- and trispectrum in the frequency domain, whereas Morris (2014) uses only the first two moments. Third, the computations are different. Since the innovation representation depends on the existence and computation of a discrete algebraic Ricatti equation, it is not possible to derive analytical derivatives. Nevertheless, both approaches come to similar conclusions, which should help build confidence across potential users of the pruned state-space for estimating nonlinear DSGE models.

8.1 Rank conditions

Since the pruned state-space (PSS) is a linear system with well-defined statistical properties, the same criteria as for linearized DSGE models can be checked, in particular for its mean, second-order moments and spectral density. In fact, I extend Iskrev (2010a)'s criteria for third- and fourth-order cumulants and Qu and Tkachenko (2012)'s criteria for the bi- and trispectrum of observables.

Proposition 6 (Iskrev PSS)**.** *Consider Assumptions 1 and 2 for the pruned state-space of a nonlinear DSGE model given in equations* (2.12) *and* (2.13)*. Let $q \leq T$ and assume that*

$$\overline{m}(\theta, q) := \begin{pmatrix} \mu_y' & m_2(\theta, q)' & m_3(\theta, q)' & m_4(\theta, q)' \end{pmatrix}'$$

is a continuously differentiable function of $\theta \in \Theta$. Let $\theta_0 \in \Theta$ be a regular point, θ is then locally identifiable at a point θ_0 from

the first four cumulants (or moments) of y_t, if and only if

$$\overline{M}(q) := \frac{\partial \overline{m}(\theta_0, q)}{\partial \theta'}$$

has a full column rank equal to the number of parameters.

Proof. Follows Iskrev (2010a, Theorem 2) and Rothenberg (1971, Theorem 6). □

In other words, I extend Iskrev (2010a)'s approach and focus on the first four moments of the pruned state-space system. The test checks whether these moments are uniquely determined by the deep parameters, given a finite number of lags. It therefore corresponds to Definition 1, i.e. identifiability from the first four cumulants or moments given a finite number of moment conditions. This immediately gives rise to a necessary condition: the number of identifiable parameters does not exceed the dimension of $\overline{m}(\theta, T)$. The criteria can also be used for conditional identification, that is, identification for only a subset of parameters. It is also possible to check whether the parameters are identifiable through (i) the mean and second or (ii) the mean and third or (iii) the mean and fourth moments of observables, separately. I denote the corresponding matrices as $\overline{M}_2(q)$, $\overline{M}_3(q)$ and $\overline{M}_4(q)$. Note that, given a first-order approximation and the Gaussian distribution for u_t, the proposition reduces to the original Theorem 2 of Iskrev (2010a), since all higher-order cumulants are zero in this case. Iskrev (2010a, Corollary 1) also proposes a necessary condition, that is, checking injectivity of the mapping from the deep parameters to the solution matrices. To do so, stack all elements of the steady state, the solution matrices as well as all parameters of the stochastic innovations into a vector $\overline{\tau}(\theta) := \left(\bar{y}'\ c'\ d'\ vec(A)'\ vec(B)'\ vec(C)'\ vec(D)'\ \Gamma'_{2,\xi}\ \Gamma'_{3,\xi}\ \Gamma'_{4,\xi}\right)'$. Consider the factorization $\overline{M}(q) = \frac{\partial \overline{m}(\theta,q)}{\partial \overline{\tau}(\theta)'} \frac{\partial \overline{\tau}(\theta)}{\partial \theta'}$. An immediate corollary im-

plies that a point θ_0 is locally identifiable, only if the rank of

$$\overline{J} := \frac{\partial \overline{\tau}(\theta_0)}{\partial \theta'}$$

at θ_0 is equal to n_θ. This condition is, however, only necessary, because $\overline{\tau}$ may be unidentifiable.

Proposition 7 (Qu & Tkachenko PSS)**.** *Consider Assumptions 1 and 2 for the pruned state-space of a nonlinear DSGE model given in equations* (2.12) *and* (2.13)*. Assume that the power spectrum* (4.9)*, bispectrum* (4.10) *and trispectrum* (4.11) *are continuous in* $\omega \in [-\pi; \pi]$ *and continuous and differentiable in* $\theta \in \Theta$*. Let*

$$\begin{aligned}\overline{G}(\theta) = {} & \mathtt{d}(\mu_y(\theta))' \, \mathtt{d}(\mu_y(\theta)) + \int_{-\pi}^{\pi} \mathtt{d}(\mathcal{S}_{2,y}(\omega_1;\theta))^* \, \mathtt{d}(\mathcal{S}_{2,y}(\omega_1;\theta)) \, d\omega_1 \\ & + \int_{-\pi}^{\pi}\int_{-\pi}^{\pi} \mathtt{d}(\mathcal{S}_{3,y}(\omega_1,\omega_2;\theta))^* \, \mathtt{d}(\mathcal{S}_{3,y}(\omega_1,\omega_2;\theta)) \, d\omega_1 d\omega_2 \\ & + \int_{-\pi}^{\pi}\int_{-\pi}^{\pi}\int_{-\pi}^{\pi} \mathtt{d}(\mathcal{S}_{4,y}(\omega_1,\omega_2,\omega_3;\theta))^* \, \mathtt{d}(\mathcal{S}_{4,y}(\omega_1,\omega_2,\omega_3;\theta)) \, d\omega_1 d\omega_2 d\omega_3\end{aligned}$$

and $\theta_0 \in \Theta$ *be a regular point. Furthermore, assume there is an open neighborhood of* θ_0 *in which* $\overline{G}(\theta_0)$ *has a constant rank. Then,* θ *is locally identifiable at a point* θ_0 *from the mean, power spectrum, bispectrum and trispectrum of* y_t*, if and only if* $\overline{G}(\theta_0)$ *is nonsingular, i.e. its rank is equal to the number of parameters.*

Proof. Follows Qu and Tkachenko (2012, Theorem 2) and Rothenberg (1971, Theorem 1). Note that I use the complex conjugate $*$, since the polyspectra are in general complex matrices. $\mathtt{d}\mathcal{S}_{k,y}^* \mathtt{d}\mathcal{S}_{k,y}$ is a Gram matrix, therefore it is Hermitian and positive semidefinite. Furthermore, there is an isomorphism between complex and real matrices such that the $(n_y^k \times 1)$ vector $\mathcal{S}_{k,y}$ can be transformed into a $(2n_y^k \times 2)$ real matrix $\mathcal{S}_{k,y}^R$ (see Brillinger (2001, p. 71) and Pintelon and Schoukens (2001,

p. 553)). The following equivalence holds: $\mathcal{S}^*_{k,y}\mathcal{S}_{k,y} \Leftrightarrow \mathcal{S}^{R'}_{k,y}\mathcal{S}^{R}_{k,y}$. Furthermore $2 \cdot rank(\mathcal{S}^*_{k,y}\mathcal{S}_{k,y}) = rank(\mathcal{S}^{R'}_{k,y}\mathcal{S}^{R}_{k,y})$. The same is true if one considers the differential of $\mathcal{S}_{k,y}$ with respect to θ_j. The proof requires $rank(\mathtt{d}\mathcal{S}^{R'}_{k,y}\mathtt{d}\mathcal{S}^{R}_{k,y})$ to be nonsingular, i.e. full rank, for θ_0 to be locally identified. This is equivalent to $rank(\mathtt{d}\mathcal{S}^*_{k,y}\mathtt{d}\mathcal{S}_{k,y}) = n_\theta$. □

This proposition corresponds to Definition 2, i.e. identifiability from the complete set of dynamic properties. Qu and Tkachenko (2012) provide several extensions, which also apply in this setting. In particular, identification can be checked from a subset of frequencies only, conditional on other parameters being fixed or including general constraints on the parameters. Moreover, one can check whether the parameters are identifiable through the mean and individual spectra. I denote the corresponding matrices as $\overline{G}_2(\theta_0)$, $\overline{G}_3(\theta_0)$ and $\overline{G}_4(\theta_0)$. Given a first-order approximation and the Gaussian distribution for u_t, this proposition reduces to the original Theorem 2 of Qu and Tkachenko (2012), since the bi- and trispectrum are zero in this case. Note that Qu and Tkachenko (2012) use numerical derivatives, whereas I am able to use analytical derivatives due to Chapter 5. For both cases, however, it is still necessary to divide the interval $[-\pi;\pi]$ into sufficient subintervals to numerically approximate the integrals. Lastly, the dimension of $\overline{G}(\theta_0)$ is always $n_\theta \times n_\theta$.

8.2 Implementation

Both propositions exploit the dynamic structure of the pruned solution of a nonlinear DSGE model, in order to define mappings and establish conditions for local injectivity of the mappings. Necessary as well as sufficient conditions for local identification, based on ranks of Jacobians, are derived. For calculating the ranks, I use the singular value decomposition and count the nonzero entries on the main di-

agonal. This requires a specification of the tolerance level, for which I use both a range from 1e-3 to 1e-15 as well as a robust level that depends on the size of the matrix $(max(size(X)) \times eps(norm(X)))$, which is also MATLAB's default value.

Strictly speaking, the criteria are a *yes* or *no* condition. Loosely speaking, however, if a parameter is identified for very large tolerance levels, then it is most likely strongly identified. If it is identified only for very low levels, this may indicate weak identification.[36] In the case of rank deficiency, one is able to pinpoint sets of problematic parameters by analyzing the nullspace. This will be a vector of zeros, if a parameter does not affect the objective at hand. Furthermore the columns that are linearly dependent indicate that these sets of parameters are indistinguishable. While this approach, followed by Iskrev (2010a), is computationally very fast, I find that in some cases, there were redundancies in the subsets, since larger subsets may include smaller ones and are not pinpointed separately. Thus, similar to Ratto and Iskrev (2011) and Qu and Tkachenko (2012, Corollary 4), a more robust method is to consider the power set and check the criteria for all possible subsets of parameters in a recursive fashion. In my experience, this *brute-force* approach yields more reliable results and is computationally just slightly slower, because, if one finds a subset of parameters that are not identified, one can exclude that subset from higher-order subsets.[37]

There are also some further numerical issues. In particular, choosing the lag order T, as well as the number of subintervals N for the frequencies, may change the results. In practice, however, this is not a question of extremely sensitive results, but rather

[36] Note that this is not based on the literature on strength of identification, but provides only a rough indication for subsets of strongly identified parameters. Nevertheless, these can be used as an initial guess for the methods used, for instance, in Koop et al. (2013) and Qu (2014) to detect weak identification.

[37] I implemented both procedures for both criteria in the code.

one of speed: the higher T or N, the more time the calculations need.[38] With this line of thought, one can make use of the inherent symmetry in the computation of cumulants and spectra. That is, I only need to focus on the plane $t_1 \leq t_2 \leq t_3$ for Proposition 6 and $s_1 \leq s_2 \leq s_3$ ($s_j = 1, \ldots, N+1; j = 1, 2, 3$) for Proposition 7, since these determine all other cumulants and spectra through permutations. The computations of the bispectrum can be accelerated further by noting that the sum $\omega_{s_1} + \omega_{s_2}$ contains many duplicate elements, since $\omega_{s_j} \in [-\pi; \pi]$. Thus, I do not need to do the computations for all $N(N+1)/2$ runs, but rather for a much smaller set. Similarly, I do not need to evaluate all $N(N+1)(N+2)/6$ possible values of $\omega_{s_1} + \omega_{s_2} + \omega_{s_3}$ for the trispectrum, but only the unique values. See Chandran and Elgar (1994) for a thorough discussion of principal domains of polyspectra.

