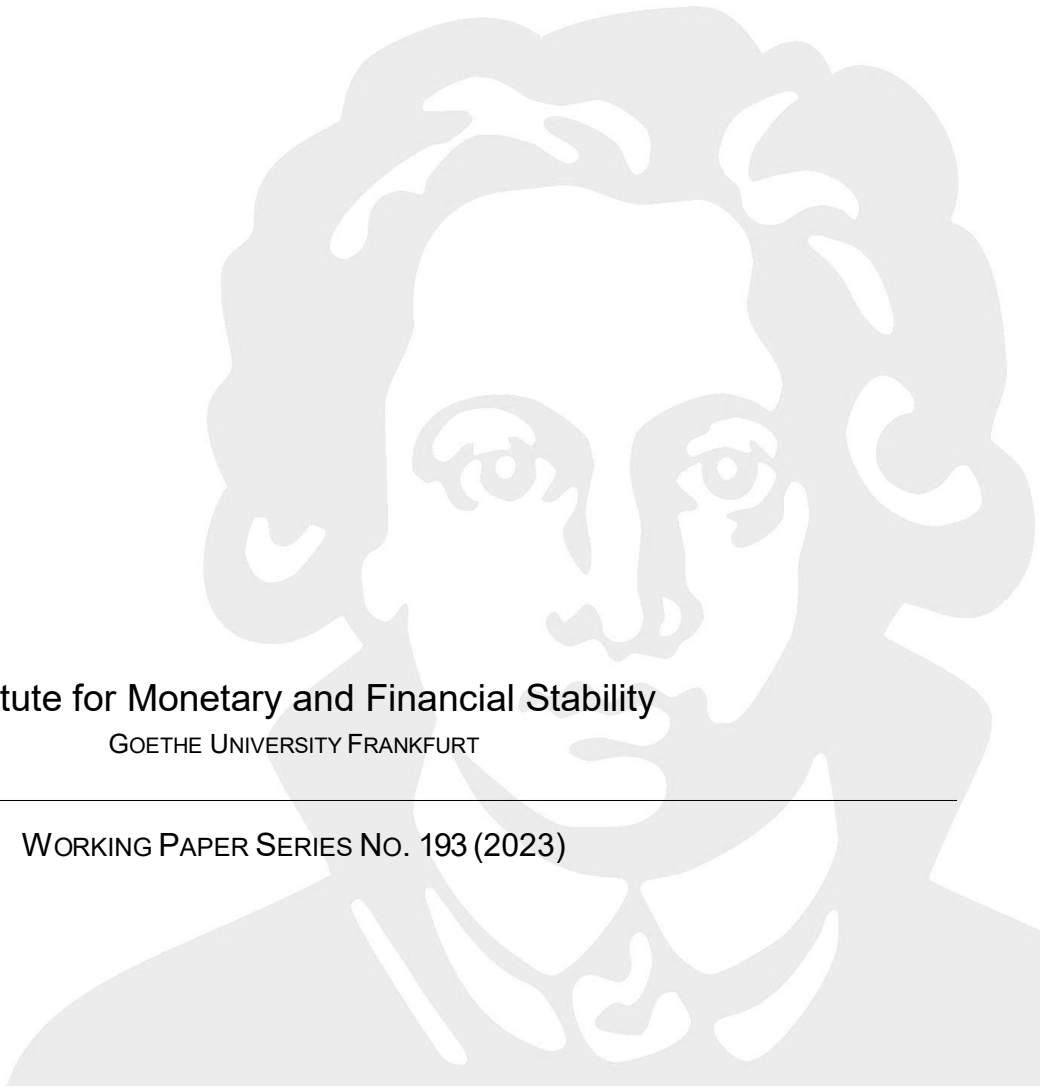
ALEXANDER MEYER-GOHDE

Numerical Stability Analysis of Linear DSGE Models – Backward Errors, Forward Errors and Condition Numbers



Institute for Monetary and Financial Stability
GOETHE UNIVERSITY FRANKFURT

WORKING PAPER SERIES NO. 193 (2023)

Institute for Monetary and Financial Stability
Goethe University Frankfurt
House of Finance
Theodor-W.-Adorno-Platz 3
D-60629 Frankfurt am Main
www.imfs-frankfurt.de | info@imfs-frankfurt.de

NUMERICAL STABILITY ANALYSIS OF LINEAR DSGE MODELS
BACKWARD ERRORS, FORWARD ERRORS AND CONDITION NUMBERS

ALEXANDER MEYER-GOHDE

ABSTRACT. This paper develops and implements a backward and forward error analysis of and condition numbers for the numerical stability of the solutions of linear dynamic stochastic general equilibrium (DSGE) models. Comparing seven different solution methods from the literature, I demonstrate an economically significant loss of accuracy specifically in standard, generalized Schur (or QZ) decomposition based solutions methods resulting from large backward errors in solving the associated matrix quadratic problem. This is illustrated in the monetary macro model of Smets and Wouters (2007) and two production-based asset pricing models, a simple model of external habits with a readily available symbolic solution and the model of Jermann (1998) that lacks such a symbolic solution - QZ-based numerical solutions miss the equity premium by up to several annualized percentage points for parameterizations that either match the chosen calibration targets or are nearby to the parameterization in the literature. While the numerical solution methods from the literature failed to give any indication of these potential errors, easily implementable backward-error metrics and condition numbers are shown to successfully warn of such potential inaccuracies. The analysis is then performed for a database of roughly 100 DSGE models from the literature and a large set of draws from the model of Smets and Wouters (2007). While economically relevant errors do not appear pervasive from these latter applications, accuracies that differ by several orders of magnitude persist.

JEL classification codes: C61, C63, E17

Keywords: Numerical accuracy; DSGE; Solution methods; Condition number; Backward error; Forward error

---

I am grateful to Michael Bauer, Frank Heinemann, Michel Juillard and Harald Uhlig as well as participants of the Schumpeter-BSE-Seminar, the Tübingen Macro Seminar, the CEF and Annual Conference of the IAAE for useful comments and suggestions and likewise to Raphael Abiry, Kyrill Fries, and Maximilian Thomin for invaluable research assistance. Any and all errors are entirely my own. This research was supported by the DFG through grant nr. 465469938 "Numerical diagnostics and improvements for the solution of linear dynamic macroeconomic models".

Goethe-Universität Frankfurt and Institute for Monetary and Financial Stability, Chair of Financial Markets and Macroeconomics, Theodor-W.-Adorno-Platz 3, 60629 Frankfurt am Main, Germany Email: meyer-gohde@econ.uni-frankfurt.de

Date: November 19, 2023.

1. Introduction

> *The machine could not adequately deal with ill conditioned equations, letting out a very sharp whistle when equilibrium could not be reached.*
> –attributed to Mary Croarken commenting on the Mallock simultaneous equation solver machine from 1931 (Higham and Hammarling, 2005)

Every user of numerical software has at least once (and likely much more frequently than that) encountered a warning that a matrix is nearly singular, badly scaled, or that a regression is nearly collinear - reminding us that these numerical problems are well-known to econometricians.[1] These warnings serve to inform the user that the limitations of finite-precision computing might have been reached and that the numerical results produced might be erroneous - the modern equivalent to the "very sharp whistle" from the epigraph. While users of numerical solution methods for linear macroeconomic models, specifically dynamic stochastic general equilibrium (DSGE) models, would be warned by the underlying software (say, Matlab) if a standard linear system of equations is numerically unstable, the solution of these linear DSGE models involves nonstandard equations, a matrix quadratic equation and linear equations that nest the solution of this quadratic. I demonstrate numerical instability in linear DSGE numerical solution methods from the literature, specially those that employ the generalized Schur or QZ decomposition, of economic consequence and engage in a backward-forward error analysis from the numerical mathematics literature to provide condition numbers and backward error bounds on the solutions and moments from these methods. That is, we lack the whistle for our methods and this paper seeks simultaneously to show that we need it and to provide it.

[1] Farrar and Glauber (1967, p. 99) noted more than half a century ago that "[t]he computer programmer's approach to singularity in regression analysis has begun to shape the econometrician's view of multicollinearity [.. and] the programmer, accordingly, is required to build checks for non-singularity into standard regression routines." See Pesaran (2015, Sec. 3.11) for highlights of the pernicious effects of multicollinearity on test statistics and the discussion of numerical versus structural causes of multicolinearity, what Spanos and McGuirk (2002, p. 365) state "constitutes one of the primary empirical modeling problems pertaining to the linear regression model." By comparing different numerical solution methods in the literature, this paper aims to analyze specifically numerical causes of ill-conditioning in linear DSGE models, but without prejudice towards potential structural causes - recalling (Farrar and Glauber, 1967, p. 99) observation that technical observations have led to theoretical developments - pointing to the Viner-Wong envelope theorem from now nearly 100 years ago.

