Identification of DSGE Models - the Effect of Higher-Order Approximation and Pruning[☆]

Willi Mutschler*

Center for Quantitative Economics, Westfälische Wilhelms-Universität Münster

Abstract

This paper shows how to check rank criteria for a local identification of nonlinear DSGE models, given higher-order approximations and pruning. This approach imposes additional restrictions on (higher-order) moments and polyspectra, which can be used to identify parameters that are unidentified in a first-order approximation. The identification procedures are demonstrated by means of the Kim (2003) and the An and Schorfheide (2007) models. Both models are identifiable with a second-order approximation. Furthermore, analytical derivatives of unconditional moments, cumulants and corresponding polyspectra up to fourth order are derived for the pruned state-space.

Keywords: identification, pruning, higher-order moments, cumulants, polyspectra, nonlinear DSGE, rank condition, analytical derivatives

JEL: C10, C51, C52, E1

[☆]This paper is an improved version of some sections in an earlier paper entitled: Identification of DSGE Models - A Comparison of Methods and the Effect of Second-Order Approximation, which was presented at the course on Identification Analysis and Global Sensitivity Analysis for Macroeconomic Models in Ispra, Italy (2013); at the Fall Macro Midwest Meeting in Minneapolis, USA (2013); at the Spring Meeting of Young Economists in Vienna, Austria (2014); at the Workshop on Empirical Methods in Macroeconomic Policy Analysis in Bucharest, Romania (2014); at the Asian Meeting of the Econometric Society in Taipei, Taiwan (2014); at the European Meeting of the Econometric Society in Toulouse, France (2014); at the Annual Conference of the Verein für Socialpolitik in Hamburg, Germany (2014); and at the DYNARE Conference in Paris, France (2014). The author thanks Nikolay Iskrev, Michel Juillard, Thomas Lubik, Junior Maih, Zhongjun Qu, Marco Ratto, Stephanie Schmitt-Grohé, Ron Smith, Mark Trede, Bernd Wilfing and three anonymous referees for critical and helpful comments and encouragement.

*Corresponding author. Center for Quantitative Economics, Westfälische Wilhelms-Universität Münster, Am Stadtgraben 9, 48143 Münster, Tel.: +49-251-83-22914, Fax: +49-251-83-22012, Email: willi@mutschler.eu.

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. Several formal methods have been proposed to check local identification in linearized DSGE models via rank criteria (Iskrev, 2010; Komunjer & Ng, 2011; Qu & Tkachenko, 2012) or Bayesian indicators (Koop et al., 2013), for a review and methodological comparison of these techniques, see Mutschler (2015b). Whereas there is a growing literature on the nonlinear estimation of DSGE models (Andreasen, 2011, 2013; Fernández-Villaverde & Rubio-Ramírez, 2007; Herbst & Schorfheide, 2014; Ivashchenko, 2014; Kollmann, 2015), all identification methods focus on the linear approximation of the DSGE model to the first order. Furthermore, since the majority of papers assume Gaussian innovations, the proposed criteria are based on first- and second-order moments only. In this paper, we relax these assumptions 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.[1] However, there is a caveat, since higher-order approximations may yield explosive or non-stationary processes. Therefore, we use the pruning scheme proposed by Kim et al. (2008) and operationalized by Andreasen et al. (2014), who show that the pruned state-space (PSS) is a linear, stationary and ergodic state-space system, but with non-Gaussian innovations. 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 with Gaussian innovations.

Exploiting these insights, the contribution of this paper is twofold. First, we extend ideas from Iskrev (2010) and Qu & Tkachenko (2012) to establish rank criteria for higher-order approximations, both in the time and in the frequency domains. To this end, we show how to analytically calculate the Jacobians of unconditional second-, third- and fourth-order moments, cumulants and corresponding polyspectra of the PSS, with respect

[1] By higher-order statistics, we mean moments or cumulants of an order exceeding two in the time domain, and their multidimensional Fourier transform, called polyspectra in the frequency domain.

to the deep parameters of the model. In this manner, we are able to check identification, given theoretical higher-order moments of observables. We limit ourselves 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, we focus 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 our exposition, all methods are applied to two models, namely those of Kim (2003) and An & Schorfheide (2007), which are known to have a lack of identification in their (log-)linearized Gaussian versions. In particular, we show that the parameters governing the adjustment costs in Kim (2003), as well as all parameters including the coefficients of the Taylor rule in An & Schorfheide (2007), 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.

The ideas and procedures derived are useful both from theoretical and applied points of view. Theoretically, this paper adds to the literature on the local identification of nonlinear DSGE models by establishing (i) rank criteria for the pruned state-space representation and (ii) procedures for computing analytical derivatives of unconditional moments, cumulants and polyspectra for higher-order approximations. Based on these findings, we believe that the suggested approach is a useful new tool before actually taking nonlinear DSGE models to data. In particular, an applied researcher can check whether (sets of) unidentified parameters in the linearized model may be estimable, using higher-order approximations, prior to actually using (tedious) nonlinear estimation methods. Furthermore, given information from higher-order moments or polyspectra the procedure may also provide guidance for moment-matching estimation approaches or particle likelihood-type estimators. The present paper is also related to Morris (2014), who likewise derives rank criteria for the pruned state-space (PSS) system, yet in the manner of Komunjer & Ng (2011). The key differences between our work and Morris (2014) can be summarized with respect to three aspects. First, the perspective and system representation is different, since Komunjer & Ng (2011)'s approach assumes a minimal system. For the PSS,

this requires the innovations representation. Our approach neither relies on a specific model framework nor on a minimal system; thus, we do not need to reparametrize the solution system. Second, our criteria 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.

Our MATLAB code is model-independent (a DYNARE implementation is underway) and can be found on the homepage of the author.

2. DSGE framework

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

$$\begin{aligned} E_t f\left(x_{t+1}, u_{t+1}, y_{t+1}, x_t, u_t, y_t|\theta\right) &= 0, \\ x_{t+1} &= h(x_t, u_{t+1}, \sigma|\theta), \\ y_{t+1} &= g(x_t, u_{t+1}, \sigma|\theta), \end{aligned}$$

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, we assume 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.[2]

[2]Because we check criteria based on the first four moments of observables in the pruned state-space, we require at least 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 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.

Apart from the existence of moments and temporal independence, we do not need to impose any distributional assumptions.[3] See Appendix A on how to squeeze the example models into this framework.[4]

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. We follow Schmitt-Grohé & Uribe (2004) and use perturbation techniques to solve the model around the nonstochastic 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, we exploit ideas of Gomme & Klein (2011) to approximate the policy functions, using the Magnus & Neudecker (1999) definition of the Hessian. Denote the Jacobian of f evaluated at the steady state as

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

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.[5]

The approximations of the policy functions are a straightforward application of Taylor series expansions in the state variables. There are several methods and algorithms for

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

[4] This is basically a mixture of the DYNARE framework (innovations enter nonlinearly, no distinction between states and controls) and the framework of Schmitt-Grohé & 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, we are able to add measurement equations and measurement errors by simply extending our 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 & Schorfheide (2007) model in Appendix A.2 for an example.

[5] For recent literature in favor of this definition, see Magnus (2010) and Pollock (2013). The online documentation of the code also contains additional material which clarifies the concept.

calculating the first-order solution matrices, since these are the coefficients of a first-order linearization or log-linearization of the model. We follow Klein (2000) to obtain the first-order solutions $\hat{x}_t = h_x \hat{x}_{t-1} + h_u u_t$ and $\hat{y}_t = g_x \hat{x}_{t-1} + g_u u_t$, using the generalized Schur decomposition.[6] $\hat{x}_t = x_t - \bar{x}$ and $\hat{y}_t = y_t - \bar{y}$ denote deviations from steady state. The second-order solution matrices can be calculated by inserting the policy functions into the model equations 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 & Neudecker (1999, p. 110), evaluating the Jacobian $\mathcal{D}\bar{f}$ and Hessian $H\bar{f}$ of f at the nonstochastic steady state, and setting it to zero yields (after some algebraic calculations, see Gomme & Klein (2011)):

$$\begin{bmatrix} vec(g_{vv}) \\ vec(h_{vv}) \end{bmatrix} = -Q^{-1} vec(R), \qquad \begin{pmatrix} h_{\sigma\sigma} \\ g_{\sigma\sigma} \end{pmatrix} = -S^{-1} U. \tag{2}$$

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 $h_{\sigma\sigma}$ the corresponding terms with respect to the perturbation parameter. Note that all terms are evaluated at the nonstochastic steady state. See Appendix E 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. We separate these using index matrices, see Appendix A for an example that clarifies the notation. A third-order approximation using Magnus-Neudecker Hessians is given in Binning (2013). For our purposes, it is sufficient to note that there are closed-form solutions that we will differentiate in Section 5 with respect to the deep parameters.