Lastly, all procedures check only local identification. Thus, it is necessary to ensure that the analysis is valid for a sufficient range of parameters. Therefore, in the applications, I check all criteria, given first a specific point (e.g. calibrated parameters or prior mean) and second, given many draws from a prespecified prior domain of θ that yield a determinate solution. In this way, one has a quasi-global flavor of the rank criteria for the pruned state-space. Also, most consistent estimators require only local identification for their asymptotic properties to hold. Nevertheless, even if all prior draws are identifiable, the model is still by no means globally identified.

[38] In most practical cases, T between 10 and 30 will be sufficient, since the higher the lag, the less informative the identification restrictions. Furthermore, I experienced with different values for N and find that the results for the applications hardly change. The reason is that, if $\theta_0 \in \Theta$ is identified using only a subset of frequencies (small N), it is also identified if considering the full spectrum ($N \to \infty$) (the converse is not true). Therefore, I recommend starting with $N = 10\,000$ for the power spectrum, $N = 1\,000$ for the bispectrum, $N = 100$ for the trispectrum and increase N if the results are unsatisfactory.

Table 8.1: Rank analysis: nonlinear Kim (2003) model

	Iskrev					*Qu and Tkachenko*			
Tol	$\overline{J}$	$\overline{M}_2$	$\overline{M}_3$	$\overline{M}_4$	$\overline{M}$	$\overline{G}_2$	$\overline{G}_3$	$\overline{G}_4$	$\overline{G}$
1e-03	7 (7)	6 (6)	7 (7)	7 (7)	7 (7)	4 (4)	3 (3)	4 (4)	4 (4)
1e-05	7 (7)	7 (7)	7 (7)	7 (7)	7 (7)	6 (6)	5 (5)	6 (6)	6 (6)
1e-07	7 (7)	7 (7)	7 (7)	7 (7)	7 (7)	7 (7)	7 (7)	7 (7)	7 (7)
1e-09	7 (7)	7 (7)	7 (7)	7 (7)	7 (7)	7 (7)	7 (7)	7 (7)	7 (7)
1e-11	7 (7)	7 (7)	7 (7)	7 (7)	7 (7)	7 (7)	7 (7)	7 (7)	7 (7)
1e-13	7 (7)	7 (7)	7 (7)	7 (7)	7 (7)	7 (7)	7 (7)	7 (7)	7 (7)
1e-15	7 (7)	7 (7)	7 (7)	7 (7)	7 (7)	7 (7)	7 (7)	7 (7)	7 (7)
robust	7 (7)	7 (7)	7 (7)	7 (7)	7 (7)	7 (7)	7 (7)	7 (7)	7 (7)
full	7	7	7	7	7	7	7	7	7

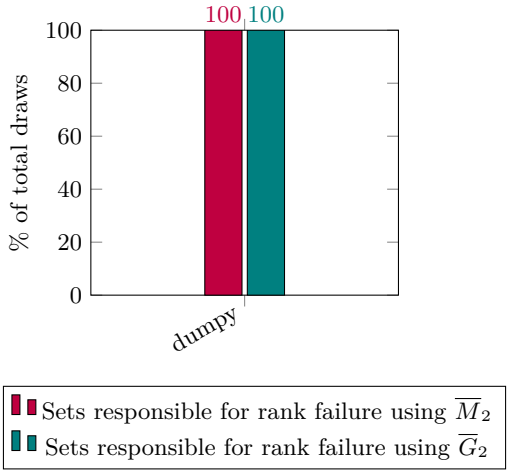
Notes: Ranks of identification tests given local point, different tolerance levels Tol and analytical derivatives. Numerical derivatives with differentiation step 1e-7 in parenthesis. Lags for Iskrev's method set to $T = 30$, and subintervals for Qu and Tkachenko's method set to $N = 10000$ for $\overline{G}_2$, $N = 1000$ for $\overline{G}_3$, $N = 100$ for $\overline{G}_4$ and $N = 100$ for $\overline{G}$.

8.3 Applications

8.3.1 The Kim (2003) model

Reconsider the result of Chapter 7 for the Kim (2003) model. That is, given a first-order approximation, the set (θ, ϕ) is observationally equivalent, since both parameters enter as a ratio $\frac{\phi+\theta}{1+\theta}$ into the solution. However, considering an approximation to the second-order yields additional restrictions on the first four moments and corresponding polyspectra, as can be seen in Table 8.1. All criteria unanimously yield the result that θ and ϕ are distinguishable using a second-order approximation. This result is robust across tolerance levels, as well as across the choice of derivatives. Note that $\overline{M}_k$ checks identification using the mean and kth order cumulants only, whereas $\overline{G}_k$ uses the mean and polyspectrum of order k only. Hence, the model is identifiable using either all information from

Figure 8.1: Non-identified sets: nonlinear Kim (2003)



Notes: Identification results for 100 draws from the prior domain using analytical derivatives with robust tolerance level, $T = 30$ and $N = 10\,000$. Sets by brute-force method.

moments and cumulants ($\overline{M}$) and corresponding polyspectra ($\overline{G}$) or using only individual statistics. Thus, including higher-order statistics may benefit identification and estimation, but it is not necessary for this model. I demonstrate this by repeating the analysis for 100 random draws from the prior domain and using only the first two moments and power spectrum in the objective functions ($\overline{M}_2$ and $\overline{G}_2$). For illustrative purposes, similar to Ratto and Iskrev (2011), a parameter *dumpy* is added into the analysis, which does not enter the model. As is evident in Figure 7.1 of Chapter 7, all criteria indicate that the set (θ, ϕ) is not identifiable in a first-order approximation. Given a second-order approximation and using the mean, autocovariogram and power spectrum of the PSS, the situation is different, see Figure 8.1. Now, in all cases, it is only dumpy that

is not identifiable. I thus conclude that an approximation to the second-order yields additional restrictions to identify θ and ϕ separately, using the nonlinear DSGE model. This result is – as far as I know – new to the literature. Ivashchenko and Mutschler (2015) estimate this model on the basis of second-order approximation and confirm that both parameters are separately estimable.

8.3.2 The An and Schorfheide (2007) model

In Chapter 7 it is shown that in the linearized An and Schorfheide (2007) model, the steady state ratio of government spending $1/g = c/y$ does not enter the solution. Moreover, the elasticity of demand ν and price stickiness parameter ϕ are not separately identifiable. The same is true for the coefficients of the Taylor rule when using the output-gap specification. However, using a second-order approximation and the particle filter, An and Schorfheide (2007) show that a nonlinear approach enriches dynamics:

> *[T]he log-likelihood is slightly sloped in $1/g = c/y$ dimension. Moreover, (...) the quadratic likelihood (...) suggests that ν and ϕ are potentially separately identifiable (An and Schorfheide, 2007, p. 164).*
>
> *[T]he nonlinear approach is able to extract more information on the structural parameters from the data. For instance, it appears that the monetary policy parameter such as ψ_1 can be more precisely estimated with the quadratic approximation (An and Schorfheide, 2007, p. 164).*

I confirm these alluring results by checking the rank criteria for a local point, as well as for the prior domain. Table 8.2 shows that across criteria, a second-order approximation yields additional restrictions on moments and polyspectra, so as to identify all parameters of the model in the vicinity of the local point. This holds for each statistic individually, as well as for the complete set of dynamic properties. In other words, breaking with certainty equivalence, one obtains

Table 8.2: Rank analysis: nonlinear An/Schorfheide (2007) model

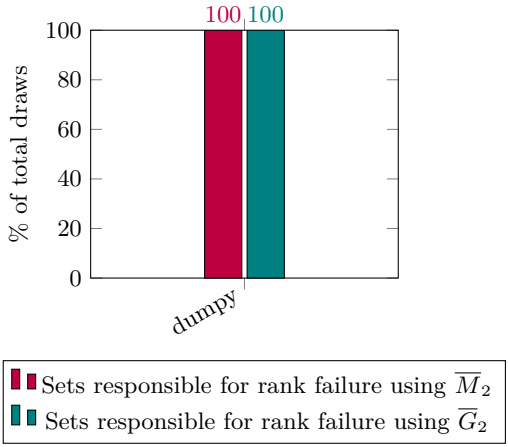
	Iskrev					*Qu and Tkachenko*			
Tol	$\overline{J}$	$\overline{M}_2$	$\overline{M}_3$	$\overline{M}_4$	$\overline{M}$	$\overline{G}_2$	$\overline{G}_3$	$\overline{G}_4$	$\overline{G}$
1e-03	15	12	15	15	15	7	7	6	7
1e-05	15	14	15	15	15	11	11	12	12
1e-07	15	15	15	15	15	13	14	14	14
1e-09	15	15	15	15	15	13	14	15	15
1e-11	15	15	15	15	15	14	15	15	15
1e-13	15	15	15	15	15	15	15	15	15
1e-15	15	15	15	15	15	15	15	15	15
robust	15	15	15	15	15	15	15	15	15
full	15	15	15	15	15	15	15	15	15

Notes: Ranks of identification tests given local point, different tolerance levels Tol and analytical derivatives. Lags for Iskrev's method set to $T = 30$, and subintervals for Qu and Tkachenko's method set to $N = 10000$ for $\overline{G}_2$, $N = 1000$ for $\overline{G}_3$, $N = 100$ for $\overline{G}_4$ and $N = 100$ for $\overline{G}$.

additional information, such that one is able to identify previously non-identifiable parameters. In fact, the first two moments of the PSS already contain enough departure from linearity and Gaussianity and therefore enough restrictions to identify all parameters. The same result holds when I repeat the analysis for 100 random draws from the prior domain, again including a parameter dumpy that does not enter the model. As can be seen in Figure 7.2 in Chapter 7 for a first-order approximation, $(\psi_1, \psi_2, \rho_R, \sigma_R)$, (ϕ, ν) and c/y enter the problematic sets, whereas in the second-order approximation, in all cases, all parameters (apart from dumpy) are identifiable, see Figure 8.2.[39]

[39] Morris (2014) also shows that ν and ϕ are separately identifiable. As a robustness check for the Taylor rule coefficients, I compared the spectral density evaluated at θ_0 with the spectral densities evaluated at one hundred points from the non-identification curve (fixing all parameters except the Taylor rule coefficients). Non-identification curves are defined in Qu and Tkachenko (2012). If parameters are not identified, points on this curve

Figure 8.2: Non-identified sets: nonlinear An/Schorfheide (2007)



Notes: Identification results for 100 draws from the prior domain using analytical derivatives with robust tolerance level, $T = 30$ and $N = 10\ 000$. Sets by brute-force method.

In summary, I confirm An and Schorfheide (2007)'s approach to estimating the model using a second-order approximation. Breaking with certainty equivalence yields additional information that can be used to identify all parameters of the model. The identifiability of the Taylor rule coefficients through the nonlinear model is – as far as I know – new to the literature.

yield the same spectral density at all frequencies apart from an approximation error; whereas if parameters are identified, the spectral densities differ. I found maximum relative and absolute deviations in the order of 10^{-4} for the first 100 points away from θ_0, which is larger than the implied approximation error of 10^{-5} (step size used in the Euler method), and keep growing. I also used the points reported in Table 1 of Qu and Tkachenko (2012) and found maximum relative and absolute deviations in the order of 10^{+4}. These findings provide further support for the result.