I demonstrate the relevance of this whistle by examining two sets of experiments. The first set comprises two macro finance models, a simple habit formation model chosen as a symbolic solution is readily available and the influential model of Jermann (1998), and the policy relevant medium scale New Keynesian model of Smets and Wouters (2007). The first two are small scale production based asset pricing models that can be used to address perhaps the most prominent puzzle in the asset pricing literature, the equity premium puzzle (Mehra and Prescott, 1985; Mehra, 2003), that seeks to understand how risky assets can command such a high excess return in the face of moderate underlying volatility. While many convincing consumption based explanations that modify assumptions on, say, the stochastic properties of the pricing kernel have been offered, production based asset pricing face the additional challenge of needing to provide a structural cause of these stochastic properties. Providing a structural explanation invariably requires solving a structural model, the most common being dynamic stochastic general equilibrium (DSGE) models, which generally need to be solved numerically. Cochrane (2008, p. 300) expressed concern regarding the accuracy of solution approximations in general equilibrium and this paper points out a surprising degradation of the accuracy of solution approximations in the simplest approximation, linear approximations, and their consequences for the equity premium reported by these methods. Using both a highly stylized production-based asset pricing model and the model of Jermann (1998), I demonstrate the novel phenomenon that standard DSGE solution methods can produce numerical inaccuracies of economic significance, delivering an equity premium in certain extreme calibrations off by several annual percentage points. I also demonstrate that theoretically equivalent alternate ways of stating the equilibrium conditions lead to different numerical consequences. Ultimately, the inaccuracy in these models with log normal asst pricing stems from inaccuracies in the underlying macro variables whose risks are being priced. I then turn to the Smets and Wouters (2007) monetary macro model and demonstrate an analogous instability contained within the authors' prior. This leads to significant disagreement among the different solution methods from the literature concerning the moments of the core New Keynesian variables, inflation, output growth, and the nominal interest rate. In all three models for the parameterizations that led to obvious numerical instabilities (differences between solutions or moments from different methods in even all significant digits), non of the methods from the literature examined here produce a warning that the solution might be inaccurate or numerically instable. In all of these cases, however, the methods I

offer here on the conditioning and backward-forward errors did provide warnings and, even more so, provided warnings consistent with the degree of the error involved.

While examining the entirety of the literature involving linear(ized) DSGE models for the numerical stability of their solutions is obviously infeasible, the second set of experiments takes a step in this direction. First I utilize the Macroeconomic Model Data Base (MMB) (see Wieland, Cwik, Müller, Schmidt, and Wolters, 2012; Wieland, Afanasyeva, Kuete, and Yoo, 2016), a model comparison initiative at the Institute for Monetary and Financial Stability (IMFS),[2] to examine the condition numbers and backward error bounds of the solution methods from the literature for this set of around 100 models, ranging from small-scale pedagogical models to large-scale models from policy institutions. Second I analyze the condition numbers and backward error bounds of the solution methods for a sample of 100,000 draws from the posterior of the medium scale model of Smets and Wouters (2007). The analysis in both experiments confirms the results from the two macro-finance models and the Smets and Wouters (2007) model above that particularly those methods that use the generalized Schur or QZ decomposition are prone to inaccuracies, yet also that these inaccuracies are not economically significant in general. That is, I confirm that numerical instability is not an omnipresent problem in DSGE models, but like their linear system cousins, theoretical confined to singularities and practically to ranges around these in the parameter spaces.

The analysis here assess the numerical stability of solutions to linear DSGE models. Providing a solution to a DSGE model involves solving a functional equation to determine an unknown function that maps the sequences of variables in the information set into the endogenous variables of the model, resolving expectations of these same endogenous variables (Judd, 1998; Fernández-Villaverde, Rubio-Ramírez, and Schorfheide, 2016). Linear DSGE models and associated linear solutions have long been studied, e.g., Blanchard (1979) and Blanchard and Kahn (1980), and modern numerical packages such as Dynare (Adjemian, Bastani, Juillard, Mihoubi, Perendia, Ratto, and Villemot, 2011), Gensys (Sims, 2001), (Perturbation) AIM (Anderson and Moore, 1985; Anderson, Levin, and Swanson, 2006), Uhlig's Toolkit (Uhlig, 1999) and Solab (Klein, 2000) not only provide tools for solving a wide range of linear models, but also provide a first step in the solution procedure for many nonlinear methods as well. The substantial hurdle in these linear methods is finding a solution to a (matrix) quadratic equation, frequently required to be

[2]See http://www.macromodelbase.com.

the unique stable solution. For multivariate models with potentially singular coefficient matrices, the standard method is to double the dimension of the problem and employ the generalized Schur or QZ decomposition of Moler and Stewart (1973). While this algorithm is backward stable for the generalized eigenvalue decompositions for which it was designed, it is not always backward stable for quadratic eigenvalue problems (Tisseur, 2000) and may yield ill-conditioned eigenvalues for quadratic matrix polynomials (Higham, Mackey, and Tisseur, 2006; Higham, Mackey, Tisseur, and Garvey, 2008). I present the backward-forward error analysis of Higham and Kim (2001) for matrix quadratic equations and extend it to apply to the shock impact matrix and variance-covariance matrix of endogenous variables to provide an assessment of the accuracy of various solution methods in the literature valid when a symbolic solution is not available for comparison. Backward error diagnostics that can be calculated at minimal additional cost and in the absence of a symbolic or analytic solution successfully warn of potential inaccuracies. This is of immediate, practical use, as none of the algorithms from the literature I explore produced any warning that their solutions might suffer from economically significant losses of accuracy.

Apart from Anderson (2008), very little attention has been paid to comparing the accuracy of different algorithms for linear models[3] and to numerically addressing the assumptions necessary for the existence of a unique stable solution.[4] Improvements in the accuracy of the solution to linear DSGE models has implications for many nonlinear solutions as well. Anderson, Levin, and Swanson (2006) demonstrate that even small inaccuracies in lower orders compound to larger errors in the computation of higher, nonlinear solutions such as in Jin and Judd (2002). In terms of a backward-forward error analysis, Judd, Maliar, and Maliar (2017) comes the closest, yet their focus is the forward error of (non)linear solutions with regards to the underlying nonlinear model, taking again the accuracy of the linear solution for granted.

The remainder of the paper is structured as follows. Section 2 introduces the class of linear DSGE models and the solutions from the literature I will analyze. In section 3, I turn to the backward error and condition number analysis of these solutions, providing

[3]This is in stark contrast to the many studies that examine the accuracy of different nonlinear methods. See Fernández-Villaverde, Rubio-Ramírez, and Schorfheide (2016) for an overview.

[4]Heilberger, Klarl, and Maußner (2015) provides an exception, showing that, theoretically, if the rank assumption for the QZ decomposition is fulfilled for one ordering of the eigenvalues that conforms to the unit circle separation, it holds for any ordering that conforms to the same.

practical forward error bounds and comparing numerical considerations behind the calculation of and expected properties of these measures. Section 4 applies the analysis to two small macro-finance models and the Smets and Wouters (2007) model in detail, a large set of DSGE models in overview, and finally a set of draws from the posterior of the model of Smets and Wouters (2007). In section 5, I conclude.

2. Solving Linear DSGE Models

Standard numerical solution packages available to economists and policy makers—e.g., Dynare (Adjemian, Bastani, Juillard, Mihoubi, Perendia, Ratto, and Villemot, 2011), Gensys (Sims, 2001), (Perturbation) AIM (Anderson and Moore, 1985; Anderson, Levin, and Swanson, 2006), Uhlig's Toolkit (Uhlig, 1999) and Solab (Klein, 2000)—all analyze models that in some way or another can be expressed in the form of the nonlinear functional equation

$$0 = E_t[f(y_{t+1}, y_t, y_{t-1}, \varepsilon_t)] \tag{1}$$

The model equations (optimality conditions, resource constraints, market clearing conditions, etc.) are represented by the n_y-dimensional vector-valued function $f : \mathbb{R}^{n_y} \times \mathbb{R}^{n_y} \times \mathbb{R}^{n_y} \times \mathbb{R}^{n_e} \to \mathbb{R}^{n_y}$; $y_t \in \mathbb{R}^{n_y}$ is the vector of n_y endogenous variables; and $\varepsilon_t \in \mathbb{R}^{n_e}$ the vector of n_e exogenous shocks with a known distribution, where n_y and n_e are positive integers ($n_y, n_e \in \mathbb{N}$).

The solution to (1) is sought as the unknown function

$$y_t = y(y_{t-1}, \varepsilon_t), \quad y : \mathbb{R}^{n_y + n_e} \to \mathbb{R}^{n_y} \tag{2}$$

a function in the time domain that maps states, y_{t-1} and ε_t, into endogenous variables, y_t. An analytic form for (2) is rarely available and researchers and practitioners are compelled to find approximative solutions. However, a steady state, $\overline{y} \in \mathbb{R}^{n_y}$ a vector such $\overline{y} = y(\overline{y}, 0)$ and $0 = f(\overline{y}, \overline{y}, \overline{y}, 0)$ can frequently be recovered, either analytically or numerically, providing a point of expansion around which local solutions may be recovered.