3. 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

[6] See Anderson (2008) for a comparison of algorithms, which are basically all equivalent and differ only (slightly) in computational burden. Furthermore, all provide and check the Blanchard & 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.

stable. This is due to artificial fixed points of the approximation, see Kim et al. (2008, p. 3408) for a univariate example. Thus, the model may be neither stationary nor imply an ergodic probability distribution, both of which 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.[7] For instance, given a second-order approximation, we decompose the state vector into first-order ($\hat{x}_t^f$) and second-order ($\hat{x}_t^s$) effects ($\hat{x}_t = \hat{x}_t^f + \hat{x}_t^s$), and set up the law of motions for these variables, preserving only effects up to second-order (see the technical appendix of Andreasen et al. (2014) for details):

$$\hat{x}_{t+1}^f = h_x \hat{x}_t^f + h_u u_{t+1} \tag{3}$$

$$\begin{aligned}\hat{x}_{t+1}^s = {} & h_x \hat{x}_t^s + \frac{1}{2} H_{xx} \left(\hat{x}_t^f \otimes \hat{x}_t^f \right) + \frac{1}{2} H_{uu} \left(u_{t+1} \otimes u_{t+1} \right) \\ & + \frac{1}{2} H_{xu} \left(\hat{x}_t^f \otimes u_{t+1} \right) + \frac{1}{2} H_{ux} \left(u_{t+1} \otimes \hat{x}_t^f \right) + \frac{1}{2} h_{\sigma\sigma} \sigma^2\end{aligned} \tag{4}$$

$$\begin{aligned}\hat{y}_{t+1} = {} & g_x (\hat{x}_t^f + \hat{x}_t^s) + g_u u_{t+1} + \frac{1}{2} G_{xx} \left(\hat{x}_t^f \otimes \hat{x}_t^f \right) + \frac{1}{2} G_{uu} \left(u_{t+1} \otimes u_{t+1} \right) \\ & + \frac{1}{2} G_{xu} \left(\hat{x}_t^f \otimes u_{t+1} \right) + \frac{1}{2} G_{ux} \left(u_{t+1} \otimes \hat{x}_t^f \right) + \frac{1}{2} g_{\sigma\sigma} \sigma^2\end{aligned} \tag{5}$$

with H_{xx} being an $n_x \times n_x^2$ matrix containing all second-order terms for the i-th state variable in the i-th row, whereas G_{xx} is an $n_y \times n_x^2$ matrix containing all second-order terms for the i-th control variable in the i-th 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.[8] Thus, terms containing $\hat{x}_t^f \otimes \hat{x}_t^s$ and $\hat{x}_t^s \otimes \hat{x}_t^s$ 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 $z_t := [(\hat{x}_t^f)',\ (\hat{x}_t^s)',\ (\hat{x}_t^f \otimes \hat{x}_t^f)']'$: equations (3), (4) and (5) can then be rewritten as a linear system of equations called the pruned

[7] This may seem an ad hoc procedure, but pruning can also be founded theoretically as a Taylor expansion in the perturbation parameter (Johnston et al., 2014; Lombardo & Uhlig, 2014) or on an infinite moving average representation (Lan & Meyer-Gohde, 2013). Importantly, the solution matrices are the same.

[8] The proposed notation is the same as in the technical appendix of Andreasen et al. (2014), regarding the DYNARE notation. The second-order solution matrices can be obtained from our Magnus-Neudecker notation, using MATLAB's permute and reshape functions, see Appendix A for a simple example that clarifies the notation used in the code.

state-space:

$$z_{t+1} = c + Az_t + B\xi_{t+1} \tag{6}$$

$$\hat{y}_{t+1} = d + Cz_t + D\xi_{t+1} \tag{7}$$

where

$$\xi_{t+1} := \begin{bmatrix} u_{t+1} \\ u_{t+1} \otimes u_{t+1} - vec(\Sigma) \\ u_{t+1} \otimes x_t^f \\ x_t^f \otimes u_{t+1} \end{bmatrix}, \qquad c := \begin{bmatrix} 0 \\ \frac{1}{2}h_{\sigma\sigma}\sigma^2 + \frac{1}{2}H_{uu}vec(\Sigma) \\ (h_u \otimes h_u)vec(\Sigma) \end{bmatrix},$$

$$d := \begin{bmatrix} \frac{1}{2}g_{\sigma\sigma}\sigma^2 + \frac{1}{2}G_{uu}vec(\Sigma) \end{bmatrix},$$

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

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

Thus, conceptually we work in 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. For instance, 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. Furthermore, if u_t has finite fourth moments, the pruned state-space system then has finite second moments (see Andreasen et al. (2014) for closed-form expressions). We show below that if u_t has finite eighth moments, the pruned state-space system then has finite fourth moments.[9] Note that apart from the existence of moments and temporal independence, we do not need to impose any distributional assumptions on u_t. Even in the (common) case of u_t being normally distributed, ξ_t is clearly non-Gaussian, therefore leaving scope for higher-order moments to contain additional information.

[9]This approach also works for higher-order approximations. That is, appending the state vector accordingly, we are always able to establish a linear system in the extended state vector. See Andreasen et al. (2014) or Mutschler (2015a) for the corresponding matrices of the third-order PSS. y_t and z_t have finite fourth moments, if u_t has finite twelfth moments.

The mean of the extended state vector is equal to

$$\mu_z := E(z_t) = (I_{n_z} - A)^{-1}c. \tag{8}$$

with $n_z = 2n_x + n_x^2$. Intuitively, the mean of the pruned state-space 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, we adjust the mean for risk, given a constant $\frac{1}{2}h_{\sigma\sigma}\sigma^2$ and the variance of the states

$$vec(\Sigma_x) := E(\hat{x}_t^f \otimes \hat{x}_t^f) = (I_{n_x^2} - h_x)^{-1}(h_u \otimes h_u)vec(\Sigma). \tag{9}$$

Since there is a linear relationship between y_t and z_{t-1} in (7), we obtain

$$\mu_y := E(y_t) = \bar{y} + C\mu_z + d. \tag{10}$$

For the derivation of moments, cumulants and spectra, we work with zero mean variables to simplify the notation and expressions in the following sections. Therefore, we denote $\widetilde{z}_t := z_t - \mu_z$ and $\widetilde{y}_t := y_t - \mu_y$ in the following section.

4. Unconditional moments, cumulants and polyspectra up to fourth-order

Most linear DSGE models assume Gaussian innovations.[10] Since a zero mean Gaussian process is characterized completely by its second moment, all identification criteria focus only on information from the autocorrelation structure in the time domain or the power spectrum in the frequency domain. However, whenever we are confronted with nonlinearities or non-Gaussian stochastic innovations as in the pruned state-space, it is natural to analyze whether we are able to exploit information from higher-order moments for the identification and estimation of parameters. Researchers in mathematics, statistics and signal processing have developed tools, called higher-order statistics (HOS), to

[10] Two notable exceptions are Curdia et al. (2014) and Chib & Ramamurthy (2014) who estimate two standard linear DSGE models with Student's t-distributed errors and conclude that these models outperform their Gaussian counterparts. Our 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 B for details.

solve detection, estimation and identification problems when the noise source is non-Gaussian or we are faced with nonlinearities.[11] However, applications in the macroeconometric literature are rather sparse.[12] 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. 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 the n_z^k vectors $\mathcal{C}_{k,z}$ as

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

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, we adopt the compact notation of Swami & Mendel (1990) and store all product-moments of a mean zero vector-valued process in a vector using Kronecker products. 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. 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 the second-order moments. In general, for cumulants higher than three, we need to know the lower-order moments or cumulants. Nevertheless, using cumulants is preferable for several reasons. For instance,

[11] For introductory literature and tutorials, see the textbooks of Brillinger (2001), Nikias & Petropulu (1993), Priestley (1983) and the references therein.

[12] 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 & Dagenais (1997); Erickson & Whited (2002)). For an application to DSGE models, see Mutschler (2015a).

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).[13]

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

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

with $\omega_j \in [-\pi; \pi]$ and imaginary i (see Swami et al. (1994) for further details). The second-, third- and fourth-order spectra 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 spectra.

Standard results from VAR(1) systems and insights from HOS can be used, regarding the computation of unconditional cumulants and polyspectra in the PSS. 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 $u_{t+1} \otimes u_{t+1}$. The kth-order cumulants of ξ_t are

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

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

and corresponding polyspectra $\mathcal{S}_{k,\xi}(\omega_1, \ldots, \omega_{k-1}) = (2\pi)^{1-k}\Gamma_{k,\xi}$ are flat. Regarding the computation of $\Gamma_{k,\xi}$, see also Appendix B.[14] There, we show 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.