Chapter 9

Conclusion

> *It is important to note that the problem is not one of the appropriateness of a particular estimation technique. In the situation described [without additional restrictions], there clearly exists no way using any technique whatsoever in which the true [model] can be estimated (Fisher, 1966, p. 5).*

Theoretically, this thesis adds to the literature on the local identification of nonlinear and non-Gaussian DSGE models. It gives applied researchers a strategy to detect identification problems and means to avoid them in practice. Therefore, the ideas and procedures are useful both from theoretical and applied points of view.

To this end, I provide a comprehensive review and comparison of existing methods for local identification of DSGE models in a consistent notation and framework. In theory, the methods should provide the same conclusions, however, the issue of numerical errors due to nonlinearities and very large matrices can make these methods tedious in practice and may lead to unreliable or contradictory conclusions. Therefore, in order to thoroughly analyze identification of a model, one has to be aware of the drawbacks of the different methods and check whether the methods come to the same conclusion. The example models show that by evaluating different crite-

ria, one also gains inside into the dynamic structure of the DSGE model, in particular sets of parameters that do not enter the mean or reduced-form restrictions.

> *Still, it appears that a large class of popular DSGE structures are only very weakly identified; observational equivalence is widespread; and reasonable estimates are obtained not because the model and the data are informative but because auxiliary restrictions make the likelihood of the data (or a portion of it) informative (Canova and Sala, 2009, p. 448).*

The methods based on ranks can be implemented without using any data, they are a *yes or no* condition. Data-based indicators may give insight into the strength of identification. However, the Bayesian learning rate indicator crucially depends on the ability of the optimizer to find the mode or the convergence of the MCMC sampling algorithm. Having non-identifiable parameters in the algorithms amplifies this problem; hence I advise to exclude these from the Bayesian analysis by first checking the rank criteria.

> *In sum, new-Keynesian models specify policy rules that are a snake pit for econometricians. There is no basis for all the obvious devices, such as excluding variables from the policy rule, using instruments, assuming that the right-hand variables of policy rules are orthogonal to the disturbance, or restricting lag length of disturbances. (...) Not only might these problems exist, but theory predicts that most of them do exist. Empiricists must throw out important elements of the theory in order to identify parameters (Cochrane, 2011, p. 568).*

> *One can only begin to get around these central problems by strong assumptions, in particular that the central bank does not respond to many variables, and to natural rate shocks in particular, in ways that would help it to stabilize the economy (Cochrane, 2011, p. 607).*

Identification of the Taylor rule is particularly problematic, as is shown for the An and Schorfheide (2007) model. This lack of identification will likely prevail in other DSGE models, since monetary

authority is almost always modelled as a feedback rule. In this thesis, I show that using an output growth rule instead of an output gap rule can add auxiliary restrictions to the linearized model in order to identify the Taylor rule.

Another approach, established in this thesis, is to consider higher-order statistics and higher-order approximations. For this reason, I derive formal rank criteria for a local identification of the deep parameters of a nonlinear or non-Gaussian DSGE model, using the pruned state-space system and higher-order statistics. The procedures can be implemented prior to actually estimating the nonlinear model. In this way, I demonstrate the identifiability of the Kim (2003) and the An and Schorfheide (2007) model, when solved by a second-order approximation. For both models, the first four moments and polyspectra contain, together and individually, additional restrictions, which can be used to estimate sets of parameters that are not identified in the first-order approximation. Unfortunately, the proposed rank conditions do not point towards a specific estimation method. An and Schorfheide (2007) show that using a particle filter weakly enhances identifiability of the parameters of their model. Ivashchenko and Mutschler (2015) use the Central Difference Kalman filter and the Quadratic Kalman filter on the pruned as well as unpruned second-order approximation of the Kim (2003) model. They also find that the problematic parameters are separately estimable. Economically speaking, this means that estimating the nonlinear model is a way to solve the functional equivalence between intertemporal and multisectoral investment adjustment costs; thus enriching model dynamics in the neoclassical growth model.

Even though the exposition is based on the second-order approximation, an extension to higher-orders is straightforward, since the pruned state-space always results in a system which is linear in an extended state vector. The propositions and code can also be used for linear DSGE models with non-Gaussian innovations. A

further extension would be to establish rank criteria for other DSGE model specifications, as long as one is able to calculate moments or the spectrum of the data-generating process. For instance, Bianchi (2015) derives analytical moments for Markov switching models, which can be used in a similar fashion to check identification via rank criteria for Markov switching DSGE models.

Lastly, another contribution of this thesis is a computational one. It provides a set of useful matrix concepts and tools for the fast computation of higher-order statistics given both the Gaussian as well as Student's t-distribution as the underlying shock process. Moreover, procedures for computing analytical derivatives of unconditional moments, cumulants and polyspectra for higher-order approximations are provided. The code is written model-independent and can be used easily to check identification of other models, as long as they can be represented in the proposed framework.

In summary, I believe that this dissertation provides useful new tools before actually taking nonlinear or non-Gaussian DSGE models to data. In particular, an applied researcher can check whether (sets of) unidentified parameters in the linearized model may be estimable for higher-order approximations, prior to actually using (tedious) estimation methods for the nonlinear model. Furthermore, given information from higher-order moments or polyspectra the procedures may also provide guidance for moment-matching estimation approaches or particle likelihood-type estimators.

References

An, S. and F. Schorfheide (2007). Bayesian Analysis of DSGE Models. *Econometric Reviews 26*(2-4), 113–172.

Anderson, B. D. and J. B. Moore (1979). *Optimal filtering*. Prentice Hall Information And System Sciences Series. New Jersey: Prentice Hall.

Anderson, G. (2008). Solving Linear Rational Expectations Models: A Horse Race. *Computational Economics 31*(2), 95–113.

Andreasen, M. M. (2010). How to Maximize the Likelihood Function for a DSGE Model. *Computational Economics 35*(2), 127–154.

Andreasen, M. M. (2011). Non-linear DSGE models and the optimized central difference particle filter. *Journal of Economic Dynamics & Control 35*(10), 1671–1695.

Andreasen, M. M. (2013). Non-linear DSGE models and the central difference Kalman filter. *Journal of Applied Econometrics 28*(6), 929–955.

Andreasen, M. M., J. Fernández-Villaverde, and J. F. Rubio-Ramírez (2014). The Pruned State-Space System for Non-Linear DSGE Models: Theory and Empirical Applications. Working paper, NBER.

Antsaklis, P. and A. Michel (2007). *A Linear Systems Primer*. Basel: Birkhäuser.

Bai, J. and S. Ng (2005). Tests for Skewness, Kurtosis, and Normality for Time Series Data. *Journal of Business & Economic Statistics 23*(1), 49–60.

Bao, Y. (2013). On Sample Skewness and Kurtosis. *Econometric Reviews 32*(1-4), 415–448.

Bianchi, F. (2015). Methods for measuring expectations and uncertainty in Markov-Switching models. *Journal of Econometrics (to appear)*, forthcoming.

Binning, A. (2013). Third-Order Approximation of Dynamic Models without the Use of Tensors. Norges Bank Working Paper Series Number 13, Norges Bank.

Blanchard, O. J. and C. M. Kahn (1980). The Solution of Linear Difference Models under Rational Expectations. *Econometrica 48*(5), 1305–1311.

Brillinger, D. R. (1965). An Introduction to Polyspectra. *The Annals of Mathematical Statistics 36*(5), 1351–1374.

Brillinger, D. R. (2001). *Time Series: Data Analysis and Theory*. Classics in Applied Mathematics (Book 36). San Francisco: S.I.A.M.

Caglar, E., J. S. Chadha, and K. Shibayama (2012). Bayesian Estimation of DSGE Models: Is the Workhorse Model Identified? Koç University-TUSIAD Economic Research Forum Working Papers 1205, Koç University.

Calvo, G. (1983). Staggered prices in a utility-maximizing framework. *Journal of Monetary Economics 12*(3), 383–398.

Canova, F. and L. Sala (2009). Back to square one: Identification issues in DSGE models. *Journal of Monetary Economics 56*(4), 431–449.

Chandran, V. and S. Elgar (1994). A general procedure for the derivation of principal domains of higher-order spectra. *IEEE Transactions on Signal Processing 42*(1), 229–233.

Chib, S. and S. Ramamurthy (2014). DSGE Models with Student-t Errors. *Econometric Reviews 33*(1-4), 152–171.

Christoffel, K., G. Coenen, and A. Warne (2008). The New Area-Wide Model of the euro area: a micro-founded open-economy model for forecasting and policy analysis. ECB Working Paper Series 0944, European Central Bank.

Cochrane, J. H. (2011). Determinacy and Identification with Taylor Rules. *Journal of Political Economy 119*(3), pp. 565–615.

Curdia, V., M. del Negro, and D. L. Greenwald (2014). Rare Shocks, Great Recessions. *Journal of Applied Econometrics 29*(7), 1031–1052.

Dagenais, M. G. and D. L. Dagenais (1997). Higher moment estimators for linear regression models with errors in the variables. *Journal of Econometrics 76*(1), 193–221.

Deistler, M. (1976). The Identifiability of Linear Econometric Models with Autocorrelated Errors. *International Economic Review 17*(1), pp. 26–46.

Deistler, M. and H.-G. Seifert (1978). Identifiability and Consistent Estimability in Econometric Models. *Econometrica 46*(4), pp. 969–980.

Dixit, A. K. and J. E. Stiglitz (1977). Monopolistic Competition and Optimum Product Diversity. *American Economic Review 67*(3), 297–308.

Driscool, J. C. (2000). On the Microfoundations of Aggregate Demand and Aggregate Supply. Manuscript, Brown University.

Erickson, T. and T. M. Whited (2002). Two-step GMM estimation of the errors-in-variables model using high-order moments. *Econometric Theory 3*, 776–799.

Fernández-Villaverde, J. and J. F. Rubio-Ramírez (2007). Estimating Macroeconomic Models: A Likelihood Approach. *Review of Economic Studies 74*(4), 1059–1087.

Fisher, F. M. (1966). *The Identification Problem in Econometrics*. New York: McGraw-Hill.

Frisch, R. (1934). *Statistical confluence analysis by means of complete regression systems*, Volume 5. Oslo and Stockholm: Klara Civiltryckeri.

Fuhrer, J. C. (2000). Habit Formation in Consumption and its Implications for Monetary-Policy Models. *American Economic Review 90*(3), 367–390.

Gomes, S., P. Jacquinot, and M. Pisani (2012). The EAGLE. A model for policy analysis of macroeconomic interdependence in the euro area. *Economic Modelling 29*(5), 1686–1714.

Gomme, P. and P. Klein (2011). Second-order approximation of dynamic models without the use of tensors. *Journal of Economic Dynamics and Control 35*(4), 604–615.

Haavelmo, T. (1944). The Probability Approach in Econometrics. *Econometrica 12*, Supplement.

Hannan, E. J. (1971). The Identification Problem for Multiple Equation Systems with Moving Average Errors. *Econometrica 39*(5), 751–765.

Hannan, E. J. (1976). The Identification and Parameterization of Armax and State Space Forms. *Econometrica 44*(4), 713–723.

Herbst, E. and F. Schorfheide (2014). Sequential Monte Carlo Sampling for DSGE Models. *Journal of Applied Econometrics 29*(7), 1073–1098.

Hsiao, C. (1983). Identification. In Z. Griliches and Intriligator M. D. (Eds.), *Handbook of Econometrics*, Volume 1, pp. 223–283. New York: Elsevier.

Ireland, P. N. (2004). Money's Role in the Monetary Business Cycle. *Journal of Money, Credit and Banking 36*(6), 969–983.