A first-order, or linear, approximation of (1) at the steady state delivers

$$0 = AE_t[y_{t+1}] + By_t + Cy_{t-1} + D\varepsilon_t \tag{3}$$

where A, B, C, and D are the derivatives of f in (1) with respect to its arguments and, recycling notation, the y's in (3) refer to (log) deviations of the endogenous variables from their steady states, $\overline{y}$.

In analogy to (2), the standard approach to finding a solution to the linearized model (3) is to find a linear solution in the form

$$y_t = P\, y_{t-1} + Q\, \varepsilon_t \tag{4}$$

a recursive solution in the time domain–solutions that posit y_t as a function of its own past, y_{t-1}, and exogenous innovations, ε_t.

2.1. **Matrix Quadratic and Linear Impact Matrix Equations.** Inserting (4) into (3) and taking expectations ($E_t[\varepsilon_{t+1}] = 0$), yields the restrictions

$$0 = AP^2 + BP + C \tag{5}$$

$$0 = (AP + B)Q + D \tag{6}$$

or expressed jointly

$$AP\begin{bmatrix} P & Q \end{bmatrix} + B\begin{bmatrix} P & Q \end{bmatrix} + \begin{bmatrix} C & D \end{bmatrix} = \begin{bmatrix} 0 & 0 \end{bmatrix} \tag{7}$$

Generally, a unique P with eigenvalues inside the closed unit circle is sought (I will address this formally below). Lan and Meyer-Gohde (2014) prove the latter can be uniquely solved for Q if such a P can be found. While the theoretical hurdle is the former, matrix quadratic equation, the solution of the latter demonstrates clearly that numerical inaccuracies in P can percolate to further components of the solution (in this case Q).

To assist in the analysis, I will formalize the matrix quadratic equation in (5). For A, B, and $C \in \mathbb{R}^{n_y \times n_y}$, a matrix quadratic $M(P) : \mathbb{C}^{n_y \times n_y} \to \mathbb{C}^{n_y \times n_y}$ is defined as

$$M(P) \equiv A\,P^2 + B\,P + C \tag{8}$$

with its solutions, called solvents,[5] given by $P \in \mathbb{C}^{n_y \times n_y}$ if and only if $M(P) = 0$. The eigenvalues of the solvent, called latent roots of the associated lambda matrix[6] $M(\lambda) : \mathbb{C} \to \mathbb{C}^{n \times n}$ (here of degree two), are given via

$$M(\lambda) \equiv A\,\lambda^2 + B\,\lambda + C \tag{9}$$

The latent roots are (i) values of $\lambda \in \mathbb{C}$ such that $\det M(\lambda) = 0$ and (ii) $n_y - \operatorname{rank}(A)$ infinite roots. An explicit link between the quadratic matrix problem and the quadratic eigenvalue

[5]The analysis proceeds in the complex plane, but the results carry over when solutions are restricted to be real valued due to the eigenvalue separation about the unit circle assumed below, see also Klein (2000).

[6]See, e.g., Dennis, Jr., Traub, and Weber (1976, p. 835) or Gantmacher (1959, vol. I, p. 228).

problem is given via

$$\lambda \in \mathbb{C} : \left(A\lambda^2 + B\lambda + C\right)x = 0 \text{ for some } x \neq 0 \tag{10}$$

which has been reviewed extensively by Tisseur and Meerbergen (2001) and for which Hammarling, Munro, and Tisseur (2013) provide a comprehensive method to improve the accuracy of its solutions. If a unique stable solution is sought or required, this can be formulated via an adaptation of Blanchard and Kahn's (1980) rank and order conditions to the matrix quadratic formulation above. First assume there exist $2n_y$ latent roots of (9) of which n_y lie inside (or on) and n_y outside the unit circle. Second, there exists an $P \in \mathbb{R}^{n_y \times n_y}$ such that $M(P) = 0$ and $|eig(P)| \leq 1$.

Given P, Q follows from (6), solving the linear equation $\mathbb{C}^{n_y \times n_y} \to \mathbb{C}^{n_y \times n_y}$

$$0 = (AP + B)Q + D \tag{11}$$

Hence the solution of (log)linear(ized) DSGE models involves solving a matrix quadratic equation in P and, given this P, a linear system in Q.

2.2. **Linear DSGE Solution Methods.** Most linear DSGE methods (including Dynare (Adjemian, Bastani, Juillard, Mihoubi, Perendia, Ratto, and Villemot, 2011), Gensys (Sims, 2001), Uhlig's Toolkit (Uhlig, 1999) and Solab (Klein, 2000)) use a generalized Schur or QZ decomposition (Moler and Stewart, 1973; Golub and van Loan, 2013) of the companion linearization of (3)[7] in some form or another. For the formulation above, the matrix quadratic (5) can be brought into the QZ form as

$$F \begin{bmatrix} I_{n_y} \\ P \end{bmatrix} P = G \begin{bmatrix} I_{n_y} \\ P \end{bmatrix}, \quad F \equiv \begin{bmatrix} I_{n_y} & 0_{n_y \times n_y} \\ 0_{n_y \times n_y} & A \end{bmatrix}, G \equiv \begin{bmatrix} 0_{n_y \times n_y} & I_{n_y} \\ -C & -B \end{bmatrix} \tag{12}$$

where I_{n_y} is an $n_y \times n_y$ identity matrix and $0_{n_y \times n_y}$ is an $n_y \times n_y$ zero matrix.

Applying the QZ or generalized Schur decomposition (unitary Q and Z and upper triangular S and T with $Q^* F Z = S$ and $Q^* G Z = T$), Higham and Kim (Theorem 3 2000) prove that all solvents or solutions of (12) are of the form $P = Z_{21} Z_{11}^{-1} = Q_{11} S_{11}^{-1} T_{11} Q_{11}^{-1}$. The decomposition is intricately related to the quadratic eigenvalue problem (10) via

$$\lambda \in \mathbb{C} : (F\lambda - G)y, \text{ where } y = \begin{bmatrix} x' & x'\lambda \end{bmatrix} \text{ for some } x \neq 0 \tag{13}$$

$$\lambda \in \mathbb{C} : Q(S\lambda - T)\tilde{y}, \text{ where } \tilde{y} = Z^* \begin{bmatrix} x' & x'\lambda \end{bmatrix} \text{ for some } x \neq 0 \tag{14}$$

[7] Instead of the method of undetermined coefficients taken for expediency here, a multivariate pivoted Blanchard (1979) approach that delivers the solution constructively is presented in the appendix.

where the eigenvalues in both lines are identical following from unitary equivalence (Moler and Stewart, 1973) and hence identical to the eigenvalues in (10) and the latent roots of (9). From the upper triangularity of S and T it follows that the spectrum or set of eigenvalues of the pencil $P_{FG}(\lambda) = F\lambda - G$ is determined by the diagonal entries of S and T

$$\rho(P_{FG}) = \{t_{ii}/s_{ii},\ \text{if } s_{ii} \neq 0;\ \infty,\ \text{if } s_{ii} = 0;\ \emptyset,\ \text{if } s_{ii} = t_{ii} = 0;\ i = 1,\ldots,2n_y\} \tag{15}$$

where s_{ii} and t_{ii} denote the i'th row and i'th column of S and T respectively.

Ordering the decomposition so that the eigenvalues outside the unit circle are in the lower right blocks of S and T (hence S_{22} and T_{22}), the necessary and sufficient assumptions for a unique stable solution for y_t of (3) to exist are (1) Regularity: $P_{FG}(z)$ is called regular if there exists a $z \in \mathbb{C}$ such that $det(Fz - G) \neq 0$; (2) Order: Of the $2n_y$ generalized eigenvalues, there are exactly n_y stable roots inside (or on) the unit circle, and consequently, exactly n_y unstable roots outside the unit circle; (3) Rank: Z_{11}, the upper right block of Z, is nonsingular. If and only if these three assumptions are fulfilled does a unique solution P stable with respect to the closed unit circle exist. Consequentially, the overwhelming majority of the linear solution methods provided to researchers and practitioners in the standard numerical solution packages enumerated at the beginning of the section can be summarized by this single matrix decomposition. The specific numerical imp(Adjemian, Bastani, Juillard, Mihoubi, Perendia, Ratto, and Villemot, 2011; Villemot, 2011), Gensys (Sims, 2001), Uhlig's Toolkit (Uhlig, 1999) and Solab (Klein, 2000).

Binder and Pesaran (1997), the cyclic reduction method in Dynare (Adjemian, Bastani, Juillard, Mihoubi, Perendia, Ratto, and Villemot, 2011), and Anderson (2010) are three prominent methods that solve for P without appealing to the generalized Schur decomposition. Binder and Pesaran's (1997) "fully recursive method" works directly with the matrix quadratic (5) and iterates on

$$\tilde{P}_k = I_{n_y} - \tilde{A}\tilde{P}_{k-1}^{-1}\tilde{C},\ \text{ where } \tilde{A} \equiv B^{-1}A,\ \tilde{C} \equiv B^{-1}C,\ \tilde{P}_0 \equiv I_{n_y} \tag{16}$$

Delivering the solution to the matrix quadratic (5) as $P = -\tilde{P}_N^{-1}\tilde{C}$ for some maximum iteration N.[8] The cyclic reduction method implemented in Dynare (Adjemian, Bastani, Juillard, Mihoubi, Perendia, Ratto, and Villemot, 2011) operates on the following recursion

[8] In a related study, Binder and Meyer-Gohde (2023) examine recursive formulations that allow N to be determined endogenously according to a convergence criterion.