Letting $[\otimes_{j=1}^{k} X(j)] = X(1) \otimes X(2) \otimes \cdots \otimes X(k)$ for objects $X(j)$, Swami & Mendel (1990) show that given a zero mean stationary, time-invariant linear state-space system, 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 C). 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 y_t and z_{t-1} in (7), we obtain closed-form expressions for the kth-order cumulants of our observables (after substracting the mean). That is, for $t_j > 0$

$$\mathcal{C}_{k,y}(0, \ldots, 0) = [\otimes_{j=1}^{k} C]\mathcal{C}_{k,z}(0, \ldots, 0) + [\otimes_{j=1}^{k} D]\Gamma_{k,\xi}, \tag{11}$$

$$\mathcal{C}_{k,y}(t_1, \ldots, t_{k-1}) = [\otimes_{j=1}^{k} C]\mathcal{C}_{k,z}(t_1, \ldots, t_{k-1}). \tag{12}$$

Lastly, for the identification criteria based on cumulants, we stack all theoretical second,

[14] $\Gamma_{k,\xi}$ 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. We 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.

third and fourth cumulants into vectors

$$
\begin{aligned}
m_2(\theta, T) &= \Big(\mathcal{C}_{2,y}(0)', \quad \ldots, \quad \mathcal{C}_{2,y}(T-1)'\Big)', \\
m_3(\theta, T) &= \Big(\mathcal{C}_{3,y}(0,0)', \quad \ldots, \quad \mathcal{C}_{3,y}(T-1,T-1)'\Big)', \\
m_4(\theta, T) &= \Big(\mathcal{C}_{4,y}(0,0,0)', \quad \ldots, \quad \mathcal{C}_{4,y}(T-1,T-1,T-1)'\Big)'.
\end{aligned}
$$

Regarding the computation of polyspectra, consider the vector moving average representation (VMA) of $\widetilde{z}_t = z_t - \mu_z = \sum_{j=0}^{\infty} A^j B \xi_{t-j}$. Using equation (7) and lag operator L, we obtain the VMA for our controls

$$
\widetilde{y}_t = y_t - \bar{y} - C\mu_z - d = \sum_{j=0}^{\infty} C A^j B \xi_{t-j-1} + D\xi_t = H_\xi(L^{-1})\xi_t
$$

with transfer function $H_\xi(\mathfrak{z}) = D + C\left(\mathfrak{z} I_{n_z} - A\right)^{-1} B$ for $\mathfrak{z} \in \mathbb{C}$. Setting $\mathfrak{z}_j = e^{-i\omega_j}$, with imaginary i and $\omega_j \in [-\pi; \pi]$, we obtain 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 $\mathcal{S}_{4,y}$:

$$
\mathcal{S}_{2,y}(\omega_1) = (2\pi)^{-1} \left[H(\mathfrak{z}_1^{-1}) \otimes H(\mathfrak{z}_1)\right] \Gamma_{2,\xi}, \tag{13}
$$

$$
\mathcal{S}_{3,y}(\omega_1, \omega_2) = (2\pi)^{-2} \left[H(\mathfrak{z}_1^{-1} \cdot \mathfrak{z}_2^{-1}) \otimes H(\mathfrak{z}_1) \otimes H(\mathfrak{z}_2)\right] \Gamma_{3,\xi}, \tag{14}
$$

$$
\mathcal{S}_{4,y}(\omega_1, \omega_2, \omega_3) = (2\pi)^{-3} \left[H(\mathfrak{z}_1^{-1} \cdot \mathfrak{z}_2^{-1} \cdot \mathfrak{z}_3^{-1}) \otimes H(\mathfrak{z}_1) \otimes H(\mathfrak{z}_2) \otimes H(\mathfrak{z}_3)\right] \Gamma_{4,\xi}. \tag{15}
$$

In summary, the pruned state-space representation is a stable system and has well-defined statistical properties, which we can exploit for our identification analysis. In particular, we see that 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, we may also gain additional information by considering higher-order statistics. In Section 6 we show how to incorporate these additional restrictions into formal identifiability criteria, but first, we discuss the derivatives of these objects.

5. Derivatives of solution matrices, cumulants and polyspectra

To establish rank criteria, we need derivatives of all solution matrices, cumulants and polyspectra with respect to the deep parameters θ. Following ideas from Iskrev (2008)

and Schmitt-Grohé & Uribe (2012, Suppl. Mat., Sec. A.3), we view f, as well as the Jacobian of f, as a function 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, we have $f(\overline{xuy}(\theta), \theta) = 0$. Differentiating yields

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

This expression can easily be obtained analytically using, for example, MATLAB's symbolic toolbox. The derivative of the Jacobian $\mathcal{D}\bar{f}(\overline{xuy}(\theta), \theta)$ 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 according to equation 1 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 $H := \mathcal{H}f(\overline{xuy}(\theta), \theta)$:[15]

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

Our MATLAB code writes all analytical derivatives, using symbolic expressions, into script files for further evaluation. For numerical derivatives, we employ the two-sided central difference method. Note that we use the following notation: $\mathtt{d}X := \frac{\partial vec(X)}{\partial \theta'}$ for the Jacobian of a matrix.

Furthermore, we repeatedly use the commutation matrix $K_{m,n}$ which transforms the $m \times n$ matrix A, such that $K_{m,n} vec(A) = vec(A')$[16], 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*

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

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

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

Proof: Magnus & Neudecker (1999, p. 175). Note that $\mathtt{d}X := \frac{\partial vec(X)}{\partial \theta'}$.

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*

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

Proof: Magnus & Neudecker (1999, p. 185). Note that $\mathtt{d}X := \frac{\partial vec(X)}{\partial \theta'}$.

Moreover, we 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* $\mathtt{d}X_1$ *and* $\mathtt{d}X_2$*;* $\mathtt{d}X_1$ *is of dimension* $(mn_1 \times n_\theta)$ *and* $\mathtt{d}X_2$ *of dimension* $(mn_2 \times n_\theta)$*.*
2. *For* $i = 1, \ldots, n_\theta$
   (a) *Denote the i-th column of* $\mathtt{d}X_1$ *and* $\mathtt{d}X_2$ *as* $\mathtt{d}X_1^i$ *and* $\mathtt{d}X_2^i$ *respectively.* $\mathtt{d}X_1^i$ *is of dimension* $(mn_1 \times 1)$ *and* $\mathtt{d}X_2^i$ *of dimension* $(mn_2 \times 1)$*.*
   (b) *Reshape* $\mathtt{d}X_1^i$ *into a* $(m \times n_1)$ *matrix* $[\mathtt{d}X_1^i]_{(m \times n_1)}$ *and* $\mathtt{d}X_2^i$ *into a* $(m \times n_2)$ *matrix* $[\mathtt{d}X_2^i]_{(m \times n_2)}$*.*
   (c) *Store* $vec(\begin{bmatrix}[\mathtt{d}X_1^i]_{(m \times n_1)} & [\mathtt{d}X_2^i]_{(m \times n_2)}\end{bmatrix})$ *into the i-th column of a matrix* $\mathtt{d}X$*.*
3. $\mathtt{d}X$ *is the derivative of* X *with respect to* θ *and is of dimension* $(mn \times n_\theta)$*.*

Note that $\mathtt{d}X := \frac{\partial vec(X)}{\partial \theta'}$.

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), $g_v = [g_x \; g_u]$,

$$h_v = \begin{pmatrix} h_x & h_u \\ \underset{n_u \times n_x}{0} & \underset{n_u \times n_u}{0} \end{pmatrix} \text{ and } F = -(h_v' g_v' \otimes I_n)\mathtt{d}f_2 - (h_v' \otimes I_n)\mathtt{d}f_1 - (g_v' \otimes I_n)\mathtt{d}f_4 - \mathtt{d}f_3,$$

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 & Uribe (2012, Suppl. Mat., Sec. A.3) for the derivation of these results. Since we 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) with respect to θ requires the analytical derivatives of Q^{-1}, R, S^{-1} and T. See Appendix E 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 D 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.

Derivative of means. Differentiating the expressions for the means of z_t (8) and y_t (10) with respect to θ requires the analytical derivatives of the pruned state-space solution matrices. We obtain

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

Derivative of cumulants. Differentiating the cumulants for y_t of order k, requires $\mathtt{d}\Sigma_\xi$ (given in Appendix E), 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 we make use of generalized Sylvester equations as outlined in Appendix C. The analytical derivatives of the cumulants in equations (11) and (12) are then a simple application of Theorem 1.

Derivative of polyspectra. To obtain the derivative of $\mathcal{S}_{k,y}(\omega_1,\ldots,\omega_{k-1})$ w.r.t. θ in equations (13) to (15) analytically, we divide the interval $[-\pi;\pi]$ into N subintervals to obtain $N+1$ frequency indices with ω_s denoting the $s-$th frequency in the partition. The following steps can be conducted simultaneously using parallel computation: For each $\mathfrak{z}_j = e^{-i\omega_{s_j}} (j = 1,\ldots,k-1; s_j = 1,\ldots N+1, s_1 \leq \cdots \leq s_{k-1}, \text{imaginary } i)$ we first compute the derivative of $H_\xi(\mathfrak{z}_j)$ and its conjugate transpose, using the expression in Appendix E. The Jacobians $\mathtt{d}\mathcal{S}_{k,y}(\omega_1,\ldots,\omega_{k-1})$ then follow according to Theorem 1.