Iskrev, N. (2008). Evaluating the information matrix in linearized DSGE models. *Economics Letters 99*(3), 607–610.

Iskrev, N. (2010a). Local identification in DSGE models. *Journal of Monetary Economics 57*(2), 189–202.

Iskrev, N. I. (2010b). Parameter identification in Dynamic Economic models. Economic Bulletin Autumn, Banco de Portugal: Economics and Research Department.

Ivashchenko, S. (2014). DSGE Model Estimation on the Basis of Second-Order Approximation. *Computational Economics 43*, 71–82.

Ivashchenko, S. and W. Mutschler (2015). Enriching Dynamics in investment adjustment costs using higher-order approximation and nonlinear estimation methods. CQE Working Paper, Center for Quantitative Economics Münster.

Johnston, M. K., R. G. King, and D. Lie (2014). Straightforward approximate stochastic equilibria for nonlinear Rational Expectations models. Crawford School of Public Policy Research Paper 59, Crawford School of Public Policy.

Kadane, J. B. (1974). The Role of Identification in Bayesian Theory. In S. Fienberg and A. Zellner (Eds.), *Studies in Bayesian econometrics and statistics: in honor of Leonard J. Savage*, Contributions to economic analysis, pp. 175–191. Amsterdam: North-Holland Pub. Co.

Kan, R. (2008). From moments of sum to moments of product. *Journal of Multivariate Analysis 99*(3), 542–554.

Kim, J. (2003). Functional equivalence between intertemporal and multisectoral investment adjustment costs. *Journal of Economic Dynamics and Control 27*(4), 533–549.

Kim, J., S. Kim, E. Schaumburg, and C. A. Sims (2008). Calculating and using second-order accurate solutions of discrete time dynamic equilibrium models. *Journal of Economic Dynamics and Control 32*(11), 3397–3414.

Kimball, M. S. (1995). The Quantitative Analytics of the Basic Neomonetarist Model. *Journal of Money, Credit and Banking 27*(4), 1241–77.

Klein, P. (2000). Using the generalized Schur form to solve a multivariate linear rational expectations model. *Journal of Economic Dynamics and Control 24*(10), 1405–1423.

Kollmann, R. (2015). Tractable Latent State Filtering for Non-Linear DSGE Models Using a Second-Order Approximation and Pruning. *Computational Economics 45*(2), 239–260.

Komunjer, I. (2012). Global Identification in Nonlinear Models With Moment Restrictions. *Econometric Theory 28*(04), 719–729.

Komunjer, I. and S. Ng (2011). Dynamic Identification of Dynamic Stochastic General Equilibrium Models. *Econometrica 79*(6), 1995–2032.

Koop, G. (2003). *Bayesian Econometrics*. New York: John Wiley & Sons.

Koop, G., M. H. Pesaran, and R. P. Smith (2013). On Identification of Bayesian DSGE Models. *Journal of Business and Economic Statistics 31*(3), 300–314.

Koopmans, T. C. (1949). Identification Problems in Economic Model Construction. *Econometrica 17*(2), 27–48.

Koopmans, T. C. (1950). *Statistical inference in dynamic economic models: Cowles Commission Monograph 10*. New York: John Wiley & Sons.

Koopmans, T. C. and W. C. Hood (1953). The Estimation of Simultaneous Linear Economic Relationships. In T. C. Koopmans and W. C. Hood (Eds.), *Studies in econometric method*, pp. 112–199. New York: John Wiley & Sons.

Koopmans, T. C. and O. Reiersøl (1950). The Identification of Structural Characteristics. *Annals of Mathematical Statistics 21*(2), 165–181.

Kotz, S. and S. Nadarajah (2004). *Multivariate t-distributions and their applications*. Cambridge: University Press.

Lan, H. and A. Meyer-Gohde (2013). Pruning in Perturbation DSGE Models - Guidance from Nonlinear Moving Average Approximations. Humbold University Berlin Discussion Papers, Humbold University Berlin.

Laxton, D., S. Mursula, M. Kumhof, and D. Muir (2010). The Global Integrated Monetary and Fiscal Model (GIMF). IMF Working Papers 10/34, International Monetary Fund.

Lindley, D. (1972). *Bayesian Statistics, A Review*. Philadelphia: S.I.A.M.

Lombardo, G. and H. Uhlig (2014). A theory of pruning. ECB Working Paper Series Number 1696, European Central Bank.

Lucas, Robert E., J. and E. C. Prescott (1971). Investment Under Uncertainty. *Econometrica 39*(5), pp. 659–681.

Magnus, J. R. (2010). On the concept of matrix derivative. *Journal of Multivariate Analysis 101*(9), 2200–2206.

Magnus, J. R. and H. Neudecker (1999). *Matrix Differential Calculus with Applications in Statistics and Econometrics* (2 ed.). New York: John Wiley & Sons.

McCallum, B. T. and E. Nelson (1999). Nominal income targeting in an open-economy optimizing model. *Journal of Monetary Economics 43*(3), 553–578.

Meijer, E. (2005). Matrix algebra for higher order moments. *Linear Algebra and its Applications 410*(0), 112–134.

Mendel, J. M. (1991). Tutorial on higher-order statistics (spectra) in signal processing and system theory: theoretical results and some applications. *Proceedings of the IEEE 79*(3), 278–305.

Milani, F. (2012). The Modeling of Expectations in Empirical DSGE Models: A Survey. *Advances in Econometrics 28*, 3–38.

Morris, S. D. (2014). Local identification in nonlinear dynamic stochastic general equilibrium models. Manuscript, Bowdoin College.

Mutschler, W. (2015a). Higher-order statistics for DSGE models. CQE Working Paper, Center for Quantitative Economics Münster.

Mutschler, W. (2015b). Identification of DSGE models - The effect of higher-order approximation and pruning. *Journal of Economic Dynamics and Control 56*, 34–54.

Nikias, C. and A. Petropulu (1993). *Higher-order Spectra Analysis: A Nonlinear Signal Processing Framework*. Prentice Hall Signal Processing Series. New Jersey: Prentice Hall.

Onatski, A. and N. Williams (2010). Empirical and policy performance of a forward-looking monetary model. *Journal of Applied Econometrics 25*(1), 145–176.

Pintelon, R. and J. Schoukens (2001). *System identification: a frequency domain approach*. New York: John Wiley & Sons.

Poirier, D. J. (1998). Revising Beliefs in Nonidentified Models. *Econometric Theory 14*(4), 483–509.

Pollock, D. S. G. (2013). On Kronecker products, tensor products and matrix differential calculus. *International Journal of Computer Mathematics 90*(11), 2462–2476.

Priestley, M. (1983). *Spectral Analysis and Time Series, Two-Volume Set: Volumes I and II*. London: Academic Press.

Qu, Z. (2014). Inference in dynamic stochastic general equilibrium models with possible weak identification. *Quantitative Economics 5*(2), 457–494.

Qu, Z. and D. Tkachenko (2012). Identification and frequency domain quasi-maximum likelihood estimation of linearized dynamic stochastic general equilibrium models. *Quantitative Economics 3*(1), 95–132.

Ratto, M. and N. Iskrev (2011). Identification analysis of DSGE models with DYNARE. MONFISPOL 225149, Joint Research Centre, European Commission.

Ratto, M. and N. Iskrev (2012). Estimation and inference in DSGE models using derivatives of the likelihood. Working paper, Joint Research Centre, European Commission.

Ríos-Rull, J.-V., F. Schorfheide, C. Fuentes-Albero, M. Kryshko, and R. Santaeulália-Llopis (2012). Methods versus substance: Measuring the effects of technology shocks. *Journal of Monetary Economics 59*(8), 826–846.

Rotemberg, J. J. (1982). Sticky Prices in the United States. *Journal of Political Economy 90*(6), 1187–1211.

Roth, M. (2013). On the multivariate t distribution. Technical report, Department of Electrical Engineering - Division of Automatic Control.

Rothenberg, T. J. (1971). Identification in Parametric Models. *Econometrica 39*(3), 577–591.

Rusticelli, E., R. A. Ashley, E. B. Dagum, and D. M. Patterson (2008). A New Bispectral Test for NonLinear Serial Dependence. *Econometric Reviews 28*(1-3), 279–293.

Schmitt-Grohé, S. and M. Uribe (2004). Solving dynamic general equilibrium models using a second-order approximation to the policy function. *Journal of Economic Dynamics and Control 28*(4), 755–775.

Schmitt-Grohé, S. and M. Uribe (2012). What's News In Business Cycles. *Econometrica 80*(6), 2733–2764.

Sims, C. A. (1989). Models and their uses. *American Journal of Agricultural Economics 71*(1), 489–494.

Smets, F. and R. Wouters (2007). Shocks and Frictions in US Business Cycles: A Bayesian DSGE Approach. *American Economic Review 97*(3), 586–606.

Stock, J. H. and F. Trebbi (2003). Retrospectives: Who Invented Instrumental Variable Regression? *The Journal of Economic Perspectives 17*(3), 177–194.

Swami, A., G. Giannakis, and S. Shamsunder (1994). Multichannel ARMA processes. *IEEE Transactions on Signal Processing 42*(4), 898–913.

Swami, A. and J. Mendel (1990). Time and lag recursive computation of cumulants from a state-space model. *IEEE Transactions on Automatic Control 35*(1), 4–17.

Taylor, J. B. (1993). Discretion versus policy rules in practice. *Carnegie-Rochester Conference Series on Public Policy 39*(1), 195–214.

Teyssiére, G. and A. P. Kirman (2011). *Long memory in economics*. Berlin: Springer.

Tobin, J. (1969). A General Equilibrium Approach to Monetary Theory. *Journal of Money, Credit and Banking 1*(1), 15–29.

Valles, J. (1997). Aggregate investment in a business cycle model with adjustment costs. *Journal of Economic Dynamics and Control 21*(7), 1181–1198.

Working, E. J. (1925). The Statistical Determination of Demand Curves. *The Quarterly Journal of Economics 39*(4), 503–543.

Working, E. J. (1927). What Do Statistical Demand Curves Show? *The Quarterly Journal of Economics 41*(2), 212–235.

Wright, P. G. (1928). *The tariff on animal and vegetable oils: Appendix B*. New York: The Macmillan company.