(see Bini, Latouche, and Meini, 2002)

$$P = -\hat{A}_i^{-1} A_0 \tag{17}$$

where

$$\hat{A}_i = \hat{A}_{i-1} - A_{2,i-1} A_{1,i-1}^{-1} A_{0,i-1} \tag{18}$$

$$A_{1,i} = A_{1,i-1} - A_{0,i-1} A_{1,i-1}^{-1} A_{2,i-1} - A_{2,i-1} A_{1,i-1}^{-1} A_{0,i-1} \tag{19}$$

$$A_{0,i} = -A_{0,i-1} A_{1,i-1}^{-1} A_{0,i-1} \tag{20}$$

$$A_{2,i} = -A_{2,i-1} A_{1,i-1}^{-1} A_{2,i-1} \tag{21}$$

with initial conditions $\hat{A}_0 = B$, $A_{2,0} = A$, $A_{1,0} = B$, and $A_{0,0} = C$ until convergence of $\hat{A}_i$.[9] Anderson (2010) applies the bi-orthogonality from the separation of stable and unstable solutions to solve for the left invariant space associated with unstable solutions via[10]

$$\begin{bmatrix} y_t \\ E_t[y_{t+1}] \end{bmatrix} = \begin{bmatrix} 0_{n_y \times n_y} & I_{n_y} \\ -A^{-1}C & -A^{-1}B \end{bmatrix} \begin{bmatrix} y_{t-1} \\ y_t \end{bmatrix} \Rightarrow \begin{bmatrix} V_1 & V_2 \end{bmatrix} \begin{bmatrix} 0_{n_y \times n_y} & I_{n_y} \\ -A^{-1}C & -A^{-1}B \end{bmatrix} = \mathscr{M} \begin{bmatrix} V_1 & V_2 \end{bmatrix} \tag{22}$$

where the vectors of V span the invariant space associated with unstable eigenvalues. This gives $y_t = -V_2^{-1} V_1 y_{t-1}$ as the solution to the homogenous problem, i.e., the matrix quadratic (5), $P = -V_2^{-1} V_1$. Essentially, by rearranging or shuffling the equations and variables, Anderson (2010) is able to reformulate a potentially singular system requiring the generalized Schur decomposition into a nonsingular system that can be solved using standard eigenvalue methods. The key commonality of these three methods is that they avoid the QZ or generalized Schur decompositon.

3. Backward-Forward Error Analysis of Linear DSGE Model Solutions

I turn now to the backward-forward analysis to provide measures for the accuracy/numerical stability of solutions to linear DSGE models. I begin by introducing the concepts of backward error and condition numbers for linear systems and how they relate to forward errors, then I turn to backward error bounds and condition numbers for the solutions of linear DSGE models. Finally, I provide ex posterior measures of the

[9] Huber, Meyer-Gohde, and Saecker (2023) examine alternative formulations and relate the cyclic reduction method to structure preserving doubling methods.

[10] This assumes that A is invertible, the general case can be found in Anderson (2010) and is merely slightly more involved, utilizing the shuffle-algorithm of Luenberger (1978) to yield an invertible A.

forward error, providing easily implementable bounds on the numerical errors of solutions provided by users' preferred method.

To fix ideas, I begin with a linear system. This provides a link to the concept of a condition number that is familiar (or at least the warning thereof provided by the numerical software being used) to practitioners and the nonlinear matrix measures necessary for the analysis of solutions to linear DSGE models. To this end, given a linear system

$$Ax = b, \quad A \in \mathscr{R}^{n\times n}, \ x \text{ and } b \in \mathscr{R}^n \tag{23}$$

I would like to know, how "good" a solution $\hat{x}$ provided numerically is. Ideally, I would like measure how far the approximate solution $\hat{x}$ is from the true solution x, the forward error, defined as

$$\|x - \hat{x}\| / \|x\| \tag{24}$$

for some norm. Apparently, being able to assess how good our numerical solution is requires me to know the true solution is in the first place. Fortunately, the forward error can be bounded and is approximately equal to the product of two quantities I can readily calculate numerically, what Higham (2002, p. 9) calls a "useful rule of thumb" and holds exactly for linear equation systems as derived by Turing (1948, p. 298)

$$\text{forward error} \lesssim \text{condition number} \times \text{backward error} \tag{25}$$

illustrating that the error in the approximate solution (forward error) can be determined through both the condition number and the backward error.

The backward error of an approximate solution $\hat{x}$ gives a measure of how much the problem (here, A and b) at a minimum needs to be changed in order for the approximate solution to be the exact solution. That is, how close to the original problem is the problem actually solved by $\hat{x}$. In terms of normwise deviations,[11] this is

$$\eta(\hat{x}) = \min\left\{\epsilon : (A + \Delta A)\hat{x} = b + \Delta B, \|\Delta A\|_F \leq \alpha\epsilon, \|\Delta b\|_F \leq \beta\epsilon\right\} \tag{26}$$

Choosing $\alpha = \|A\|_F$ and $\beta = \|b\|_F$ gives $\eta(\hat{x})$ as the normwise relative backward error. Defining the residual $r \equiv b - A\hat{x}$, the constraint in the foregoing can be used to bound the

[11] The results for the linear system here hold for consistent norms, I choose this presentation to be consistent with the analysis of the quadratic problem. This combines Higham and Kim (2000), Higham (1993), Kågström (1994) and Higham (2002, Ch.7&16).

backward error from below as

$$\|r\|_F = \|\Delta A\hat{x} - \Delta B\|_F \leq \|\Delta A\|_F \|\hat{x}\|_F + \|\Delta B\|_F \leq \left(\alpha \|\hat{x}\|_F + \beta\right)\eta(\hat{x}) \tag{27}$$

where the last equality uses the optimal perturbations from (26). Expressing this in terms of the relative residual, $rr(\hat{x}) \equiv \|r\|_F / \left(\alpha \|\hat{x}\|_F + \beta\right)$ gives

$$rr(\hat{x}) \leq \eta(\hat{x}) \tag{28}$$

or that the backward error is bounded below by the relative residual. This highlights the importance of the backward error, as it states that a small backward error necessarily implies a small relative residual. That is, if the nearest problem exactly solved by $\hat{x}$ is close to the original problem (small backward error), then the residual induced by solving the original problem with $\hat{x}$ will be small.

Of at least equal importance is the reverse implication: whether a small residual necessarily implies a small backward error - particularly in the context of residual based accuracy checks from the literature (see section 3.6). To establish this, rewrite the constraint from (26) using the Kronecker product rule $\text{vec}(ABC) = \left(C' \otimes A\right)\text{vec}(B)$ as

$$r = \Delta A\hat{x} - \Delta B \tag{29}$$

$$= \left(\hat{x}' \otimes I_n\right)\text{vec}(\Delta A) - I_n \Delta b \tag{30}$$

$$= \begin{bmatrix} \alpha\left(\hat{x}' \otimes I_n\right) & -\beta I_n \end{bmatrix} \begin{bmatrix} \alpha^{-1}\text{vec}(\Delta A) \\ \beta^{-1}\Delta b \end{bmatrix} \tag{31}$$

$$= Hz \tag{32}$$

$'$ indicates transposition and vec columnwise vectorization. This is an underdetermined system in z as H is a matrix of size $n \times n(n+1)$. The minimum 2-norm solution[12] is

$$z = H^+ r \tag{33}$$

where $H^+ = H^*(HH^*)^{-1}$, * indicates conjugate transposition and H is assumed to be of full row rank. Using the properties of the 2 and Frobeneus norms[13]

$$\|z\|_2 = \left\| \begin{bmatrix} \alpha^{-1}\Delta A & \beta^{-1}\Delta b \end{bmatrix} \right\|_F = \left(\alpha^{-2}\|\Delta A\|_F^2 + \beta^{-2}\|\Delta b\|_F^2\right)^{1/2} \tag{34}$$

[12] See the appendix.