6. Identification criteria based on rank conditions

In the literature, three formal methods based on ranks have been proposed to check identification via (i) observational equivalent first and second moments (Iskrev, 2010), (ii) observational equivalent spectral densities (Qu & Tkachenko, 2012) and (iii) implications from control theory for observational equivalent minimal systems (Komunjer & Ng, 2011) for linearized DSGE models to the first-order. Since the pruned state-space (PSS) is a linear system with well-defined statistical properties, the same criteria can be checked, in particular for its mean, second-order moments and spectral density. In fact, we even extend Iskrev (2010)'s criteria for third- and fourth-order cumulants and Qu & Tkachenko (2012)'s criteria for the bi- and trispectrum of observables. But first, we state the underlying assumptions and definitions of local identifiability.

Assumption 1. Let Θ be the parameter space that yields the determinacy region of the DSGE model. Furthermore, assume y_t (t=1,...,T) is stationary to at least order four.

The first part of Assumption 1 is standard in the DSGE literature due to the rational expectation hypothesis, see Milani (2012) for a discussion. The second part needs some clarification. This concept 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. Teyssiere & 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.

We follow Deistler (1976) and define identifiability as a concept in moments, cumulants and polyspectra, not in probability laws.[17] We call $\theta_0 \in \Theta$ and $\theta_1 \in \Theta$ observationally equivalent (with respect to $\{y_t\}$), if they generate the same first four moments, cumulants or polyspectra of y_t.

Definition 1 (Identifiability from first four cumulants or moments). $\theta_0 \in \Theta$ is said to be locally identifiable from the first four moments of y_t, if there is an open neighborhood of θ_0 in which $\mu_y(\theta_0) = \mu_y(\theta_1)$, $m_2(\theta_0, T) = m_2(\theta_1, T)$, $m_3(\theta_0, T) = m_3(\theta_1, T)$ and $m_4(\theta_0, T) = m_4(\theta_1, T)$ 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 $\mu_y(\theta_0) = \mu_y(\theta_1)$, $\mathcal{S}_{2,y}(\omega_1; \theta_0) = \mathcal{S}_{2,y}(\omega_1; \theta_1)$, $\mathcal{S}_{3,y}(\omega_1, \omega_2; \theta_0) = \mathcal{S}_{3,y}(\omega_1, \omega_2; \theta_1)$ and $\mathcal{S}_{4,y}(\omega_1, \omega_2, \omega_3; \theta_0) = \mathcal{S}_{4,y}(\omega_1, \omega_2, \omega_3; \theta_1)$ for all $\omega_j \in [-\pi, \pi]$ (j=1,2,3) imply $\theta_0 = \theta_1$ for any $\theta_1 \in \Theta$.

The first definition corresponds to Iskrev (2010) in the time domain and the second definition to Qu & Tkachenko (2012) in the frequency domain. Apart from the different perspectives, Definition 1 states identifiability for a finite number of moment conditions, whereas Definition 2 corresponds to the complete set of dynamic properties. The criteria we derive all stem essentially from Rothenberg (1971), who proposes identifiability conditions based on injectivity of functions. The mappings we consider are the unconditional mean, unconditional cumulants (or moments) and corresponding polyspectra. Therefore, the fundamental idea is to determine, whether these mappings are unique for θ. 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 & 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.

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

Proposition 1 (Iskrev PSS)**.** *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

$$\overline{M}(q) := \frac{\partial \overline{m}(\theta_0, q)}{\partial \theta'} \tag{16}$$

has a full column rank equal to the number of parameters for $q \leq T$.

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

Remark 1. In other words, we extend Iskrev (2010)'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. 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, we can fix a subset of parameters. In our applications, we also check whether the parameters are identifiable through the mean, and second, third or fourth moments of observables, separately. We denote the corresponding matrices as $\overline{M}_2(q)$, $\overline{M}_3(q)$ and $\overline{M}_4(q)$. Iskrev (2010, 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, we stack all elements of the steady state, the solution matrices as well as all parameters of the stochastic innovations that depend on θ into a vector τ:

$$\tau(\theta) := \begin{pmatrix} \bar{y}' & c' & d' & vec(A)' & vec(B)' & vec(C)' & vec(D)' & \Gamma_{2,\xi}' & \Gamma_{3,\xi}' & \Gamma_{4,\xi}' \end{pmatrix}'$$

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

$$J := \frac{\partial \tau(\theta_0)}{\partial \theta'} \tag{17}$$

at θ_0 is equal to n_θ. This condition is, however, only necessary, because τ may be unidentifiable. Lastly, note that, given a first-order approximation and the Gaussian distribution for u_t, the proposition reduces to Iskrev (2010, Theorem 2), since all higher-order cumulants are zero in this case.

Proposition 2 (Qu & Tkachenko PSS)**.** *Assume that the power spectrum* (13)*, bispectrum* (14) *and trispectrum* (15) *are continuous in* $\omega \in [-\pi;\pi]$ *and continuous and differentiable in* $\theta \in \Theta$*. Let*

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

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 & Tkachenko (2012, Theorem 2) and Rothenberg (1971, Theorem 1).[18] □

Remark 2. Similar to Iskrev (2010)'s approach, Qu & Tkachenko (2012) focus on the dynamic structure of the DSGE model, but they work in the frequency domain. We 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, using the pruned state-space representation. Note that even when using analytical derivatives, we still have to divide the interval $[-\pi;\pi]$ into sufficient subintervals to numerically approximate the integrals. The dimension of $\overline{G}(\theta_0)$, however, is always $n_\theta \times n_\theta$. Moreover, in the applications, we check whether the parameters are identifiable through the mean and individual spectra. We denote the corresponding matrices as $\overline{G}_2(\theta_0)$,

[18] Note that we 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 $\mathcal{S}^R_{k,y}$ (see Brillinger (2001, p. 71) and Pintelon & Schoukens (2001, p. 553)). We have the following equivalence: $\mathcal{S}^*_{k,y}\mathcal{S}_{k,y} \Leftrightarrow \mathcal{S}^{R'}_{k,y}\mathcal{S}^R_{k,y}$. 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 we consider 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 $rank(\mathtt{d}\mathcal{S}^*_{k,y}\mathtt{d}\mathcal{S}_{k,y}) = n_\theta$.

$\overline{G}_3(\theta_0)$ and $\overline{G}_4(\theta_0)$. Qu & Tkachenko (2012) also provide several extensions, which also apply in our setting. In particular, we can check identification only from a subset of frequencies, conditional on other parameters being fixed, or we include general constraints on the parameters. Given a first-order approximation and the Gaussian distribution for u_t, this proposition reduces to Qu & Tkachenko (2012, Theorem 2), since the bi- and trispectrum are zero in this case.

7. 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. For all procedures, we are able to derive necessary as well as sufficient conditions for identification, based on ranks of Jacobians. For calculating the ranks, we 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 we use, on the one hand, a range from 1e-3 to 1e-17, 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.[19] In the case of rank deficiency, we are 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 (2010), is computationally very fast, we 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 & Iskrev (2011) and Qu & Tkachenko (2012, Corollary 4), a more robust method is to consider the power set and check the

[19] 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.

criteria for all possible subsets of parameters in a recursive fashion. In our experience, this *brute-force* approach yields more reliable results and is computationally just slightly slower, because, if we find a subset of parameters that are not identified, we can exclude that subset from higher-order subsets.[20]

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 one of speed: the higher T or N, the more time the calculations need.[21] With this line of thought, we can make use of the inherent symmetry in the computation of cumulants and spectra. That is, we only need to focus on the plane $t_1 \leq t_2 \leq t_3$ for Proposition 1 and $s_1 \leq s_2 \leq s_3$ (s_j=1,...,N+1; j=1,2,3) for Proposition 2, 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, we do not need to do the computations for all $N(N+1)/2$ runs, but rather for a much smaller set. Similarly, we 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 & 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 identification analysis is valid for a sufficient range of parameters. Therefore, in our applications, we 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, we have a quasi-global flavor of our 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

[20] We implemented both procedures for both criteria in the code.

[21] In most practical cases, T between 10 and 30 will be sufficient, since the higher the lag, the less informative the identification restrictions. Furthermore, we experienced with different values for N and find that the results for our 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, we recommend starting with $N = 10000$ for the power spectrum, $N = 1000$ for the bispectrum, $N = 100$ for the trispectrum and increase N if the results are unsatisfactory.

draws are identifiable, the model is still by no means globally identified.

8. Applications

8.1. The Kim (2003) model

This model extends the neoclassical growth model to include investment adjustment costs twofold. First, intertemporal adjustment costs, which involve a nonlinear substitution between capital and investment, are introduced into the capital accumulation equation governed by a parameter ϕ. Second, multisectoral costs, which are captured by a nonlinear transformation between consumption and investment, enter the budget constraint given a parameter θ. See Appendix A.1 for the model equations and setup. In the original paper, Kim (2003) log-linearizes the model and shows analytically that there is observational equivalence between these two specifications:

> [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, 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 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 k-th 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 moments and cumulants ($\overline{M}$) and corresponding polyspectra ($\overline{G}$) or we could use only individual statistics. Thus, including higher-order statistics may benefit identification and estimation, but it is not necessary for this model. We 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 & Iskrev (2011), we add a parameter *dumpy* into the analysis, which does not enter the model. As is evident in Figure 1, all crite-

Table 1: Identification analysis of the Kim (2003) model, 2nd-order approximation

	Iskrev					*Qu and Tkachenko*			
tol	J $T{=}30$	$\overline{M}_2$ $T{=}30$	$\overline{M}_3$ $T{=}30$	$\overline{M}_4$ $T{=}30$	$\overline{M}$ $T{=}30$	$\overline{G}_2$ $N{=}10000$	$\overline{G}_3$ $N{=}1000$	$\overline{G}_4$ $N{=}100$	$\overline{G}$ $N{=}100$
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)
1e-17	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)
Require	7	7	7	7	7	7	7	7	7

Ranks of identification tests given local point, different tolerance levels tol and analytical derivatives.
Numerical derivatives with differentiation step 1e-7 in parenthesis.

ria indicate that *dumpy* and (θ, ϕ) are not identifiable in a first-order approximation.[22] Given a second-order approximation and using the mean, autocovariogram and power spectrum of the PSS, the situation is different: now, in all cases, it is only dumpy that is not identifiable. We 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 we know – new to the literature.