Appendix A

Magnus-Neudecker definition of Hessian

The nonstochastic steady state is given by $\bar{x} = h(\bar{x}, 0, 0|\theta)$, $\bar{y} = g(\bar{x}, 0, 0|\theta)$, $\bar{u} = 0$ and $\bar{f} = f(\bar{x}, \bar{u}, \bar{y}|\theta) = 0$. Define the steady state vector $\bar{\mathrm{z}} := (\overline{x}', \overline{u}', \overline{y}', \overline{x}', \overline{u}', \overline{y}')'$, then the Jacobian $\mathcal{D}f(\bar{\mathrm{z}})$ and Hessian $\mathcal{H}f(\bar{\mathrm{z}})$ of f evaluated at the steady state are defined as:

$$\mathcal{D}f(\bar{\mathrm{z}}) := \begin{pmatrix} \frac{\partial f^1(\bar{\mathrm{z}})}{\partial x'_{t+1}} & \frac{\partial f^1(\bar{\mathrm{z}})}{\partial u'_{t+1}} & \frac{\partial f^1(\bar{\mathrm{z}})}{\partial y'_{t+1}} & \frac{\partial f^1(\bar{\mathrm{z}})}{\partial x'_t} & \frac{\partial f^1(\bar{\mathrm{z}})}{\partial u'_t} & \frac{\partial f^1(\bar{\mathrm{z}})}{\partial y'_t} \\ \vdots & \vdots & \vdots & \vdots & & \\ \frac{\partial f^n(\bar{\mathrm{z}})}{\partial x'_{t+1}} & \frac{\partial f^n(\bar{\mathrm{z}})}{\partial u'_{t+1}} & \frac{\partial f^n(\bar{\mathrm{z}})}{\partial y'_{t+1}} & \frac{\partial f^n(\bar{\mathrm{z}})}{\partial x'_t} & \frac{\partial f^n(\bar{\mathrm{z}})}{\partial u'_t} & \frac{\partial f^n(\bar{\mathrm{z}})}{\partial y'_t} \end{pmatrix} =: \begin{pmatrix} \mathcal{D}f^1(\bar{\mathrm{z}}) \\ \vdots \\ \mathcal{D}f^n(\bar{\mathrm{z}}) \end{pmatrix}$$

$$\mathcal{H}f(\bar{\mathrm{z}}) := \mathcal{D}vec((\mathcal{D}f(\bar{\mathrm{z}}))') := \begin{pmatrix} \mathcal{H}f^1(\bar{\mathrm{z}}) \\ \vdots \\ \mathcal{H}f^n(\bar{\mathrm{z}}) \end{pmatrix}$$

f is of dimension $n \times 1$, the Jacobian $Df(\bar{\mathrm{z}})$ of dimension $n \times (2n_x + 2n_u + 2n_y)$ and the Hessian $Hf(\bar{\mathrm{z}})$ of dimension $n(2n_x + 2n_u + 2n_y) \times (2n_x + 2n_u + 2n_y)$.

Appendix B

Auxiliary matrices and derivatives

Let $n_v = n_x + n_u, n = n_v + n_y$, $g_v = \begin{pmatrix} g_x & g_u \end{pmatrix}$, and

$$h_v = \begin{pmatrix} h_x & h_u \\ 0_{n_u \times n_x} & 0_{n_u \times n_u} \end{pmatrix}.$$

Then the auxiliary solution matrices are given by

$$Q = \begin{bmatrix} h_v' \otimes f_2 \otimes h_v' + I_{n_v} \otimes f_4 \otimes I_{n_v} & I_{n_v} \otimes (f_1 \otimes I_{n_v} + f_2 g_v \otimes I_{n_v}) \end{bmatrix}$$

$$R = (I_n \otimes M')HM, \quad S = \begin{bmatrix} f_1 + f_2 g_v & f_2 + f_4 \end{bmatrix},$$

$$U = f_2 trm\left((I_{n_y} \otimes (\tilde{\eta}\tilde{\eta}'))g_{vv}\right) + trm\left((I_n \otimes N')HN(\tilde{\eta}\tilde{\eta}')\right),$$

$$M = \begin{bmatrix} h_v \\ g_v h_v \\ I_{n_v} \\ g_v \end{bmatrix}, \quad N = \begin{bmatrix} I_{n_v} \\ g_v \\ 0_{n \times (n_v)} \end{bmatrix}, \quad \tilde{\eta} = \begin{pmatrix} 0_{n_x \times n_u} \\ \eta \end{pmatrix}$$

and trm defines the matrix trace of an $nm \times n$ matrix $[Y_1' \, Y_2' \, \ldots \, Y_m']'$ as the $m \times 1$ vector $[tr(Y_1) \; tr(Y_2) \; \ldots \; tr(Y_m)]'$.

Derivative of $\mathbf{Q^{-1}}$ Notice that Q is partitioned into $Q = [Q_1\ Q_2]$,

$$Q_1 = h_v' \otimes f_2 \otimes h_v' + I_{n_v} \otimes f_4 \otimes I_{n_v},\ Q_2 = I_{n_v} \otimes (f_1 + f_2 g_v) \otimes I_{n_v}.$$

Deriving $\mathtt{d}(f_2 g_v)$ using Theorem 1 and mechanically applying Theorem 2 repeatedly, I obtain the derivatives $\mathtt{d}Q_1$ and $\mathtt{d}Q_2$. Now I can use Algorithm 1 to compute $\mathtt{d}Q$. However, I am interested in $\mathtt{d}Q^{-1}$, thus in step 2(b) of Algorithm 1 I also compute the derivative of the inverse using $-Q^{-1}\left[[\mathtt{d}Q_1^i]\ \ [\mathtt{d}Q_2^i]\right]Q^{-1}$ (Magnus and Neudecker, 1999, p. 184) and store it in step 2(c) in the ith column of $\mathtt{d}(Q^{-1})$.

Derivative of R Regarding the derivative of R I first have to derive $\mathtt{d}M$. This can be done in the same fashion, since M is partitioned into $M = (h_v, g_v h_v, I_{n_v}, g_v)'$. $\mathtt{d}h_v$ and $\mathtt{d}g_v$ are known, whereas $\mathtt{d}(g_v h_v)$ can be derived using Theorem 1. Applying Algorithm 1 I get $\mathtt{d}M$, whereas for the transpose one has the following relationship $\mathtt{d}M' = K_{2(n_x+n_y),n_x}\mathtt{d}M$. Now I am able to compute the derivative of R using Theorems 1 and 2.

Derivative of $\mathbf{S^{-1}}$ Since S is similarly partitioned as Q, i.e. $S = [S_1\ S_2]$, the derivative $\mathtt{d}(S^{-1})$ can be calculated analogously to $\mathtt{d}(Q^{-1})$.

Derivative of T T is the sum of two matrices, for which I will derive the derivatives separately. Consider the first part, $f_2 \cdot trm[(I_{n_y} \otimes (\tilde{\eta}\tilde{\eta}'))g_{vv}]$. Since the derivatives of $(\tilde{\eta}\tilde{\eta}')$ and g_{vv} are known, it is straightforward to compute $\mathtt{d}((I_{n_y} \otimes (\tilde{\eta}\tilde{\eta}'))g_{vv})$ applying Theorems 1 and 2. The only slightly difficult part is the matrix trace function. However, Algorithm 1 can be used to overcome this difficulty. In fact, there is only one partition, for which the derivative is known. Now taking the trm of the reshaped matrix in step 2(b) and storing this in step 2(c), I get $\mathtt{d}(trm[(I_{n_y} \otimes (\tilde{\eta}\tilde{\eta}'))g_{vv}])$. Theorem 1 then yields the derivative of $f_2 \cdot trm[(I_{n_y} \otimes (\tilde{\eta}\tilde{\eta}'))g_{vv}]$. The

same steps can be used to derive the derivative of the second part, $trm[(I_n \otimes N')HN(\tilde{\eta}\tilde{\eta}')]$. However, I first have to derive an expression for $\mathtt{d}N$ and $\mathtt{d}N'$. Since N is partitioned, I can use Algorithm 1 to compute $\mathtt{d}N$ and $\mathtt{d}N' = K_{2n,n_v}\mathtt{d}N$.

Derivative of $\mathbf{H_\xi(\mathfrak{z})}$ $H_\xi(\mathfrak{z})$ is given by $D + C(\mathfrak{z} \cdot I_{n_z^2} - A)^{-1}B$ with $\mathfrak{z} = e^{-i\omega} \in \mathbb{C}$. Closed form expressions for $\mathtt{d}A$, $\mathtt{d}B$, $\mathtt{d}C$ and $\mathtt{d}D$ are given in Chapter 5 using Algorithm 1 for partitioned matrices. Thus, I only need the derivative of the inverted expression which is given by

$$\mathtt{d}\left((\mathfrak{z}I_{n_z^2} - A)^{-1}\right) = \left(-(\mathfrak{z}I_{n_z^2} - A)^{'-1} \otimes (\mathfrak{z}I_{n_z^2} - A)^{-1}\right)(-\mathtt{d}A)$$

where I used $\mathtt{d}(X^{-1}) = (-(X')^{-1} \otimes X^{-1})\mathtt{d}X$, see Magnus and Neudecker (1999, p. 184). Thus, computing $\mathtt{d}H_\xi$ is a straightforward application of Theorem 1. The derivative of the conjugate transpose is given by $\mathtt{d}H_\xi^*(\mathfrak{z}) = K_{n_y,n_\xi} conj(\mathtt{d}H_\xi(\mathfrak{z}))$, where $conj$ returns the complex conjugate.

Appendix C

Example for notation and index matrices

When separating matrices and especially Jacobians into states and shocks, I use index matrices to keep track of the corresponding positions of terms. For illustration, consider only the transition of states with $n_x = 2$ and $n_u = 1$. For $i, j = 1, 2$ denote $h^j_{x_i} := \frac{\partial h^j(\bar{x}_1, \bar{x}_2, 0)}{\partial x_{i,t-1}}$, $h^j_{x_i u} := \frac{\partial^2 h^j(\bar{x}_1, \bar{x}_2, 0)}{\partial x_{i,t-1} \partial u_t}$, where j corresponds to the jth row of h_v. Similar notation applies to $h^j_u, h^j_{ux_i}, h^j_{x_i u}$ and h^j_{uu}. The solution matrices for states are given by

$$h_v = \begin{bmatrix} h^1_{x_1} & h^1_{x_2} & h^1_u \\ h^2_{x_1} & h^2_{x_2} & h^2_u \\ 0 & 0 & 0 \end{bmatrix}, \qquad h_{vv} = \begin{bmatrix} h^1_{x_1 x_1} & h^1_{x_1 x_2} & h^1_{x_1 u} \\ h^1_{x_2 x_1} & h^1_{x_2 x_2} & h^1_{x_2 u} \\ h^1_{u x_1} & h^1_{u x_2} & h^1_{uu} \\ h^2_{x_1 x_1} & h^2_{x_1 x_2} & h^2_{x_1 u} \\ h^2_{x_2 x_1} & h^2_{x_2 x_2} & h^2_{x_2 u} \\ h^2_{u x_1} & h^2_{u x_2} & h^2_{uu} \\ 0 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{bmatrix}.$$

In order to use notation of Andreasen et al. (2014) I get rid of the zeros and reshape and permute these matrices to get

$$H_{xx} = \begin{bmatrix} h^1_{x_1x_1} & h^1_{x_2x_1} & h^1_{x_1x_2} & h^1_{x_2x_2} \\ h^2_{x_1x_1} & h^2_{x_2x_1} & h^2_{x_1x_2} & h^2_{x_2x_2} \end{bmatrix},$$

$$H_{xu} = \begin{bmatrix} h^1_{x_1u} & h^1_{x_2u} \\ h^2_{x_1u} & h^2_{x_2u} \end{bmatrix}, H_{ux} = \begin{bmatrix} h^1_{ux_1} & h^1_{ux_2} \\ h^2_{ux_1} & h^2_{ux_2} \end{bmatrix}, H_{uu} = \begin{bmatrix} h^1_{uu} \\ h^2_{uu} \end{bmatrix}.$$

This can be accomplished by using the following index matrices:

$$idx_{H_{xx}} = \begin{bmatrix} 1 & 2 & 10 & 11 \\ 4 & 5 & 13 & 14 \end{bmatrix}, \quad idx_{H_{uu}} = \begin{bmatrix} 21 \\ 24 \end{bmatrix}$$

$$idx_{H_{xu}} = \begin{bmatrix} 19 & 20 \\ 22 & 23 \end{bmatrix}, \quad idx_{H_{ux}} = \begin{bmatrix} 3 & 12 \\ 6 & 15 \end{bmatrix}$$

That is, in order to compute e.g. H_{xx} I simply select the corresponding terms from h_{vv} using $idx_{H_{xx}}$. Since I now know the exact positions, I am further able to select the correct rows of $\mathtt{d}h_{vv}$ to compute $\mathtt{d}H_{xx}$.