[13] See the appendix.

and using the definition of $\eta(\hat{x})$ from (26, which is the minimum value of the larger of $\alpha\|\Delta A\|_F$ and $\beta\|\Delta b\|_F$,[14] gives

$$\frac{1}{\sqrt{2}}\|z\|_2 \leq \eta(\hat{x}) \leq \|z\|_2 \tag{35}$$

focusing on the upper bound

$$\eta(\hat{x}) \leq \|z\|_2 = \left\|H^+ r\right\|_2 \leq \left\|H^+\right\|_2 \|r\|_2 = \|r\|_F/\sigma_{min}(H) \tag{36}$$

where σ_{min} is the smallest singular value of its argument or the smallest eigenvalue, λ_{min}, of its argument and conjugation,

$$\begin{aligned}
\sigma_{min}(H) &= \lambda_{min}\left(HH^*\right)^{1/2} && (37)\\
&= \lambda_{min}\left(HH^*\right)^{1/2} && (38)\\
&= \lambda_{min}\left(\begin{bmatrix}\alpha\left(\hat{x}'\otimes I_n\right) & -\beta I_n\end{bmatrix}\begin{bmatrix}\alpha\left(\bar{\hat{x}}\otimes I_n\right) & -\beta I_n\end{bmatrix}\right)^{1/2} && (39)\\
&= \lambda_{min}\left(\alpha^2\hat{x}'\bar{\hat{x}}I_n + \beta^2 I_n\right)^{1/2} && (40)\\
&= \lambda_{min}\left(\left(\alpha^2\hat{x}^*\hat{x} + \beta^2\right)I_n\right)^{1/2} && (41)\\
&= \left(\alpha^2\|\hat{x}\|_F^2 + \beta^2\right)^{1/2} && (42)
\end{aligned}$$

Combining (28), (36), and (42) gives

$$rr(\hat{x}) \leq \eta(\hat{x}) \leq \frac{\alpha\|\hat{x}\|_F + \beta}{\left(\alpha^2\|\hat{x}\|_F^2 + \beta^2\right)^{1/2}} rr(\hat{x}) \leq \sqrt{2}rr(\hat{x}) \tag{43}$$

and hence the backward error can be bounded up to $\sqrt{2}$ to the relative residual.[15] This tells us that a small backward error implies a small relative residual and vice-versa. If the condition number of the problem is also small, then a small backward error implies a small forward error and we can conclude that small relative residuals, small backward errors, and small forward errors are synonymous.

Hence, the condition number of the problem needs to be established. To do this, consider the following perturbation of (23)

$$(A + \Delta A)(x + \Delta x) = b + \Delta b \tag{44}$$

[14] See Higham (2002, p. 310).

[15] Actually for this problem, the lower bound can be achieved, see Higham (2002, p. 120) and Rigal and Gaches (1967). For the purposes here and in line with exposition for the DSGE model that follows, providing bounds suffices to make the argument.

where perturbations will be measured normwise (congruently to the backward errors above so that choosing $\alpha = \|A\|_F$ and $\beta = \|b\|_F$ gives normwise relative perturbations) by

$$\epsilon = \max\left\{\alpha^{-1}\|\Delta A\|_F, \beta^{-1}\|\Delta b\|_F\right\} \tag{45}$$

Expanding (44) gives

$$Ax + A\Delta x + \Delta A x + \Delta A \Delta x = b + \Delta b \tag{46}$$

Noting that $\Delta A \Delta x$ is of the order $\mathcal{O}(\epsilon^2)$ and $AX = b$ gives

$$A\Delta x = -\Delta A x + \Delta b + \mathcal{O}(\epsilon^2) \tag{47}$$

using the Kronecker product rule $\operatorname{vec}(ABC) = \left(C' \otimes A\right)\operatorname{vec}(B)$ this can be written as

$$A\Delta x = -\left(x' \otimes I_n\right)\operatorname{vec}(\Delta A) + I_n \Delta b + \mathcal{O}(\epsilon^2) \quad = \begin{bmatrix} -\alpha\left(x' \otimes I_n\right) & \beta I_n \end{bmatrix} \begin{bmatrix} \alpha^{-1}\operatorname{vec}(\Delta A) \\ \beta^{-1}\Delta b \end{bmatrix} + \mathcal{O}(\epsilon^2) \tag{48}$$

and so

$$\Delta x = A^{-1}\begin{bmatrix} -\alpha\left(x' \otimes I_n\right) & \beta I_n \end{bmatrix} \begin{bmatrix} \alpha^{-1}\operatorname{vec}(\Delta A) \\ \beta^{-1}\Delta b \end{bmatrix} + \mathcal{O}(\epsilon^2) \tag{49}$$

$$\|\Delta x\|_F = \left\|A^{-1}\begin{bmatrix} -\alpha\left(x' \otimes I_n\right) & \beta I_n \end{bmatrix}\right\|_2 \left\|\begin{bmatrix} \alpha^{-1}\operatorname{vec}(\Delta A) \\ \beta^{-1}\Delta b \end{bmatrix}\right\|_2 + \mathcal{O}(\epsilon^2) \tag{50}$$

as $\left\|\begin{bmatrix} \alpha^{-1}\operatorname{vec}(\Delta A) \\ \beta^{-1}\Delta b \end{bmatrix}\right\|_2 = \left\|\begin{bmatrix} \alpha^{-1}\Delta A & \beta^{-1}\Delta b \end{bmatrix}\right\|_F \leq \sqrt{2}\epsilon$, the foregoing can be written as

$$\frac{\|\Delta x\|_F}{\|x\|_F} \leq \sqrt{2}\Psi\epsilon \tag{51}$$

where

$$\Psi \equiv \left\|A^{-1}\begin{bmatrix} -\alpha\left(x' \otimes I_n\right) & \beta I_n \end{bmatrix}\right\|_2 / \|x\|_F \tag{52}$$

is the condition number giving the bound above on the forward error, $\frac{\|\Delta x\|_F}{\|x\|_F}$, sharp to first order in ϵ. Note that the above conforms to Higham's (2002, p. 9) "useful rule of thumb"

$$\underbrace{\frac{\|\Delta x\|_F}{\|x\|_F}}_{\text{forward error}} \lesssim \underbrace{\Psi}_{\text{condition number}} \times \underbrace{\epsilon}_{\text{backward error}} \tag{53}$$

where the meaning of $\lesssim$ becomes clear: sharp to first order in ϵ and up to a factor $\sqrt{2}$.

This bound can be weakened to

$$\frac{\|\Delta x\|_F}{\|x\|_F} \leq \sqrt{2}\Psi\epsilon \leq \sqrt{2}\Phi\epsilon \tag{54}$$

where

$$\Psi \leq \left\|A^{-1}\right\|_2 \left\| \begin{bmatrix} -\alpha\left(x' \otimes I_n\right) & \beta I_n \end{bmatrix} \right\|_2 / \|x\|_F \tag{55}$$

$$= \left\|A^{-1}\right\|_2 \sigma_{max}\left(\begin{bmatrix} -\alpha\left(x' \otimes I_n\right) & \beta I_n \end{bmatrix} \right) / \|x\|_F \tag{56}$$

$$= \left\|A^{-1}\right\|_2 \left(\alpha^2 \|x\|_F^2 + \beta^2\right)^{1/2} / \|x\|_F \tag{57}$$

$$\leq \left\|A^{-1}\right\|_2 \left(\alpha \|x\|_F + \beta\right) / \|x\|_F \equiv \Phi \tag{58}$$

Finally, this can be further weakened, as $\|x\|_F \geq \|b\|_F / \|A\|_F$, to yield

$$\Phi = \left\|A^{-1}\right\|_2 \left(\alpha \|x\|_F + \beta\right) / \|x\|_F \tag{59}$$

$$\leq \left\|A^{-1}\right\|_2 \|A\|_F \left(\alpha \|x\|_F / \|b\|_F + \beta / \|b\|_F\right) \tag{60}$$

$$\leq \left\|A^{-1}\right\|_2 \|A\|_F \left(\alpha / \|A\|_F + \beta / \|b\|_F\right) \tag{61}$$

$$\leq \left\|A^{-1}\right\|_F \|A\|_F \left(\alpha / \|A\|_F + \beta / \|b\|_F\right) \tag{62}$$

$\left\|A^{-1}\right\|_F \|A\|_F$ is the condition number of A, $\kappa(A)$, and choosing $\alpha = \|A\|_F$ and $\beta = \|b\|_F$ gives normwise relative perturbations in A and b.

I now turn to the matrix quadratic and matrix impact equations, (5) and (6), that the solution methods for linear DSGE models from the previous section solve. Higham and Kim (2001) provide bounds on the backward error and a condition number for the solvent, P, of a quadratic matrix equation - I extend their analysis to normwise relative perturbations and a posteriori forward error bounds consistent with Higham (1993) and Kågström (1994). The latter is particularly useful for practitioners as it provides easy to calculate metrics od the conditioning and forward errors of the solution provided by the existing DSGE literature - the "very sharp whistle" from the epigraph for solutions of DSGE models. While the matrix impact equation is a system of linear equations and simply a matrix-matrix extension of the introduction to the methods above,[16] the coefficients in this equation depend on the solution to the matrix quadratic and the explicit consideration of this demonstrates that inaccuracies in the matrix quadratic contaminate the accuracy of solutions for impact coefficients as observed by Anderson, Levin, and Swanson (2006).