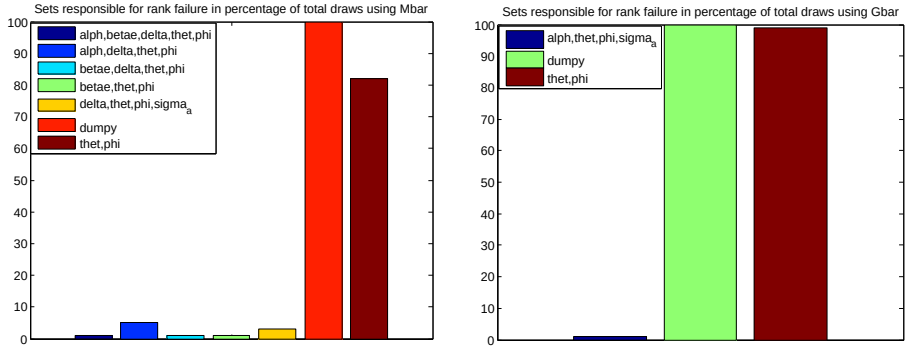
8.2. The An and Schorfheide (2007) model

This model is a prototypical DSGE model often cited in the literature concerning a lack of identification and nonlinear estimation. The authors show that (in the version we use in Appendix A.2) the set of parameters (ν, ϕ) and the steady state ratio $1/g = c/y$ do not enter the log-linearized solution. However, using a second-order approximation and the particle filter, they conclude that

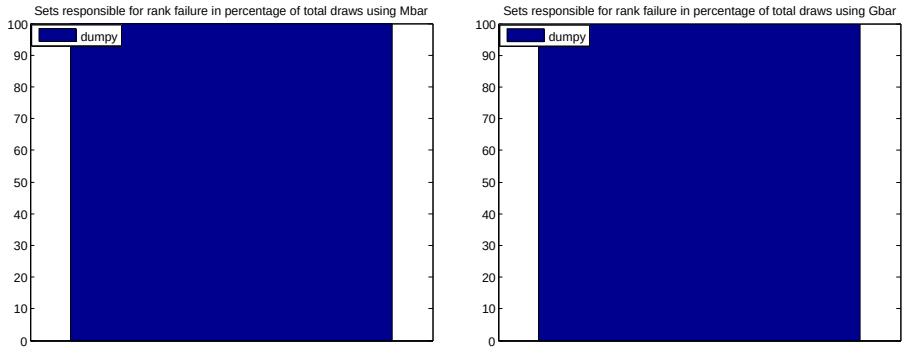
> the log-likelihood is slightly sloped in $1/g = c/y$ dimension. Moreover, (...) the quadratic likelihood (...) suggests that ν and ϕ are potentially separately identifiable (An & Schorfheide, 2007, p. 164).

[22] For a first-order approximation, we set all second-order terms in the pruned state-space to zero. Note that this is computationally not efficient, see Mutschler (2015b) for a better approach.

Figure 1: Sets responsible for nonidentification in the Kim (2003) model



(a) First-order approximation



(b) Second-order approximation

Identification results for $\overline{M}_2$ (left) and $\overline{G}_2$ (right) for 100 draws from the prior domain using analytical derivatives with robust tolerance level, $T = 30$ and $N = 10000$. Sets by brute-force method.

Furthermore, Komunjer & Ng (2011), Mutschler (2015b), Ratto & Iskrev (2011) and Qu & Tkachenko (2012) show that the coefficients entering the Taylor-rule $(\psi_1, \psi_2, \rho_R, \sigma_R)$ are not separately identifiable in the log-linearized model. However, An and Schorfheide argue that

> the 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 & Schorfheide, 2007, p. 164).

We confirm these alluring results by checking our rank criteria for a local point, as well

Table 2: Identification analysis of the An and Schorfheide (2007) model, 2^{nd}-order approximation

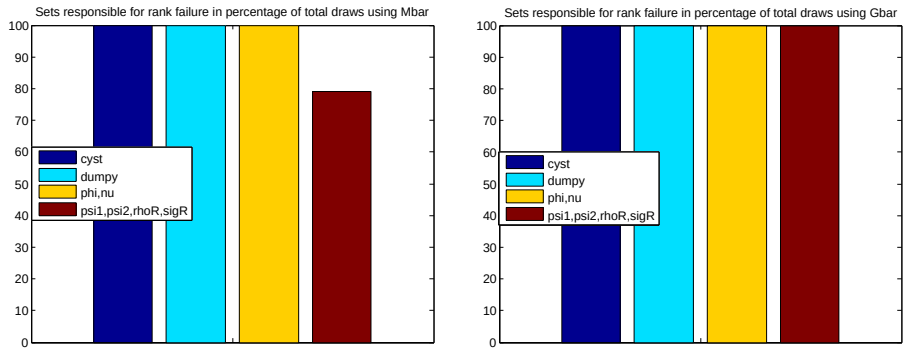
	Iskrev					*Qu and Tkachenko*			
tol	$J_{T=30}$	$\overline{M}_{2,T=30}$	$\overline{M}_{3,T=30}$	$\overline{M}_{4,T=30}$	$\overline{M}_{T=30}$	$\overline{G}_{2,N=10000}$	$\overline{G}_{3,N=1000}$	$\overline{G}_{4,N=100}$	$\overline{G}_{N=100}$
1e-03	15	12	15	15	15	7	7	6	7
1e-05	15	14	15	15	15	11	11	12	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
1e-17	15	15	15	15	15	15	15	15	15
robust	15	15	15	15	15	15	15	15	15
Require	15	15	15	15	15	15	15	15	15

Ranks of identification tests given local point, different tolerance levels tol and analytical derivatives.

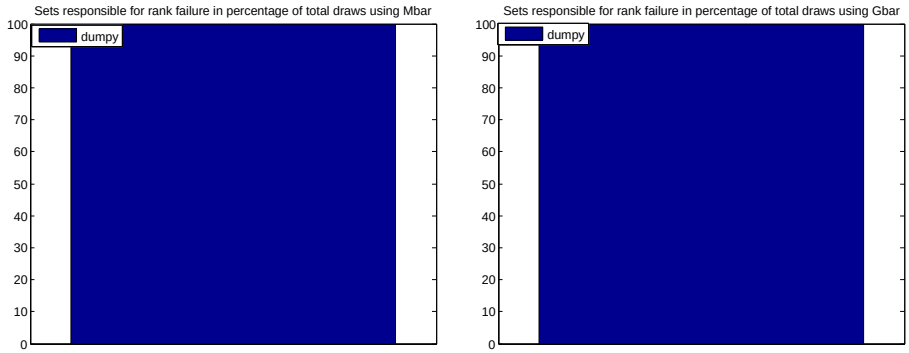
as for the prior domain. Table 2 shows that across criteria, a second-order approximation yields additional restrictions on moments and polyspectra, so as to identify all parameters of the model in the vicinity of the local point. This holds for each statistic individually, as well as for the complete set of dynamic properties. In other words, breaking with certainty equivalence, we obtain additional information, such that we are able to identify previously nonidentifiable 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 we 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 2 for a first-order approximation[23], the Taylor rule coefficients, (ϕ, ν) and c/y enter the problematic sets, whereas in the second-order approximation, in all cases, we are able to identify all parameters (apart from dumpy). In summary, we confirm An & Schorfheide (2007)'s approach to estimating the model using a second-order approximation and nonlinear estimation methods. 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

[23]See footnote 22.

Figure 2: Sets responsible for nonidentification in the An & Schorfheide (2007) model



(a) First-order approximation



(b) Second-order approximation

Identification results for $\overline{M}_2$ (left) and $\overline{G}_2$ (right) for 100 draws from the prior domain using analytical derivatives with robust tolerance level, $T = 30$ and $N = 10000$. Sets by brute-force method.

nonlinear model is – as far as we know – new to the literature.[24]

9. Conclusion

We establish formal rank criteria for a local identification of the deep parameters of a nonlinear DSGE model, using the pruned state-space system and higher-order statistics. Our procedures can be implemented prior to actually using nonlinear estimation methods. In this way, we demonstrate the identifiability of the Kim (2003) and the An & 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 & Schorfheide (2007) show that using a particle filter weakly enhances identifiability of the parameters of their model. Ivashchenko & Mutschler (2015) use the Central Difference and Quadratic Kalman filter on the pruned as well as unpruned version of the Kim (2003) model. They also find that the problematic parameters are separately estimable. Furthermore, the results of the present paper indicate that including higher-order statistics in a moment-matching approach or likelihood-type estimation may indeed improve the estimation of parameters.