Appendix D

Product-moments of innovations

First-order approximation

Given a first-order approximation, the innovations are defined as the $n_\xi \times 1$ vector $\xi_{t+1} = u_{t+1}$ with $n_\xi = n_u$ elements. I am interested in product moments $M_{2,\xi} := E(\xi_t \otimes \xi_t)$, $M_{3,\xi} := E(\xi_t \otimes \xi_t \otimes \xi_t)$ and $M_{4,\xi} := E(\xi_t \otimes \xi_t \otimes \xi_t \otimes \xi_t)$ with n_ξ^2, n_ξ^3 and n_ξ^4 elements, respectively. These, however, contain many duplicate elements. Denote with $\widetilde{M}_{k,\xi}$ the unique elements of $M_{k,\xi}$, one has the following relationships:

$$M_{2,\xi} = DP_{n_\xi} \cdot \widetilde{M}_{2,\xi}, \quad M_{3,\xi} = TP_{n_\xi} \cdot \widetilde{M}_{3,\xi}, \quad M_{4,\xi} = QP_{n_\xi} \cdot \widetilde{M}_{4,\xi},$$

with the duplication matrix DP_{n_ξ} defined by Magnus and Neudecker (1999, Ch. 3, Sec. 8), and the triplication matrix TP_{n_ξ} and quadruplication matrix QP_{n_ξ} similarly defined by Meijer (2005). These matrices are independent of θ and their Moore-Penrose pseudoinverse always exists, e.g. $(QP'_{n_\xi} QP_{n_\xi})^{-1} QP'_{n_\xi} \cdot M_{4,\xi} = \widetilde{M}_{4,\xi}$. Furthermore, DP_{n_ξ}, TP_{n_ξ} and QP_{n_ξ} are constructed such that there is a unique ordering in $\widetilde{M}_{k,\xi}$, see Meijer (2005) for an example and more details.

To compute the product-moments of ξ_t symbolically I therefore use the following procedure in MATLAB given the number of shocks n_u and the order of product moments $k = 2, 3, 4$.

1. Define $u_{t+1} = (u_{t+1,1}, \dots u_{t+1,n_u})'$ and $\Sigma = [\Sigma_{ij}]_{nu \times nu}$ symbolically with $i, j = 1, \dots n_u$.

2. Get all integer permutations of $[i_1, i_2, \dots i_{n_\xi}]$ that sum up to k, with $i_j = 1, \dots, k$ and $j = 1, \dots, n_\xi$. Sort them in the ordering of Meijer (2005).

3. For each permutation $[i_1, i_2, \dots i_{n_\xi}]$ evaluate symbolically

$$E\left[(\xi_{1,t})^{i_1} \cdot (\xi_{2,t})^{i_2} \cdot \dots (\xi_{n_\xi,t})^{i_{n_\xi}}\right]$$

   and store it in the vector $\widetilde{M}_{k,\xi}$.

The expressions in step 3 contain terms of the form

$$const. \cdot E[(u_{1,t+1})^{i_{u_1}} \cdot (u_{2,t+1})^{i_{u_2}} \cdot \cdot (u_{n_u,t+1})^{i_{u_{n_u}}}],$$

that is joint product moments of the elements of u_{t+1}. Given a function that evaluates the moment structure of u_{t+1} either analytically or numerically, I am able to calculate these terms individually and save them into script files. Note, that these computations need only to be done once for a model, after that one simply evaluates the script files numerically given model parameters θ. The code can evaluate product moments from the Gaussian distribution as well as Student's t-distribution.

Normal distribution In the case that u_t is normally distributed, the joint product moments are functions of the variances and covariances in Σ and can be computed analytically. To this end, I use the very efficient method and MATLAB function of Kan (2008) to derive these joint product moments symbolically. The cumulants can then be computed as outlined in Chapter 4.

Student's t-distribution In the case that u_t is t-distributed with v degrees of freedom, u_t is rewritten in terms of an Inverse-Gamma distributed variable $W = v^{-1/2} \sim IGAM(v/2, v/2)$, and a normally distributed variable $\varepsilon_t \sim N(0, \Sigma)$, $u_t = v^{-1/2}\varepsilon_t$ (similar to Kotz and Nadarajah (2004) or Roth (2013)). W and ε_t are independent, i.e. $E(u_t u_t') = E(W)E(\varepsilon_t \varepsilon_t') = \frac{v}{v-2}\Sigma$. Whereas all odd product moments of u_t are zero, the even product moments ($n = \sum_{j=1}^{n_u} i_{u_j}$ is an even number) are given by

$$E[(u_{1,t})^{i_{u_1}} \cdot (u_{2,t})^{i_{u_2}} \cdot \ldots \cdot (u_{n_u,t})^{i_{u_{n_u}}}] = E[W^{\frac{n}{2}}] \cdot E[(\varepsilon_{1,t})^{i_{u_1}} \cdot (\varepsilon_{2,t})^{i_{u_2}} \cdot \ldots \cdot (\varepsilon_{n_u,t})^{i_{u_{n_u}}}].$$

The first term is equal to $E[W^n] = \frac{v/2}{(v/2-1)\ldots(v/2-n)}$ and since ε_t is multivariate normal, I can use Kan (2008)'s procedure and MATLAB function for the second product. The cumulants can then be computed as outlined in Chapter 4.

Second-order approximation

Given a second-order approximation[40], the innovations are defined as the $n_\xi \times 1$ vector

$$\xi_{t+1} = \Big(u_{t+1}' \quad (u_{t+1} \otimes u_{t+1} - vec(\Sigma))' \quad (u_{t+1} \otimes \hat{x}_t^f)' \quad (\hat{x}_t^f \otimes u_{t+1})' \Big)'$$

with $n_\xi = n_u + n_u^2 + 2n_x n_u$ elements. I am interested in product-moments $M_{2,\xi} := E(\xi_t \otimes \xi_t)$, $M_{3,\xi} := E(\xi_t \otimes \xi_t \otimes \xi_t)$ and $M_{4,\xi} := E(\xi_t \otimes \xi_t \otimes \xi_t \otimes \xi_t)$ with n_ξ^2, n_ξ^3 and n_ξ^4 elements, respectively. In order to compute these objects efficiently, I first reduce the dimension of ξ_t, since it has some duplicate elements. That is, I compute product-moments for the $n_{\tilde{\xi}} = n_u + n_u(n_u+1)/2 + n_u n_x$ vector

$$\tilde{\xi}_{t+1} := \Big(u_{t+1}' \quad vech(u_{t+1}u_{t+1}' - \Sigma)' \quad (u_{t+1} \otimes \hat{x}_t^f)' \Big)'$$

[40] For a third-order approximation, see Mutschler (2015a).

since

$$\xi_t = \begin{pmatrix} I & 0 & 0 \\ 0 & DP_{n_u} & 0 \\ 0 & 0 & I \\ 0 & 0 & K_{n_x,n_u} \end{pmatrix} \tilde{\xi}_t := F_\xi \cdot \tilde{\xi}_t$$

with DP_{n_u} being the duplication matrix and K_{n_x,n_u} the commutation matrix such that $K_{n_x,n_u}(u_{t+1} \otimes \hat{x}_t^f) = (\hat{x}_t^f \otimes u_{t+1})$. Then $M_{k,\xi} := [\otimes_{j=1}^k F_\xi] \cdot M_{k,\tilde{\xi}}$ with $M_{k,\tilde{\xi}}$ denoting the kth-order product-moment of $\tilde{\xi}_t$. Since $[\otimes_{j=1}^k F_\xi]$ does not change with θ, I can focus on $M_{k,\tilde{\xi}}$. $M_{k,\tilde{\xi}}$, however, contains also many duplicate elements. Denote with $\widetilde{M}_{k,\tilde{\xi}}$ the unique elements of $M_{k,\tilde{\xi}}$, for which one has the following relationships:

$$M_{2,\tilde{\xi}} = DP_{n_{\tilde{\xi}}} \cdot \widetilde{M}_{2,\tilde{\xi}}, \quad M_{3,\tilde{\xi}} = TP_{n_{\tilde{\xi}}} \cdot \widetilde{M}_{3,\tilde{\xi}}, \quad M_{4,\tilde{\xi}} = QP_{n_{\tilde{\xi}}} \cdot \widetilde{M}_{4,\tilde{\xi}},$$

with the duplication matrix $DP_{n_{\tilde{\xi}}}$ defined by Magnus and Neudecker (1999, Ch. 3, Sec. 8), and the triplication matrix $TP_{n_{\tilde{\xi}}}$ and quadruplication matrix $QP_{n_{\tilde{\xi}}}$ similarly defined by Meijer (2005).[41] These matrices are independent of θ and their Moore-Penrose pseudoinverse always exists, e.g. $(QP'_{n_{\tilde{\xi}}} QP_{n_{\tilde{\xi}}})^{-1} QP'_{n_{\tilde{\xi}}} \cdot M_{4,\tilde{\xi}} = \widetilde{M}_{4,\tilde{\xi}}$. Furthermore, $DP_{n_{\tilde{\xi}}}$, $TP_{n_{\tilde{\xi}}}$ and $QP_{n_{\tilde{\xi}}}$ are constructed such that there is a unique ordering in $\widetilde{M}_{k,\tilde{\xi}}$, see Meijer (2005) for an example and more details.

To compute the product-moments of $\tilde{\xi}_t$ symbolically I therefore use the following procedure in MATLAB given the number of shocks n_u, the number of state variables n_x and the order of product-moments $k = 2, 3, 4$. Note that these computations can be used for *any* DSGE model with n_u shocks and n_x states.

[41] Actually $\widetilde{M}_{k,\tilde{\xi}}$ has some further duplicate terms for $n_u, n_x > 1$ due to higher-order cross terms of u_{t+1} and $\hat{x}_t^f$, which I can further reduce using indices from the `unique` function of MATLAB.

1. Define $u_{t+1} = (u_{t+1,1}, \dots u_{t+1,n_u})'$, $\hat{x}_t^f = (\hat{x}_{t,1}^f, \dots \hat{x}_{t,n_x}^f)'$ and $\Sigma = [\Sigma_{ij}]_{n_u \times n_u}$ symbolically with $i, j = 1, \dots n_u$. Set up

$$\tilde{\xi}_t = (u_t', vech(u_{t+1}u_{t+1}' - \Sigma)', (u_{t+1} \otimes \hat{x}_t^f)')'.$$

2. Get all integer permutations of $[i_1, i_2, \dots i_{n_{\tilde{\xi}}}]$ that sum up to k, with $i_j = 1, \dots, k$ and $j = 1, \dots, n_{\tilde{\xi}}$. Sort them in the ordering of Meijer (2005).

3. For each permutation $[i_1, i_2, \dots i_{n_{\tilde{\xi}}}]$ evaluate symbolically

$$E\left[(\tilde{\xi}_{1,t})^{i_1} \cdot (\tilde{\xi}_{2,t})^{i_2} \cdot \ \dots \ \cdot (\tilde{\xi}_{n_{\tilde{\xi}},t})^{i_{n_{\tilde{\xi}}}}\right]$$

   and store it in the vector $\widetilde{M}_{k,\tilde{\xi}}$.