I will proceed as follows. I will address the backward errors, the condition numbers and finally a posteriori or practical forward error bounds, beginning first with the matrix quadratic $0 = AP^2 + BP + C$ and then the impact matrix Q that solves $(AP + B)Q + D$, firstly

[16] See, e.g., Higham and Higham (1998).

conditioning on an infinite precision solution for P and then successively considering the effects of numerical inaccuracies in P on Q. These results on solutions for Q, the impact matrix of shocks, from (6),

$$0 = (AP + B)Q + D \tag{63}$$

proceeds first by formulating this equation to conform to standard linear numerical analyses

$$FQ = -D \tag{64}$$

where $F \equiv AP + B$ - perturbations in F will not be equivalent to the perturbations in A and B above and arbitrary perturbations in P ignore the analysis above. To demonstrate this, I will consider three approaches to the conditioning of the linear system in Q: (1) The perturbed equation $FQ = -D$

$$(F + \Delta F)(Q + \Delta Q) = -D - \Delta D \tag{65}$$

the perturbed version of (63) with arbitrary perturbations in P

$$0 = [(A + \Delta A)(P + \Delta P) + B + \Delta B](Q + \Delta Q) + D + \Delta D \tag{66}$$

and finally, the perturbed version of (63) with perturbations in P as result from solving (5) under finite precision

$$0 = \{[A + \Delta A][P(A + \Delta A, B + \Delta B, C + \Delta C)] + B + \Delta B\}(Q + \Delta Q) + D + \Delta D \tag{67}$$

Finally, I will consider solving for P and Q jointly instead of successively, particularly as calculating the backward errors for Q from (67) is a nonlinear problem that either must be linearized or solved via optimization, whereas the joint problem

$$(A + \Delta A)(P + \Delta P)\begin{bmatrix} P + \Delta P & Q + \Delta Q \end{bmatrix} \tag{68}$$

$$+ (B + \Delta B)\begin{bmatrix} P + \Delta P & Q + \Delta Q \end{bmatrix} \tag{69}$$

$$+ \begin{bmatrix} C + \Delta C & D + \Delta D \end{bmatrix} = \begin{bmatrix} 0 & 0 \end{bmatrix} \tag{70}$$

yields problems for backward errors, conditioning numbers, and a posteriori forward errors that take the interdependence into account and use the methods from the matric quadratic problem in a straightforward manner.

3.1. **Backward Error Analysis of Linear DSGE Model Solutions.** I begin with bounds on the backward errors of the matrix quadratic $0 = AP^2 + BP + C$ from (5). Defining the backward error of an approximate solvent P in terms of the relative normwise deviations in the matrix of coefficients,

$$\eta_P(\hat{P}) = \min\Big\{\epsilon : (A + \Delta A)\hat{P}^2 + (B + \Delta B)\hat{P} + C + \Delta C = 0,$$
$$\|\Delta A\|_F \leq \epsilon\alpha, \|\Delta B\|_F \leq \epsilon\beta, \|\Delta C\|_F \leq \epsilon\gamma\Big\}$$

this error can be bounded from above and below as follows

Theorem 1 (Bounds on the Backward Error of P)

The backward error is bounded by

$$RR(\hat{P}) \leq \eta_P(\hat{P}) \leq \left\| \begin{bmatrix} \alpha\hat{P}^{2\prime} \otimes I_{n_y} & \beta\hat{P}' \otimes I_{n_y} & \gamma I_{n_{y^2}} \end{bmatrix}^+ \mathrm{vec}(R) \right\|_2 \leq \mu_P(\hat{P}) RR(\hat{P})$$

where $RR(\hat{P})$ is the relative residual

$$RR(\hat{P}) = \frac{\|R\|_F}{\alpha\left\|\hat{P}^2\right\|_F + \beta\left\|\hat{P}\right\|_F + \gamma}$$

and $\mu_P(\hat{P})$ is a factor given by

$$1 \leq \mu_P(\hat{P}) = \frac{\alpha\left\|\hat{P}^2\right\|_F + \beta\left\|\hat{P}\right\|_F + \gamma}{\left(\alpha^2\sigma^2_{min}(\hat{P}^2) + \beta\sigma^2_{min}(\hat{P}) + \gamma^2\right)^{1/2}}$$

where σ_{min} is the smallest non-zero singular value.

Proof. See the appendix here. □

Higham and Kim (2001) point out that their backward error analysis demonstrates that a small relative residual (the absolute residual being $AP^2 + BP + C$ for an approximate P returned by a numerical algorithm) does not necessarily imply a small backward error for the matrix quadratic problem. The former follows from the relative residual being bounded above by the backward error and the latter, the converse, is hampered by the presence of $\mu_P(\hat{P})$, a "growth" (Kågström, 1994) or "amplification" (Higham, 1993) factor that measures by how much the backward error can exceed the relative residual. Examining $\mu_P(\hat{P})$ and following Higham (1993) and Ghavimi and Laub (1995), it follows that this factor can be arbitrarily larger than one, particularly when $\left\|\hat{P}\right\|_F >> \sigma_{min}(\hat{P})$ - i.e., when $\sigma_{max}(\hat{P}) >> \sigma_{min}(\hat{P}) \Rightarrow \left\|\hat{P}\right\| \left\|\hat{P}^+\right\| >> 1$ and, hence, $\hat{P}$ is an ill-conditioned solution to the matrix quadratic. The intermediate upper bound takes the structure of the system into account and, e.g., see Ghavimi and Laub (1995), can be substantially

lower than the larger upper bound if large elements in the residual counteract (near) zero entries in $\left[\alpha\hat{P}^{2\prime}\otimes I_{n_y} \quad \beta\hat{P}^{\prime}\otimes I_{n_y} \quad \gamma I_{n_{y^2}}\right]^{+}$.

Turning to the shock impact matrix Q and beginning with the formulation in (64), the equation $FQ=-D$ with $F\equiv AP+B$ where the backward error is defined in terms of the relative normwise deviations in the matrix of coefficients of the linear equation it solves

$$\eta_{Q_1}(\hat{Q})=\min\left\{\epsilon:(F+\Delta F)\hat{Q}=-D-\Delta D,\quad \|\Delta F\|_F\leq\epsilon\phi,\quad \|\Delta D\|_F\leq\epsilon\delta\right\}$$

This error can be bounded from above and below as follows

Theorem 2 (Bounds on the Backward Error of Q via (64))
The backward error is bounded by

$$RR_{Q_1}(\hat{Q})\leq\eta_{Q_1}(\hat{Q})\leq\left\|\left[\phi\hat{Q}^{\prime}\otimes I_{n_e} \quad \delta I_{n_y\cdot n_e}\right]^{+}\mathrm{vec}(R)\right\|_2\leq\mu_{Q_1}(\hat{Q})RR_{Q_1}(\hat{Q})$$

where $RR_{Q_1}(\hat{Q})$ is the relative residual

$$RR_{Q_1}(\hat{Q})=\frac{\phi\|R\|_F}{\phi\|\hat{Q}\|_F+\delta}$$

and μ_{Q_1} is given by

$$1\leq\mu_{Q_1}=\frac{\phi\|\hat{Q}\|_F+\delta}{\left(\phi^2\sigma^2_{min}(\hat{Q})+\delta^2\right)^{1/2}}$$

where σ_{min} is the smallest singular value.

Proof. See the appendix here. □

Again we have a "growth" (Kågström, 1994) or "amplification" (Higham, 1993) factor that measures by how much the backward error can exceed the relative residual. In contrast to the vector case $Ax=b$ used to introduce the concepts above, we see that the backward error can be larger than the relative error when $\|\hat{Q}\|_F >> \sigma_{min}(\hat{Q})$. Recalling that $\|\hat{Q}\|_F=\left(\sum_i\sigma_i(Q)^2\right)^{1/2}$, this corresponds to (Higham and Higham, 1992b) generalization that the sensitive of linear systems with multiple right-hand sides corresponds approximately to the worst-case sensitivity of the individual systems.

Taking the specific coefficient errors into account, I now present results on the conditioning and backward errors of solutions for Q, the impact matrix of shocks, from (63) where the backward error is defined in terms of the relative normwise deviations in the

matrix of coefficients A, B, P, and D,

$$\eta_{Q_2}(\hat{P},\hat{Q}) = \min\Big\{\epsilon : \big((A+\Delta A)\hat{P}+B+\Delta B\big)\hat{Q} = -D-\Delta D, \\ \|\Delta A\|_F \le \epsilon\alpha, \quad \|\Delta B\|_F \le \epsilon\beta, \quad \|\Delta D\|_F \le \epsilon\delta\Big\}$$

this error can be bounded from above and below as follows

Theorem 3 (Bounds on the Backward Error of Q via (63))

The backward error is bounded by

$$RR_{Q_2}(\hat{P},\hat{Q}) \le \eta_{Q_2}(\hat{P},\hat{Q}) \le \left\| \begin{bmatrix} \alpha(\hat{P}\hat{Q})' \otimes I_{n_y} & \beta\hat{Q}' \otimes I_{n_y} & \delta I_{n_y \cdot n_e} \end{bmatrix}^+ \mathrm{vec}(R) \right\|_2 \le \mu_{Q_2}(\hat{P},\hat{Q}) RR_{Q_2}(\hat{P},\hat{Q})$$

where $RR_{Q_2}(\hat{P},\hat{Q})$ is the relative residual

$$RR_{Q_2}(\hat{P},\hat{Q}) = \frac{\|R\|_F}{\alpha\left\|\hat{P}\hat{Q}\right\|_F + \beta\left\|\hat{Q}\right\|_F + \delta}$$

and $\mu_P(\hat{P})$ is given by

$$1 \le \mu_{Q_2}(\hat{P},\hat{Q}) = \frac{\alpha\left\|\hat{P}\hat{Q}\right\|_F + \beta\left\|\hat{Q}\right\|_F + \delta}{\left(\alpha^2\sigma^2_{min}(\hat{P}\hat{Q}) + \beta^2\sigma^2_{min}(\hat{Q}) + \delta^2\right)^{1/2}}$$

where σ_{min} is the smallest singular value.