Even though our 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. Our propositions and code can also be used for linear DSGE models with non-Gaussian innovations. A further extension would be

[24] Morris (2014) also shows that ν and ϕ are separately identifiable. As a robustness check for the Taylor rule coefficients, we compared the spectral density evaluated at θ_0 with the spectral densities evaluated at a hundred points from the nonidentification curve (fixing all parameters except the Taylor rule coefficients). Nonidentification curves are defined in Qu & Tkachenko (2012). If parameters are not identified, points on this curve yield the same spectral density at all frequencies apart from an approximation error; whereas if parameters are identified, the spectral densities differ. We found maximum relative and absolute deviations in the order 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. We also used the points reported in Table 1 of Qu & Tkachenko (2012) and found maximum relative and absolute deviations in the order of 10^{+4}. These findings provide further support for our result.

to establish rank criteria for other DSGE model specifications, as long as we are able to calculate moments or the spectrum of the data-generating process. For instance, Bianchi (2013) 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.

10. References

An, S., & Schorfheide, F. (2007). Bayesian analysis of DSGE models. *Econometric Reviews*, *26*, 113–172. doi:10.1080/07474930701220071.

Anderson, G. (2008). Solving linear rational expectations models: A horse race. *Computational Economics*, *31*, 95–113. doi:10.1007/s10614-007-9108-0.

Andreasen, M. M. (2011). Non-linear DSGE models and the optimized central difference particle filter. *Journal of Economic Dynamics & Control*, *35*, 1671–1695. doi:10.1016/j.jedc.2011.04.007.

Andreasen, M. M. (2013). Non-linear DSGE models and the central difference Kalman filter. *Journal of Applied Econometrics*, *28*, 929–955. doi:10.1002/jae.2282.

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

Bao, Y. (2013). On sample skewness and kurtosis. *Econometric Reviews*, *32*, 415–448. doi:10.1080/07474938.2012.690665.

Bianchi, F. (2013). *Methods for Measuring Expectations and Uncertainty in Markov-Switching Models*. Working Paper.

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

Blanchard, O. J., & Kahn, C. M. (1980). The solution of linear difference models under rational expectations. *Econometrica*, *48*, 1305–1311. doi:10.2307/1912186.

Brillinger, D. R. (1965). An introduction to polyspectra. *The Annals of Mathematical Statistics*, *36*, 1351–1374. doi:10.1214/aoms/1177699896.

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

Chandran, V., & Elgar, S. (1994). A general procedure for the derivation of principal domains of higher-order spectra. *IEEE Transactions on Signal Processing*, *42*, 229–233. doi:10.1109/78.258147.

Chib, S., & Ramamurthy, S. (2014). DSGE Models with Student-t Errors. *Econometric Reviews*, *33*, 152–171. doi:10.1080/07474938.2013.807152.

Curdia, V., del Negro, M., & Greenwald, D. L. (2014). Rare shocks, great recessions. *Journal of Applied Econometrics*, *29*, 1031–1052. doi:10.1002/jae.2395.

Dagenais, M. G., & Dagenais, D. L. (1997). Higher moment estimators for linear regression models with errors in the variables. *Journal of Econometrics*, *76*, 193–221. doi:10.1016/0304-4076(95)01789-5.

Deistler, M. (1976). The identifiability of linear econometric models with autocorrelated errors. *International Economic Review*, *17*, pp. 26–46. doi:10.1080/01969727408546081.

Deistler, M., & Seifert, H.-G. (1978). Identifiability and consistent estimability in econometric models. *Econometrica*, *46*, pp. 969–980. doi:10.2307/1909759.

Erickson, T., & Whited, T. M. (2002). Two-step GMM estimation of the errors-in-variables model using high-order moments. *Econometric Theory*, *3*, 776–799. doi:10.1017/S0266466602183101.

Fernández-Villaverde, J., & Rubio-Ramírez, J. F. (2007). Estimating Macroeconomic Models: A Likelihood Approach. *Review of Economic Studies*, *74*, 1059–1087. doi:10.1111/j.1467-937X.2007.00437.x.

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

Gomme, P., & Klein, P. (2011). Second-order approximation of dynamic models without the use of tensors. *Journal of Economic Dynamics and Control*, *35*, 604–615. doi:10.1016/j.jedc.2010.10.006.

Hannan, E. J. (1976). The Identification and Parameterization of Armax and State Space Forms. *Econometrica*, *44*, 713–723. doi:10.2307/1913438.

Herbst, E., & Schorfheide, F. (2014). Sequential Monte Carlo Sampling for DSGE Models. *Journal of Applied Econometrics*, *29*, 1073–1098. doi:10.1002/jae.2397.

Iskrev, N. (2008). Evaluating the information matrix in linearized DSGE models. *Economics Letters*, *99*, 607–610. doi:10.1016/j.econlet.2007.10.016.

Iskrev, N. (2010). Local identification in DSGE models. *Journal of Monetary Economics*, *57*, 189–202. doi:10.1016/j.jmoneco.2009.12.007.

Ivashchenko, S. (2014). DSGE model estimation on the basis of second-order approximation. *Computational Economics*, *43*, 71–82. doi:10.1007/s10614-013-9363-1.

Ivashchenko, S., & Mutschler, W. (2015). *Identifying DSGE models using nonlinear estimation methods - an example*. Working Paper Center for Quantitative Economics and St. Petersburg Institute for Economics and Mathematics.

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

Kan, R. (2008). From moments of sum to moments of product. *Journal of Multivariate Analysis*, *99*, 542 – 554. doi:10.1016/j.jmva.2007.01.013.

Kim, J. (2003). Functional equivalence between intertemporal and multisectoral investment adjustment costs. *Journal of Economic Dynamics and Control*, *27*, 533–549. doi:10.1016/S0165-1889(01)00060-4.

Kim, J., Kim, S., Schaumburg, E., & Sims, C. A. (2008). Calculating and using second-order accurate solutions of discrete time dynamic equilibrium models. *Journal of Economic Dynamics and Control*, *32*, 3397–3414. doi:10.1016/j.jedc.2008.02.003.

Klein, P. (2000). Using the generalized Schur form to solve a multivariate linear rational expectations model. *Journal of Economic Dynamics and Control*, *24*, 1405–1423. doi:10.1016/S0165-1889(99)00045-7.

Kollmann, R. (2015). Tractable latent state filtering for non-linear DSGE models using a second-order ap-

proximation and pruning. *Computational Economics*, *45*, 239–260. doi:10.1007/s10614-013-9418-3.

Komunjer, I., & Ng, S. (2011). Dynamic identification of dynamic stochastic general equilibrium models. *Econometrica*, *79*, 1995–2032. doi:10.3982/ECTA8916.

Koop, G., Pesaran, M. H., & Smith, R. P. (2013). On identification of Bayesian DSGE models. *Journal of Business and Economic Statistics*, *31*, 300–314. doi:10.1080/07350015.2013.773905.

Koopmans, T. C., & Reiersøl, O. (1950). The identification of structural characteristics. *Annals of Mathematical Statistics*, *21*, 165–181. doi:10.1214/aoms/1177729837.

Kotz, S., & Nadarajah, S. (2004). *Multivariate t-distributions and their applications*.

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

Lombardo, G., & Uhlig, H. (2014). *A theory of pruning*. ECB Working Paper Series Number 1696.

Magnus, J. R. (2010). On the concept of matrix derivative. *Journal of Multivariate Analysis*, *101*, 2200–2206. doi:10.1016/j.jmva.2010.05.005.

Magnus, J. R., & Neudecker, H. (1999). *Matrix Differential Calculus with Applications in Statistics and Econometrics*. (2nd ed.).

Meijer, E. (2005). Matrix algebra for higher order moments. *Linear Algebra and its Applications*, *410*, 112 – 134. doi:10.1016/j.laa.2005.02.040.

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*, 278–305. doi:10.1109/5.75086.

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.

Mutschler, W. (2015b). *Identification of linearized DSGE Models: A Review*. CQE Working Paper Center for Quantitative Economics.

Nikias, C., & Petropulu, A. (1993). *Higher-order Spectra Analysis: A Nonlinear Signal Processing Framework*. Prentice Hall signal processing series.

Pintelon, R., & Schoukens, J. (2001). *System identification : a frequency domain approach*.

Pollock, D. S. G. (2013). On Kronecker products, tensor products and matrix differential calculus. *International Journal of Computer Mathematics*, *90*, 2462–2476. doi:10.1080/00207160.2013.783696.

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

Qu, Z. (2014). Inference in dynamic stochastic general equilibrium models with possible weak identification. *Quantitative Economics*, *5*, 457–494. doi:10.3982/QE287.