4. Optionally: Use MATLAB's `unique` function to further reduce the dimension of $\widetilde{M}_{k,\tilde{\xi}}$.

The expressions in step 3 contain terms of the general form

$$const. \cdot E[(u_{1,t+1})^{i_{u_1}} \cdot (u_{2,t+1})^{i_{u_2}} \cdot \ldots \cdot (u_{n_u,t+1})^{i_{u_{n_u}}}] \cdot \\ \cdot E[(\hat{x}_{1,t}^f)^{i_{x_1}} \cdot (\hat{x}_{2,t}^f)^{i_{x_2}} \cdot \ldots \cdot (\hat{x}_{n_x,t}^f)^{i_x^{n_x}}],$$

that is joint product-moments of the elements of u_{t+1} and $\hat{x}_t^f$ (keeping in mind that $\hat{x}_t^f$ and u_{t+1} are independent due to the temporal independence of u_t). For instance, for $n_u = n_x = 1$ the third-order product-moment of $\tilde{\xi}_t$ is equal to

$$\widetilde{M}_{3,\tilde{\xi}} = vec\left(E\begin{bmatrix} u^3 & u^4 - \sigma_u^2 u^2 \\ u^3 x & \sigma_u^4 u - 2\sigma_u^2 u^3 + u^5 \\ xu^4 - \sigma_u^2 x u^2 & u^3 x^2 \\ -\sigma_u^6 + 3\sigma_u^4 u^2 - 3\sigma_u^2 u^4 + u^6 & x\sigma_u^4 u - 2x\sigma_u^2 u^3 + xu^5 \\ u^4 x^2 - \sigma_u^2 u^2 x^2 & u^3 x^3 \end{bmatrix}' \right)$$

where sub- and superscripts are dropped and $\Sigma = E(u_t^2) = \sigma_u^2$. Given a function that evaluates the moment structure of $\hat{x}_t^f$ and u_{t+1} either analytically or numerically, I am able to calculate these terms individually and save them into script files. The code can evaluate product-moments from the Gaussian distribution as well as Student's t-distribution analytically. Note, that these computations need only to be done once for a model, after that one simply evaluates the script files numerically given model parameters θ.

Normal distribution In the case that u_t is normally distributed, $\hat{x}_t^f$ is also Gaussian with covariance matrix Σ_x given in equation (2.15). Therefore,

$$\begin{pmatrix} u_{t+1} \\ \hat{x}_t^f \end{pmatrix} \sim N\left(\begin{pmatrix} 0 \\ 0 \end{pmatrix}, \begin{pmatrix} \Sigma & 0 \\ 0 & \Sigma_x \end{pmatrix} \right)$$

is multivariate normal. All joint product-moments are functions of the variances and covariances in Σ and Σ_x and can be computed analytically. To this end, I use the very efficient method and MATLAB function of Kan (2008) to derive these joint product-moments symbolically before storing them into script files. For the example with $n_u = n_x = 1$ and Gaussian u_t, one gets the unique entries

$$\widetilde{M}_{2,\xi} = \left[\sigma_u^2,\ 0,\ 0,\ 2\sigma_u^4,\ 0,\ \sigma_u^2\sigma_x^2\right]'$$
$$\widetilde{M}_{3,\xi} = \left[0,\ 2\sigma_u^4,\ 0,\ 0,\ 0,\ 0,\ 8\sigma_u^6,\ 0,\ 2\sigma_u^4\sigma_x^2,\ 0\right]'$$
$$\widetilde{M}_{4,\xi} = \left[3\sigma_u^4, 0, 0, 10\sigma_u^6, 0, 3\sigma_u^4\sigma_x^2, 0, 0, 0, 0, 60\sigma_u^8, 0, 10\sigma_u^6\sigma_x^2, 0, 9\sigma_u^4\sigma_x^4\right]'$$

where $\Sigma_x = E(\hat{x}_t^{f2}) = \sigma_x^2$. The cumulants can then be computed as outlined in Chapter 4. Since the third-order cumulant of a Gaussian process must be zero, ξ_t is clearly non-Gaussian, even if the underlying distribution for u_t is Gaussian.

Student's t-distribution In the case that u_t is t-distributed with v degrees of freedom and covariance matrix $\frac{v}{v-2}\Sigma$, u_t is rewritten in terms of an Inverse-Gamma distributed variable $W = v^{-1/2} \sim IGAM(v/2, v/2)$, and a normally distributed variable $\varepsilon_t \sim N(0, \Sigma)$, $u_t = v^{-1/2}\varepsilon_t$ (see Kotz and Nadarajah (2004, Ch. 1.7) and Roth (2013)). W and ε_t are independent, i.e. $E(u_t u_t') = E(W)E(\varepsilon_t\varepsilon_t') = \frac{v}{v-2}\Sigma$. Whereas all odd product-moments of u_t are zero, the even product-moments ($n = \sum_{j=1}^{n_u} i_{u_j}$ is an even number) are given by

$$E[(u_{1,t})^{i_{u_1}} \cdot (u_{2,t})^{i_{u_2}} \cdot \ldots \cdot (u_{n_u,t})^{i_{u_{n_u}}}] = E[W^{\frac{n}{2}}] \cdot E[(\varepsilon_{1,t})^{i_{u_1}} \cdot (\varepsilon_{2,t})^{i_{u_2}} \cdot \ldots \cdot (\varepsilon_{n_u,t})^{i_{u_{n_u}}}].$$

The first term is equal to $E[W^{\frac{n}{2}}] = \frac{v/2}{(v/2-1)\cdot \ldots \cdot (v/2-n/2)}$ and since ε_t is multivariate normal, one can use Kan (2008)'s procedure and MATLAB function for the second product. Similar arguments apply to the product-moments of $\hat{x}_t^f$, for instance the variance is given by $vec(\Sigma_x) = E[\hat{x}_t^f \otimes \hat{x}_t^f] = E[W] \cdot (I_{n_x^2} - h_x \otimes h_x)^{-1}(h_u \otimes h_u) \cdot E[\varepsilon_t \otimes \varepsilon_t]$. Thus, odd product-moments are also zero, whereas even product-moments can be computed symbolically by Kan (2008)'s procedure and MATLAB function, however, adjusted for $E[W^{n/2}]$.

Analytical derivatives

For analytical derivatives with respect to the model parameters θ, I first collect all auxiliary parameters for $\widetilde{M}_{k,\tilde{\xi}}$ symbolically in a vector θ^{aux} and store the symbolic Jacobians $\frac{\partial \widetilde{M}_{k,\tilde{\xi}}}{\partial \theta^{aux\prime}}$ into script files. In the Gaussian case, $\theta^{aux} = vech(\Sigma)$ for a first-order approximation and $\theta^{aux} = [vech(\Sigma)', vech(\Sigma_x)]'$ for a second-order approximation. In the Student's t case, $\theta^{aux} = \left[\theta^{aux\prime}, v\right]'$. Since $\frac{\partial \theta^{aux}}{\partial \theta'}$ is given by the model, I can evaluate the script files numerically given the identity $\frac{\partial \widetilde{M}_{k,\tilde{\xi}}}{\partial \theta^{aux\prime}} \cdot \frac{\partial \theta^{aux}}{\partial \theta'}$.

Appendix E

Generalized Sylvester equations for cumulants

The zero-lag cumulants

$$\mathcal{C}_{k,z} = (I_{n_z^k} - [\otimes_{j=1}^{k} A])^{-1} \cdot [\otimes_{j=1}^{k} B] \cdot \Gamma_{k,\xi}$$

require the inversion of the matrix $\left(I_{n_z^k} - [\otimes_{j=1}^{k} A]\right)$. Since $\mathcal{C}_{k,z}$ and $\Gamma_{k,\xi}$ are vectors, one can use properties of the Kronecker product and rewrite the equations to

$$\begin{aligned}
[\underset{n_z \times n_z}{\mathcal{C}_{2,z}}] &= A[\underset{n_z \times n_z}{\mathcal{C}_{2,z}}]A' + B[\underset{n_\xi \times n_\xi}{\Gamma_{2,\xi}}]B', \\
[\underset{n_z^2 \times n_z}{\mathcal{C}_{3,z}}] &= (A \otimes A)[\underset{n_z^2 \times n_z}{\mathcal{C}_{3,z}}]A' + (B \otimes B)[\underset{n_\xi^2 \times n_\xi}{\Gamma_{3,\xi}}]B', \\
[\underset{n_z^2 \times n_z^2}{\mathcal{C}_{4,z}}] &= (A \otimes A)[\underset{n_z^2 \times n_z^2}{\mathcal{C}_{4,z}}](A \otimes A)' + (B \otimes B)[\underset{n_\xi^2 \times n_\xi^2}{\Gamma_{4,\xi}}](B \otimes B)',
\end{aligned}$$

where $[\underset{n \times m}{\ }]$ reshapes a $n \cdot m$ vector into a $n \times m$ matrix. In other words, we reduce the inversion problem to a generalized Sylvester equation, which can be efficiently solved.

To compute the analytical derivatives of $\mathcal{C}_{k,z}$, I also use generalized Sylvester equations, an idea similar to Ratto and Iskrev (2012). That is, for each θ_i one takes the differential

$$d[\underset{n_z\times n_z}{\mathcal{C}_{2,z}}] - A\cdot d[\underset{n_z\times n_z}{\mathcal{C}_{2,z}}]\cdot A' = dA[\underset{n_z\times n_z}{\mathcal{C}_{2,z}}]A' + A[\underset{n_z\times n_z}{\mathcal{C}_{2,z}}](dA') + d(B[\underset{n_\xi\times n_\xi}{\Gamma_{2,\xi}}]B'),$$

$$d[\underset{n_z^2\times n_z}{\mathcal{C}_{3,z}}] - (A\otimes A)\cdot d[\underset{n_z^2\times n_z}{\mathcal{C}_{2,z}}]\cdot A' = d(A\otimes A)\cdot[\underset{n_z^2\times n_z}{\mathcal{C}_{3,z}}]\cdot A' + (A\otimes A)[\underset{n_z^2\times n_z}{\mathcal{C}_{3,z}}]\cdot(dA') + d((B\otimes B)[\underset{n_\xi^2\times n_\xi}{\Gamma_{3,\xi}}]B'),$$

$$d[\underset{n_z^2\times n_z^2}{\mathcal{C}_{4,z}}] - (A\otimes A)\cdot d[\underset{n_z^2\times n_z^2}{\mathcal{C}_{4,z}}]\cdot(A\otimes A)' = d(A\otimes A)\cdot[\underset{n_z^2\times n_z^2}{\mathcal{C}_{4,z}}]\cdot(A\otimes A)' + (A\otimes A)[\underset{n_z^2\times n_z^2}{\mathcal{C}_{4,z}}]\cdot d(A'\otimes A') + d((B\otimes B)[\underset{n_\xi^2\times n_\xi^2}{\Gamma_{4,\xi}}](B'\otimes B')),$$

which are also generalized Sylvester equations in the differential on the left hand side. Note that, contrary to the rest of the paper, here I use as notation $dX = \frac{\partial X}{\partial \theta_i}$ to denote the derivative of X w.r.t. to a specific $\theta_i (i = 1, \ldots, n_\theta)$ which has the same shape as X. All terms on the right hand side can be derived using the expressions of Chapter 5.