Proof. See the appendix here. □

Now, the "growth" (Kågström, 1994) or "amplification" (Higham, 1993) factor that measures by how much the backward error can exceed the relative residual can be arbitrarily large depending not only on the relation of $\left\|\hat{Q}\right\|_F$ to $\sigma_{min}(\hat{Q})$, but also $\left\|\hat{P}\hat{Q}\right\|_F$ to $\sigma_{min}(\hat{P}\hat{Q})$ - that is, there is a potential for the transmission of errors in the solution of P to the solution of Q.

Taking errors in the solution of P from (5) specifically into account, that is, that P is a function of A, B, and C, would result in a nonlinear optimization problem

$$\eta_{Q_3}(\hat{Q}) = \min\Big\{\epsilon : \big((A+\Delta A)\hat{P}+B+\Delta B\big)\hat{Q} = -D-\Delta D, \\ \hat{P} : (A+\Delta A)\hat{P}^2 + (B+\Delta B)\hat{P} + C + \Delta C = 0 \\ \|\Delta A\|_F \le \epsilon\alpha, \quad \|\Delta B\|_F \le \epsilon\beta, \quad \|\Delta C\|_F \le \epsilon\gamma, \quad \|\Delta D\|_F \le \epsilon\delta\Big\}$$

While the problem might be solved numerically Higham and Higham (1992a), linearity can be restored by considering P and Q jointly as in the following, which is more than sufficient as it then captures exactly this dependance of Q on P into account.

For approximate solutions to $\hat{P}$ and $\hat{Q}$ the backward error can be defined via

$$\eta_{PQ}(\hat{P},\hat{Q}) = \min\Big\{\epsilon : (A+\Delta A)\hat{P}\begin{bmatrix}\hat{P} & \hat{Q}\end{bmatrix} + (B+\Delta B)\begin{bmatrix}\hat{P} & \hat{Q}\end{bmatrix} + \begin{bmatrix}C+\Delta C & D+\Delta D\end{bmatrix} = 0,$$
$$\|\Delta A\|_F \le \epsilon\alpha, \quad \|\Delta B\|_F \le \epsilon\beta, \quad \|\Delta C\|_F \le \epsilon\gamma, \quad \|\Delta D\|_F \le \epsilon\delta\Big\}$$

this error can be bounded from above and below as follows

Theorem 4 (Bounds on the Joint Backward Error of P and Q)

The backward error is bounded by

$$RR_{PQ}(\hat{P},\hat{Q}) \le \eta_{PQ}(\hat{P},\hat{Q})$$

$$\eta_{PQ}(\hat{P},\hat{Q}) \le \left\| \begin{bmatrix} \alpha \begin{bmatrix} (\hat{P}^2)' \\ (\hat{P}\hat{Q})' \end{bmatrix} \otimes I_n & \beta \begin{bmatrix} \hat{P}' \\ \hat{Q}' \end{bmatrix} \otimes I_n & \gamma \begin{bmatrix} I_{n_y} \\ 0_{n_e \times n_y} \end{bmatrix} \otimes I_{n_y} & \delta \begin{bmatrix} 0_{n_y \times n_e} \\ I_{n_e} \end{bmatrix} \otimes I_{n_y} \end{bmatrix}^{+} \begin{bmatrix} \mathrm{vec}(R_P) \\ \mathrm{vec}(R_Q) \end{bmatrix} \right\|_2$$
$$\le \mu_{PQ}(\hat{P},\hat{Q}) RR_{PQ}(\hat{P},\hat{Q})$$

where $RR_{PQ}(\hat{P},\hat{Q})$ is the relative residual

$$RR_{PQ}(\hat{P},\hat{Q}) = \frac{\left\| \begin{bmatrix} R_P & R_Q \end{bmatrix} \right\|_F}{\alpha\left(\|\hat{P}^2\|_F + \|\hat{P}\hat{Q}\|_F\right) + \beta\left(\|\hat{P}\|_F + \|\hat{Q}\|_F\right) + \left(\gamma^2+\delta^2\right)^{1/2}}$$

and $\mu_{PQ}(\hat{P},\hat{Q})$ is given by

$$\mu_{PQ}(\hat{P},\hat{Q}) = \frac{\alpha\left(\|\hat{P}^2\|_F + \|\hat{P}\hat{Q}\|_F\right) + \beta\left(\|\hat{P}\|_F + \|\hat{Q}\|_F\right) + \left(\gamma^2+\delta^2\right)^{1/2}}{\left(\alpha^2\left[\sigma^2_{min}(\hat{P}^2) + \sigma^2_{min}(\hat{P}\hat{Q})\right] + \beta^2\left[\sigma^2_{min}(\hat{P}) + \sigma^2_{min}(\hat{Q})\right] + \gamma^2 + \delta^2\right)^{1/2}}$$

where σ_{min} is the smallest singular value.

Proof. See the appendix here. □

Again we have a "growth" (Kågström, 1994) or "amplification" (Higham, 1993) factor that measures by how much the backward error can exceed the relative residual. Examining $\mu_{PQ}(\hat{P},\hat{Q})$ and following the analysis above, it follows that this factor can be arbitrarily larger than one, particularly if $\|\hat{P}\|_F >> \sigma_{min}(\hat{P})$, $\|\hat{Q}\|_F$ to $\sigma_{min}(\hat{Q})$, or $\|\hat{P}\hat{Q}\|_F$ to $\sigma_{min}(\hat{P}\hat{Q})$ - that is, this measures encapsulates all the measures from above. Having a single metric is, of course, a double edged sword, as it by itself us unable to pinpoint the source P, Q, or PQ. The different backward errors bounds are juxtaposed in table 1 and the both the analogue between the different measures as well as the joint measure being encompassing are readily apparent.

This joint backward error can be bounded by the individual relative residuals as follows

	Lower Bound	Upper Bound	
		Sharp	Weak
P	$RR_P(\hat{P})$	$\left\| \begin{bmatrix} \alpha \hat{P}^{2\prime} \otimes I_{n_y} & \beta \hat{P}' \otimes I_{n_y} & \gamma I_{n_y{}^2} \end{bmatrix}^+ \mathrm{vec}(R_P) \right\|_2$	$\frac{\alpha \\|\hat{P}^2\\|_F + \beta \\|\hat{P}\\|_F + \gamma}{\left(\alpha^2 \sigma^2_{min}(\hat{P}^2) + \beta \sigma^2_{min}(\hat{P}) + \gamma^2\right)^{1/2}} RR_P(\hat{P})$
Q_1	$RR_{Q_1}(\hat{Q})$	$\left\| \begin{bmatrix} \phi \hat{Q}' \otimes I_{n_e} & \delta I_{n_y \cdot n_e} \end{bmatrix}^+ \mathrm{vec}(R_Q) \right\|_2$	$\frac{\phi \\|\hat{Q}\\|_F + \delta}{\left(\phi^2 \sigma^2_{min}(\hat{Q}) + \delta^2\right)^{1/2}} RR_{Q_1}(\hat{Q})$
Q_2	$RR_{Q_2}(\hat{Q})$	$\left\| \begin{bmatrix} \alpha (\hat{P}\hat{Q})' \otimes I_{n_y} & \beta \hat{Q}' \otimes I_{n_y} & \delta I_{n_y \cdot n_e} \end{bmatrix}^+ \mathrm{vec}(R_Q) \right\|_2$	$\frac{\alpha \\|\hat{P}\hat{Q}\\|_F + \beta \\|\hat{Q}\\|_F + \delta}{\left(\alpha^2 \sigma^2_{min}(\hat{P}\hat{Q}) + \beta^2 \sigma^2_{min}(\hat{Q}) + \delta^2\right)^{1/2}} RR_{Q_2}(\hat{Q})$
P, Q	$RR_{PQ}(\hat{P}, \hat{Q})$	$\left\| X^+ \begin{bmatrix} \mathrm{vec}(R_P) \\ \mathrm{vec}(R_Q) \end{bmatrix} \right\|_2$	$\frac{\alpha(\\|\hat{P}^2\\|_F + \\|\hat{P}\hat{Q}\\|_F) + \beta(\\|\hat{P}\\|_F + \\|\hat{Q}\\|_F) + (\gamma^2 + \delta^2)^{1/2}}{\left(\alpha^2 \left[\sigma^2_{min}(\hat{P}^2) + \sigma^2_{min}(\hat{P}\hat{Q})\right] + \beta^2 \left[\sigma^2_{min}(\hat{P}) + \sigma^2_{min}(\hat{Q})\right] + \gamma^2 + \delta^2\right)^{1/2}} RR_{PQ}(\hat{P}, \hat{Q}))$