Qu, Z., & Tkachenko, D. (2012). Identification and frequency domain quasi-maximum likelihood estimation of linearized dynamic stochastic general equilibrium models. *Quantitative Economics*, *3*, 95–132. doi:10.3982/QE126.

Ratto, M., & Iskrev, N. (2011). *Identification analysis of DSGE models with DYNARE*. MONFISPOL

225149.

Ratto, M., & Iskrev, N. (2012). *Estimation and inference in DSGE models using derivatives of the likelihood*. Working Paper.

Roth, M. (2013). *On the multivariate t distribution*. Technical Report.

Rothenberg, T. J. (1971). Identification in parametric models. *Econometrica*, *39*, 577–591. doi:10.2307/1913267.

Rusticelli, E., Ashley, R. A., Dagum, E. B., & Patterson, D. M. (2008). A new bispectral test for nonlinear serial dependence. *Econometric Reviews*, *28*, 279–293. doi:10.1080/07474930802388090.

Schmitt-Grohé, S., & Uribe, M. (2004). Solving dynamic general equilibrium models using a second-order approximation to the policy function. *Journal of Economic Dynamics and Control*, *28*, 755–775. doi:10.1016/S0165-1889(03)00043-5.

Schmitt-Grohé, S., & Uribe, M. (2012). What's news in business cycles. *Econometrica*, *80*. doi:10.3982/ECTA8050.

Swami, A., Giannakis, G., & Shamsunder, S. (1994). Multichannel ARMA processes. *IEEE Transactions on Signal Processing*, *42*, 898–913. doi:10.1109/78.285653.

Swami, A., & Mendel, J. (1990). Time and lag recursive computation of cumulants from a state-space model. *IEEE Transactions on Automatic Control*, *35*, 4–17. doi:10.1109/9.45138.

Teyssiere, G., & Kirman, A. P. (2011). *Long memory in economics*.

Appendix A. Example Models

Appendix A.1. The Kim (2003) model

Given auxiliary parameters $s = \frac{\beta\delta\alpha}{1-\beta+\delta\beta}$ and auxiliary variable $\lambda_t = \frac{(1-s)^\theta}{(1+\theta)c_t^{1+\theta}}$ the model is given by the following five equations f:

$$\lambda_t(1+\theta)\left(\frac{i_t}{s}\right)^\theta\left(\frac{i_t}{\delta k_t}\right)^\phi =$$

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

$$\left[(1-s)\left(\frac{c_t}{1-s}\right)^{1+\theta} + s\left(\frac{i_t}{s}\right)^{1+\theta}\right]^{\frac{1}{1+\theta}} = a_{t-1}k_{t-1}^\alpha, \quad log(a_t) = \rho_a log(a_{t-1}) + \varepsilon_{a,t},$$

$$k_t = \left[\delta\left(\frac{i_t}{\delta}\right)^{1-\phi} + (1-\delta)\left(k_{t-1}\right)^{1-\phi}\right]^{\frac{1}{1-\phi}}, \quad E_t\varepsilon_{a,t+1} = 0.$$

There are two exogenous states k_t and a_t, and no endogenous states. The controls are c_t and i_t and are both observable. There is one shock on technology $\varepsilon_{a,t}$ with

standard deviation σ_a, which we set equal to the perturbation parameter. Thus, given our definition and ordering of variables we have

$$u_t = \varepsilon_{a,t}, \qquad x_t = (k_{t-1}, a_{t-1})', \qquad y_t = (c_t, i_t)', \qquad \sigma = \sigma_A, \qquad \eta = (0 \;\; 1)'$$

The steady state of this model is given by

$$a = 1,\; k = \left(\frac{\delta}{sa}\right)^{\frac{1}{\alpha-1}},\; i = \delta k,\; c = (1-s)\left[\frac{(\alpha k^{\alpha})^{1+\theta} - s\left(\frac{i}{s}\right)^{1+\theta}}{1-s}\right]^{\frac{1}{1+\theta}},\; \varepsilon_a = 0.$$

We will consider identification of the parameter vector θ at the local point θ_0 and prior specification given in the following overview:

Parameters		*Prior specification*			*Bounds*		
θ	θ_0	Distr.	Par1	Par2	Range	Lower	Upper
α	0.60	GAMMA	0.60	0.30	$\mathbb{R}^+$	1e-5	1
β	0.99	UNIFORM	0.95	0.9999	$[a, b)$	0.9	0.99999
δ	0.0125	UNIFORM	0.01	0.02	$[a, b)$	0.01	0.02
θ	1	NORMAL	1.00	0.50	$\mathbb{R}$	-5	5
ρ_a	0.7	BETA	0.50	0.20	$[a, b)$	1e-5	0.99999
ϕ	2	NORMAL	2.00	0.50	$\mathbb{R}$	-5	5
σ_a	0.5	INVGAMMA	0.50	4.00	$\mathbb{R}^+$	1e-8	5

The code contains three model specifications: (1) The shock on technology is Gaussian. (2) We add Gaussian measurement errors to the measurement equations and extend the state vector for these additional stochastic innovations. (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.

Appendix A.2. The An and Schorfheide (2007) model

First we define auxiliary parameters $\beta = exp\left(-\frac{r^{(A)}}{400}\right)$, $\pi^* = exp\left(\frac{\pi^{(A)}}{400}\right)$ and $g^* = \frac{1}{(c/y)^*}$, then the model f consists of thirteen equations:

$$0 = \frac{1-\nu}{\nu\phi\pi^{*2}}\left(e^{\tau c_t}-1\right) - \left(e^{\pi_t}-1\right)\left[\left(1-\frac{1}{2\nu}\right)e^{\pi_t}+\frac{1}{2\nu}\right] + \beta\left(e^{E_t\pi_{t+1}}-1\right)e^{-\tau E_t c_{t+1}+\tau c_t+E_t dy_{t+1}+E_t\pi_{t+1}},$$

$$0 = 1 - e^{-\tau E_t c_{t+1}+\tau c_t+R_t-\rho_z z_t-E_t\pi_{t+1}}, \qquad 0 = e^{c_t-y_t} - e^{-g_t} + \frac{\phi\pi^{*2}g^*}{2}\left(e^{\pi_t}-1\right)^2,$$

$$0 = R_t - \rho_R R_{t-1} - (1-\rho_R)\psi_1\pi_t - (1-\rho_R)\psi_2\left(y_t - g_t\right) - \varepsilon_{R,t},$$

$$0 = dy_t - y_t + y_{t-1}, \qquad 0 = g_t - \rho_g g_{t-1} - \varepsilon_{g,t}, \qquad 0 = z_t - \rho_z z_{t-1} - \varepsilon_{z,t},$$

$$0 = YGR_t - \gamma^{(Q)} - 100(dy_t + z_t), \qquad 0 = INFL_t - \pi^{(A)} - 400\pi_t,$$

$$0 = INT_t - \pi^{(A)} - r^{(A)} - 4\gamma^{(Q)} - 400R_t,$$

$$0 = E_t\varepsilon_{R,t+1}, \qquad 0 = E_t\varepsilon_{g,t+1}, \qquad 0 = E_t\varepsilon_{z,t+1}.$$

There are three exogenous states R_t, g_t and z_t, and one endogenous state variable y_t.[25] The controls are c_t, dy_t and π_t, and the observables are $YGR_t, INFL_t$ and INT_t. There are three stochastic innovations: a monetary $\varepsilon_{R,t}$, a fiscal $\varepsilon_{g,t}$ and a technological shock $\varepsilon_{z,t}$. Furthermore we set the perturbation parameter equal to the standard deviation of the shock on technology. Thus, given our definition and ordering of variables we have

$$u_t = (\varepsilon_{R,t}, \varepsilon_{g,t}, \varepsilon_{z,t})', \quad x_t = (y_{t-1}, R_{t-1}, g_{t-1}, z_{t-1})',$$

$$y_t = (c_t, dy_t, \pi_t, YGR_t, INFL_t, INT_t)', \quad \sigma = \sigma_z$$

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

with observables equal to $S_{obs} \cdot y_t$. The steady state of this model is given by

$$y = R = g = z = \varepsilon = c = dy = \pi = 0,$$

$$YGR = \gamma^{(Q)}, \quad INFL = \pi^{(A)}, \quad INT = \pi^{(A)} + r^{(A)} + 4\gamma^{(Q)}.$$

We will consider identification of the parameter vector θ at the local point θ_0 and prior specification given in the following overview:

[25] Note that y_t is not the control vector y_t.