Appendix F

Deriving numerical derivatives

In order to derive the Jacobian of a function or matrix $F(\theta)$ at a point θ_0 with respect to θ, I use a two-sided finite difference method (also known as central differences). That is:
For each $j = 1, \ldots, n_\theta$

1. Select a step size h_j.
2. Solve the DSGE model twice using $\overline{\theta} := \theta_0 + e_j h_j$ and $\underline{\theta} := \theta_0 - e_j h_j$ with e_j a vector with the jth element equal to 1.
3. Compute

$$\mathtt{d}F^j := \frac{\partial vec(F(\theta_0))}{\partial \theta_j} \approx vec\left(\frac{F(\theta_0 + e_j h_j) - F(\theta_0 - e_j h_j)}{2h_j}\right)$$

4. Store $\mathtt{d}F^j$ as the jth column of $\mathtt{d}F$.

Appendix G

Deriving the minimal state

Consider the general framework by Schmitt-Grohé and Uribe (2004) for the linearized DSGE model (see also footnote 5):

$$\hat{x}_t = h_x \hat{x}_{t-1} + h_u u_t \quad \text{and} \quad \hat{y}_t = g_x \hat{x}_t$$

The vector of states x_t consists of n_{x_2} exogenous states $x_{2,t}$ and n_{x_1} endogenous states $x_{1,t}$. The solution can be rearranged to get

$$\begin{pmatrix} \hat{x}_{1,t} \\ \hat{x}_{2,t} \end{pmatrix} = \begin{pmatrix} 0 & h_{x_1} \\ 0 & h_{x_2} \end{pmatrix} \begin{pmatrix} \hat{x}_{1,t-1} \\ \hat{x}_{2,t-1} \end{pmatrix} + \begin{pmatrix} h_{u_1} \\ h_{u_2} \end{pmatrix} u_t$$

For the controls we therefore get

$$\begin{aligned} \hat{y}_t &= \begin{pmatrix} g_{x_1} & g_{x_2} \end{pmatrix} \begin{pmatrix} 0 & h_{x_1} \\ 0 & h_{x_2} \end{pmatrix} \begin{pmatrix} \hat{x}_{1,t-1} \\ \hat{x}_{2,t-1} \end{pmatrix} + \begin{pmatrix} g_{x_1} & g_{x_2} \end{pmatrix} \begin{pmatrix} h_{u_1} \\ h_{u_2} \end{pmatrix} u_t \\ &= \left(g_{x_1} h_{x_1} + g_{x_2} h_{x_2} \right) \hat{x}_{2,t-1} + \left(g_{x_1} h_{u_1} + g_{x_2} h_{u_2} \right) u_t \end{aligned}$$

Obviously, the driving force of the model is the vector of exogenous states $x_{2,t}$, which is called the minimal state vector. Together

with the evolution of the stochastic innovations it determines the evolution of the endogenous states and the control variables. The minimal representation is thus given by

$$\begin{aligned}\hat{x}_{2,t} &= \overbrace{h_{x_2}}^{\widetilde{h}_x}\hat{x}_{2,t-1} + \overbrace{h_{u_2}}^{\widetilde{h}_u}u_t \\ \hat{y}_t &= \underbrace{(g_{x_1}h_{x_1} + g_{x_2}h_{x_2})}_{\widetilde{g}_x}\hat{x}_{2,t-1} + \underbrace{(g_{x_1}h_{u_1} + g_{x_2}h_{u_2})}_{\widetilde{g}_u}u_t.\end{aligned}$$

Formal conditions for minimality require that there exists a controllable and observable system for every $\theta \in \Theta$. That is, for any initial state it is always possible to design an input sequence that puts the system in the desired final state, i.e. the matrix $CC = \begin{bmatrix}\widetilde{h}_u & \widetilde{h}_x\widetilde{h}_u & \dots & \widetilde{h}_x^{n_{x_2}-1}\widetilde{h}_u\end{bmatrix}$ has full row rank (controllability). Furthermore, given the evolution of the input sequence it is always possible to reconstruct the initial state by observing the evolution of the output sequence, i.e. the matrix $OO = \begin{bmatrix}\widetilde{g}_x' & \widetilde{h}_x'\widetilde{g}_x' & \dots & \widetilde{h}_x^{n_{x_2}-1'}\widetilde{g}_x'\end{bmatrix}'$ has full column rank (observability). See e.g. Anderson and Moore (1979, App. C) for further details about minimality.

For small and medium-sized DSGE models the distinction between endogenous and exogenous states is given through theory: some variables are clearly endogenous (like output) and some are clearly exogenous (like capital or technology). Thus, when setting up the example models, I already order the state vector accordingly. Then, I check the rank conditions for minimality and observability given the full state vector. For big DSGE models the distinction of endogenous and exogenous states is often not as clear. A failsafe approach for deriving the minimal state vector is to consider all possible subsets of combinations of states and check the rank conditions for minimality and controllability in each case. I follow this approach in the code, for a different (computational) approach handling the minimal state in big DSGE models see Kim et al. (2008).

Appendix H

Metropolis-Hastings algorithm

The Bayesian estimation is based on the likelihood, that is a complete characterization of the data generating process. The major distinction to a frequentist estimation is the assumption that the parameter vector θ is a random and not a fixed variable. The Bayesian idea is then to combine information given in the data with additional believes about the parameters, so-called (*priors*), to get an expression for the conditional probability distribution of the parameters. For this reason, one is able to put more weight on a certain span of the parameter space. Bayesian methods can hence be thought of as a bridge between calibration and the *Maximum-Likelihood* approach. In the Bayesian context, the likelihood function $\mathcal{L}(y|\theta)$ is a conditional density of observed data given the parameters: $\wp(y|\theta) = \mathcal{L}(y|\theta)$. Let $\wp(\theta)$ denote the prior density of the vector of parameters θ, then according to Bayes-rule

$$\wp(\theta|y) = \frac{\mathcal{L}(y|\theta)\wp(\theta)}{\wp(y)} = \frac{\mathcal{L}(y|\theta)\wp(\theta)}{\int \wp(\theta)\mathcal{L}(y|\theta)\ d\theta} \propto \mathcal{L}(y|\theta)\wp(\theta),$$

with $\propto$ meaning "proportional to". $\wp(y)$ is the so-called *marginal likelihood* of the data and ultimately just a constant that normalizes the integral to unity. It is independent of the vector of parameters. Removing it, hence, does not change the shape of the posterior density $\wp(\theta|y)$, it merely does not integrate to one. This non-normalized density is often called *posterior-kernel* or, in logs, *log-posterior-kernel*. The mode is a possible Bayesian estimator $\widehat{\theta}_B$ for the true vector of parameters:

$$\widehat{\theta}_B = \underset{\theta}{\operatorname{argmax}} \{\log \wp(\theta|y)\} = \underset{\theta}{\operatorname{argmax}} \{\log \mathcal{L}(y|\theta) + \log \wp(\theta)\}$$

The procedure is to compute the log-likelihood using filtering methods. In linear and Gaussian models one uses the Kalman filter, in nonlinear or non-Gaussian models one uses an extension of the Kalman filter (Andreasen, 2011, 2013; Ivashchenko, 2014; Kollmann, 2015) or the particle filter (An and Schorfheide, 2007; Fernández-Villaverde and Rubio-Ramírez, 2007; Herbst and Schorfheide, 2014). Then the *log-posterior-kernel* can be simulated through sampling or MCMC methods. Once draws from the posterior are available, one can conduct inference on the parameters.

In the literature - and most prominently in DYNARE - the *Random-Walk Metropolis-Hastings (RWMH)* approach has become the workhorse algorithm.[42] The algorithm uses the fact that, under very general regularity conditions, the moments of a distribution are asymptotically normally distributed:

> *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 (An and Schorfheide, 2007, p. 132).*

[42] The presentation of the algorithm follows Koop (2003, pp. 92–99) and An and Schorfheide (2007, p. 131).

The algorithm constructs a sequence of draws, called Markov chains, from a proposal density. This proposal distribution does not need to be identical with the posterior density, but it is only required that the algorithm can draw samples from the whole range of the posterior density. The current candidate draw θ^* is dependent on the previous candidate draw $\theta^{(s-1)}$. Weights for all candidates are the same, however, they are accepted only with a certain probability α. This probability is defined as the ratio of the *posterior-kernel* of the current candidate to the *posterior-kernel* of the previous candidate. It can be shown that given this structure, the algorithm moves from areas of low posterior probability to areas of high probability: Is $\theta^{(s-1)}$ in an area with high probability, it is likely that only candidates in the same area are accepted. Is $\theta^{(s-1)}$ in an area with low probability, it is very likely that new candidates are accepted. The covariance matrix of the proposal distribution plays an important role such that α is neither to large nor to small. Common practice is to use the inverse of the Hessian evaluated at the mode $\widehat{\theta}_B$ and to scale it with a factor c such, that the average acceptance ratio lies in between 20% and 50% (Koop, 2003, p. 98). The algorithm has the following structure (An and Schorfheide, 2007, p. 131):

Algorithm 2 (Random-Walk Metropolis-Hastings)**.**

1. *Specify* c_0, c *and* S*.*
2. *Maximize* $\log \mathcal{L}(y|\theta) + \log \wp(\theta)$ *using numerical methods.* $\widehat{\theta}_B$ *denotes the mode.*
3. *Calculate the inverse of the Hessian evaluated at the mode, denoting it as* Σ_B*.*
4. *Specify an initial value* $\theta^{(0)}$ *or draw it from* $\mathcal{N}(\widehat{\theta}_B, c_0^2\Sigma_B)$*.*
5. *For* $s = 1, \ldots, S$*:*
   - *Draw* θ^* *from the candidate-generating distribution (proposal density)* $\mathcal{N}(\theta^{(s-1)}, c^2\Sigma_B)$*.*

- *Compute the acceptance probability* α*:*

$$\alpha \equiv \alpha\left(\theta^{(s-1)}, \theta^{*}\right) = \frac{\mathcal{L}\left(\theta^{*}|y\right) \ \wp\left(\theta^{*}\right)}{\mathcal{L}\left(\theta^{(s-1)}|y\right) \ \wp\left(\theta^{(s-1)}\right)}$$

- *With probability* $min\{\alpha, 1\}$ *accept the jump to* θ^{*} *from* $\theta^{(s-1)}$*. In other words, if* $\alpha \geq 1$*, set* $\theta^{(s)} = \theta^{*}$*.*
- *Otherwise draw a uniformly distributed variable* r *between 0 and 1.*
  - *If* $r \leq \alpha$ *set* $\theta^{(s)} = \theta^{*}$ *(jump).*
  - *If* $r > \alpha$ *set* $\theta^{(s)} = \theta^{(s-1)}$ *(don't jump).*

6. *Estimate the posterior expectation of a function* $\hbar(\theta)$ *with* $\frac{1}{S}\sum_{s=1}^{S} \hbar\left(\theta^{(s)}\right)$.

If the average acceptance probability does not yield a desirable value (typically between 20% and 50%) or the algorithm does not converge, change c_0, c or S.

Local identification of nonlinear and non-Gaussian DSGE models

Willi Mutschler

This thesis adds to the literature on the local identification of nonlinear and non-Gaussian DSGE models. It gives applied researchers a strategy to detect identification problems and means to avoid them in practice. A comprehensive review of existing methods for linearized DSGE models is provided and extended to include restrictions from higher-order moments, cumulants and polyspectra. Another approach, established in this thesis, is to consider higher-order approximations. Formal rank criteria for a local identification of the deep parameters of nonlinear or non-Gaussian DSGE models, using the pruned state-space system are derived. The procedures can be implemented prior to estimating the nonlinear model. In this way, the identifiability of the Kim (2003) and the An and Schorfheide (2007) model are demonstrated, when solved by a second-order approximation.