Parameters		*Prior specification*			*Bounds*		
θ	θ_0	Distr.	Par1	Par2	Range	Lower	Upper
τ	2.00	GAMMA	2.00	0.50	$\mathbb{R}^+$	1e-5	10
ϕ	50	GAMMA	50	20	$\mathbb{R}^+$	1e-5	100
ψ_1	1.50	GAMMA	1.50	0.25	$\mathbb{R}^+$	1e-5	10
ψ_2	0.125	GAMMA	0.50	0.25	$\mathbb{R}^+$	1e-5	10
ρ_R	0.75	BETA	0.50	0.20	$[a,b)$	1e-5	0.99999
ρ_g	0.95	BETA	0.80	0.10	$[a,b)$	1e-5	0.99999
ρ_z	0.90	BETA	0.66	0.15	$[a,b)$	1e-5	0.99999
$r^{(A)}$	1.00	GAMMA	0.80	0.50	$\mathbb{R}^+$	1e-5	10
$p^{(A)}$	3.20	GAMMA	4.00	2.00	$\mathbb{R}^+$	1e-5	20
γ_Q	0.55	NORMAL	0.40	0.20	$\mathbb{R}$	-5	5
σ_R	0.002	INVGAMMA	0.30	4.00	$\mathbb{R}^+$	1e-8	5
σ_g	0.006	INVGAMMA	0.40	4.00	$\mathbb{R}^+$	1e-8	5
σ_z	0.003	INVGAMMA	0.40	4.00	$\mathbb{R}^+$	1e-8	5
ν	0.10	BETA	0.10	0.05	$[a,b)$	1e-5	0.99999
c/y	0.85	BETA	0.85	0.10	$[a,b)$	1e-5	0.99999

The code contains three model specifications: (1) Structural shocks are Gaussian. (2) We add Gaussian measurement errors to the measurement equations and extend the state vector for these additional stochastic innovations. (3) The 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.

Appendix B. Product-moments of innovations

Given a second-order approximation[26], the innovations are defined as the $n_\xi \times 1$ vector

$$\xi_{t+1} = \left(u'_{t+1} \quad (u_{t+1} \otimes u_{t+1} - vec(\Sigma))' \quad (u_{t+1} \otimes x_t^f)' \quad (x_t^f \otimes u_{t+1})' \right)'$$

with $n_\xi = n_u + n_u^2 + 2n_x n_u$ elements. We are 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

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

n_ξ^4 elements, respectively. In order to compute these objects efficiently, we first reduce the dimension of ξ_t, since it has some duplicate elements. That is, we compute product-moments for the $n_{\tilde{\xi}} = n_u + n_u(n_u+1)/2 + n_u n_x$ vector

$$\tilde{\xi}_{t+1} := \begin{pmatrix} u_{t+1}' & vech(u_{t+1}u_{t+1}' - \Sigma)' & (u_{t+1} \otimes x_t^f)' \end{pmatrix}'$$

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 $K_{n_x,n_u}(u_{t+1} \otimes x_t^f) = (x_t^f \otimes u_{t+1})$. Then we have $M_{k,\xi} := [\otimes_{j=1}^k F_\xi] \cdot M_{k,\tilde{\xi}}$ 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, we 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 we have 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 & 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).[27] Note that 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 we 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.

[27] 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 x_t^f, which we can further reduce using indices from the `unique` function of MATLAB.

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

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

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 we get in step 3 contain terms of the general form

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

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

Gaussian distribution. In the case that u_t is normally distributed, x_t^f is also Gaussian with covariance matrix Σ_x given in equation (9). Therefore,

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

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, we 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 our example with $n_u = n_x = 1$ and Gaussian u_t, we get the unique entries

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

where $\Sigma_x = E(x_t^{f2}) = \sigma_x^2$. The cumulants can then be computed as outlined in Section 4. Since the third-order cumulant of a Gaussian process must be zero, we now see, that 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 $\frac{v}{v-2}\Sigma$, we rewrite u_t 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 & Nadarajah (2004, Ch. 1.7) and Roth (2013)). Since W and ε_t are independent, we have $E(u_t u_t') = E(W)E(\varepsilon_t\varepsilon_t') = \frac{v}{v-2}\Sigma$. 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, we can use Kan (2008)'s procedure and MATLAB function for the second product. Similar arguments apply to the product-moments of x_t^f, for instance the variance is given by

$$vec(\Sigma_x) = E[x_t^f \otimes x_t^f] = \underbrace{E[W]}_{\frac{v}{v-2}} \cdot (I_{n_x^2} - h_x \otimes h_x)^{-1}(h_u \otimes h_u) \cdot \underbrace{E[\varepsilon_t \otimes \varepsilon_t]}_{vec(\Sigma)}.$$

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, we 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 $\theta^{aux} = [vech(\Sigma)', vech(\Sigma_x)]'$, whereas in the Student's t case $\theta^{aux} = [vech(\Sigma)', vech(\Sigma_x), v]'$. Since $\frac{\partial \theta^{aux}}{\partial \theta'}$ is given by the model, we 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 C. Using 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, we 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 $\mathcal{C}_{k,z}$, we also use generalized Sylvester equa-

tions, an idea similar to Ratto & Iskrev (2012). That is, we take for each θ_i 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 we use as notation $dX = \frac{\partial X}{\partial \theta_i}$ to denote the derivative of X w.r.t. to a specific θ_i (i=1,...,n_θ) which has the same shape as X. All terms on the right hand side can be derived using the expressions and theorems of Section 5.

Appendix D. Example for notation and index matrices

When separating matrices and especially Jacobians into states and shocks, we 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 j-th row of h_v. Similar notation applies for $h^j_u, h^j_{ux_i}, h^j_{x_i u}$ and h^j_{uu}. The solution matrices for states are

given by

$$h_v = \begin{bmatrix} h^1_{x_1} & h^1_{x_2} & h^1_u \\ h^2_{x_1} & h^2_{x_2} & h^2_u \\ 0 & 0 & 0 \end{bmatrix}, \qquad h_{vv} = \begin{bmatrix} h^1_{x_1x_1} & h^1_{x_1x_2} & h^1_{x_1u} \\ h^1_{x_2x_1} & h^1_{x_2x_2} & h^1_{x_2u} \\ h^1_{ux_1} & h^1_{ux_2} & h^1_{uu} \\ h^2_{x_1x_1} & h^2_{x_1x_2} & h^2_{x_1u} \\ h^2_{x_2x_1} & h^2_{x_2x_2} & h^2_{x_2u} \\ h^2_{ux_1} & h^2_{ux_2} & h^2_{uu} \\ 0 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{bmatrix}.$$

In order to use notation of Andreasen et al (2014) we 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 matrices indicating the positions in h_{vv}:

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

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

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

Appendix E. Auxiliary solution matrices and auxiliary analytical derivatives

Let $n_v = n_x + n_u, n = n_v + n_y$ and

$$g_v = \begin{pmatrix} g_x & g_u \end{pmatrix}, \quad h_v = \begin{pmatrix} h_x & h_u \\ 0_{n_u \times n_x} & 0_{n_u \times n_u} \end{pmatrix}$$

then the auxiliary solution matrices are given by

$$Q = \left[h_v' \otimes f_2 \otimes h_v' + I_{n_v} \otimes f_4 \otimes I_{n_v} \quad I_{n_v} \otimes (f_1 \otimes I_{n_v} + f_2 g_v \otimes I_{n_v}) \right],$$
$$R = (I_n \otimes M')HM, \quad S = \left[f_1 + f_2 g_v \quad f_2 + f_4 \right],$$
$$U = f_2 trm\left((I_{n_y} \otimes (\tilde{\eta}\tilde{\eta}'))g_{vv}\right) + trm\left((I_n \otimes N')HN(\tilde{\eta}\tilde{\eta}')\right),$$
$$M = \begin{bmatrix} h_v \\ g_v h_v \\ I_{n_v} \\ g_v \end{bmatrix}, \quad N = \begin{bmatrix} I_{n_v} \\ g_v \\ 0_{n \times (n_v)} \end{bmatrix}, \quad \tilde{\eta} = \begin{pmatrix} 0_{n_x \times n_u} \\ \eta \end{pmatrix}$$

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 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, we obtain the derivatives $\mathtt{d}Q_1$ and $\mathtt{d}Q_2$. Now we can use Algorithm 1 to compute $\mathtt{d}Q$. However, we are interested in $\mathtt{d}Q^{-1}$, thus in step 2(b) we 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 & Neudecker, 1999, p. 184) and store it in step 2(c) in the i-th column of $\mathtt{d}(Q^{-1})$.

Derivative of R. Regarding the derivative of R we 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 we get $\mathtt{d}M$, whereas for the transpose we have the following relationship $\mathtt{d}M' = K_{2(n_x+n_y),n_x}\mathtt{d}M$. Now we are able to compute the derivative of R using Theorems 1 and 2.

Derivative of 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 we 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, we only have one partition, for which we know the derivative. Now taking the trm of the reshaped matrix in step 2(b) and storing this in step 2(c), we 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, we first have to derive an expression for $\mathtt{d}N$ and $\mathtt{d}N'$. Since N is partitioned, we can use Algorithm 1 to compute $\mathtt{d}N$ and $\mathtt{d}N' = K_{2n,n_v}\mathtt{d}N$.

Derivative of $H_\xi(\mathfrak{z})$. $H_\xi(\mathfrak{z})$ 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 Section 5 using Algorithm 1 for partitioned matrices. Thus, we 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 we used $\mathtt{d}(X^{-1}) = (-(X')^{-1} \otimes X^{-1})\mathtt{d}X$, see Magnus & 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.