TABLE 1. Backward Error Bounds for Linear DSGE Model Solutions

- R_P and R_Q are the residuals and RR_P, and RR_Q, RR_{PQ} are the relative residuals of (5) and (6) respectively
- $\sigma_{min}(\cdot)$ is the smallest singular value
- $X = \begin{bmatrix} \alpha \begin{bmatrix} P^{2\prime} \\ Q'P' \end{bmatrix} \otimes I_{n_y} & \beta \begin{bmatrix} P' \\ Q' \end{bmatrix} \otimes I_{n_y} & \gamma \begin{bmatrix} I_{n_y} \\ 0_{n_e \times n_y} \end{bmatrix} \otimes I_{n_y} & \delta \begin{bmatrix} 0_{n_y \times n_e} \\ I_{n_e} \end{bmatrix} \otimes I_{n_y} \end{bmatrix}$

Corollary 1 (Individual and Joint Backward Error of P and Q)
The joint backward error $\eta_{PQ}(\hat{P},\hat{Q})$ can be bounded by the individual relative residuals of P and Q by

$$\max\left\{RR_P(\hat{P}),RR_{Q_2}(\hat{Q})\right\} \leq \eta_{PQ}(\hat{P},\hat{Q}) \leq \sqrt{2}\max\left\{\mu_P(\hat{P})RR_P(\hat{P}),\mu_{Q_2}(\hat{P},\hat{Q})RR_{Q_2}(\hat{Q})\right\}$$

Proof. See the appendix here. □

Hence the joint backward error is at least as large as the larger individual relative residual and can only be bounded above (up to a factor of $\sqrt{2}$) by the larger of the two individual upper bounds.

Having bounded the backward errors of the calculations of the solutions for linear DSGE models, I now turn to their condition numbers.

3.2. **Condition Numbers for Linear DSGE Model Solutions.** I now turn to the condition numbers for the solution of linear DSGE models. The condition number, as in the case of the linear model $Ax = b$, measures the sensitivity of the solution, x, with respect to the data, A and b. For linear DSGE models, we have not only the matrix form of the solution (or multiple right-hand sides) as a consideration, but also the nonlinearity in both the matrix quadratic for the transition matrix P as well as in the impact matrix Q through its dependence on P. Just as the condition number of A plays a pivotal role for the conditioning of the linear problem $Ax = b$, so too will the homogenous matrix coefficients in the two sets of equations that need to be solved for P and Q. To take the structures of the resulting equations into account, I begin by laying out the separation between two matrix pencils, before using this to derive the conditions numbers for P and Q.

For normwise perturbations in the parameter matrices

$$\epsilon = \max\left\{\frac{\|\Delta A\|_F}{\alpha},\frac{\|\Delta B\|_F}{\beta},\frac{\|\Delta C\|_F}{\gamma}\right\}$$

normwise relative perturbations in P, $\frac{\|\Delta P\|_F}{\|P\|_F}$, from the perturbed matrix quadratic equation

$$(A+\Delta A)(P+\Delta P)^2+(B+\Delta B)(P+\Delta P)+C+\Delta C = 0$$

can be bounded to first order in ϵ as follows

Theorem 5 (Condition Number of P)

The relative perturbation in P is bounded to first order in ϵ by

$$\frac{\|\Delta P\|_F}{\|P\|_F} \leq \sqrt{3}\Psi_P(P)\epsilon + \mathcal{O}(\epsilon^2)$$

where $\Psi_P(P)$, the condition number of P, is given by

$$\Psi_P(P) = \left\| V^{-1} \left[\alpha \left(P^2\right)' \otimes I_{n_y} \quad \beta P' \otimes I_{n_y} \quad \gamma I_{n_{y^2}} \right] \right\|_2 / \|P\|_F$$

where

$$V = I_{n_y} \otimes (AP + B) + P' \otimes A$$

Proof. See the appendix here and Higham and Kim (2001), noting the different measurement of perturbations. □

This sharp bound can be weakened to

Corollary 2 (Bound of Condition Number of P)

The condition number of P, $\Psi_P(P)$, can be bounded by

$$\Psi_P(P) \leq \left\| V^{-1} \right\|_2 \frac{\alpha \left\| P^2 \right\|_F + \beta \|P\|_F + \gamma}{\|P\|_F} = \Phi_P(P)$$

with

$$\left\| V^{-1} \right\|_2 = \sigma_{min}^{-1}(V) = \text{Sep}^{-1}[(A, AP + B), (I, P)]$$

where σ_{min} is the smallest singular value and Sep *is the difference measure or separation between the pencils $(A, -(AP + B))$ and (I, P).*

Proof. See the appendix here. □

The weaker bound $\Phi_P(P)$, which corresponds to a standard condition number bound as I demonstrated in the linear model example above, depends on $\left\| V^{-1} \right\|_2$, which corresponds here to the inverse of the smallest singular value of V. The sharper bound $\Psi_P(P)$ takes the interaction of coefficient, solution, and perturbed matrices via the special Kronecker structure into account, that $\Phi_P(P)$ does not. The weaker bound $\Phi_P(P)$ separates V^{-1} is insightful as to the source of ill conditioning in P, whether the system in V is well conditioned, and will be more easily computed, which is particularly useful for large models.

The smallest singular value behind the condition number is directly given by a concept that relates the underlying coefficient matrices in V, the pencil separation. Following

Stewart (1973), Stewart and Sun (1990, Theorems 2.3 and 2.5, pp. 233-234), and Demmel and Kågström (1987), the separation between two pencils (A,B) and (C,D) is given by

$$\text{Sep}[(A,B),(C,D)] \equiv \min_{\|X\|_F=1} \|AXD - BXC\|_F \tag{71}$$

which is applied to computing stable eigendecomposiitons in the references above. For Sylvester equations, $AXD - BXC = E$, the eigenvalues of these two pencils must form disjoint spectra, see Chu (1987) and Lan and Meyer-Gohde (2014) for its application to perturbation in DSGE models, and Higham (1993) and Kågström (1994) for its role in the sensitivity of the solution of Sylvester equations. The separation can be related to the minimal singular value of V via the relationship between the Frobenius and Euclidean norms $\|X\|_F = \|\text{vec}(X)\|_2$ and the relationship between the Kronecker product and columnwise vectorization $\text{vec}(ABC) = \left(C' \otimes A\right)\text{vec}(B)$

$$\min_{\|X\|_F=1} \|AXD - BXC\|_F = \min_{\|\text{vec}(X)\|_2=1} \left\| \left[\left(D' \otimes A\right) - \left(C' \otimes B\right)\right] \text{vec}(X) \right\|_2 \equiv \sigma_{min}\left[\left(D' \otimes A\right) - \left(C' \otimes B\right)\right] \tag{72}$$

which follows from the Kronecker refomulation of the Sylvester equation, $AXD - BXC = E$, to a standard linear system $Zx = v$ via $\left[\left(D' \otimes A\right) - \left(C' \otimes B\right)\right]\text{vec}(X) = \text{vec}(E)$.

For the specific DSGE problem in P this is

$$\sigma_{min}(V) = \text{Sep}[(A, -(AP+B)),(I,P)] \tag{73}$$

and hence this is the separation between two pencils $(A, -(AP+B))$ and (I,P) is given by

$$\text{Sep}[(A, -(AP+B)),(I,P)] \equiv \min_{\|X\|_F=1} \|AXP + (AP+B)X\|_F \tag{74}$$

As proven by Lan and Meyer-Gohde (2014), (9) can be factored with a solvent P as

$$M(\lambda) \equiv A\lambda^2 + B\lambda + C = (A\lambda + AP + B)(I_{ny}\lambda - P) \tag{75}$$

I.e. factoring the entire set of eigenvalues of the quadratic problem into those of the solvent P and the remaining eigenvalues contained in $A\lambda + AP + B$. That is, two pencils above are the two pencils the union of whose spectra is the spectrum of the underlying DSGE problem. Following Chu (1987), these two pencils must form disjoint spectra for the Sylvester equation $AXP + (AP+B)X$ to be solvable. That the quantitites of pencil separation and disjoint spectra (eigenvalue separation) are related should be apparent. These two measures, however, can differ arbitrarily, see Stewart (1973, pp. 754-755) and the comparison later here, and the appropriate measure is the separation.

Following the approach for P and beginning with Q defined as in (64) through $FQ = -D$ where $F \equiv AP + B$ and for normwise perturbations

$$\epsilon = \max\left\{\frac{\|\Delta F\|_F}{\phi}, \frac{\|\Delta D\|_F}{\delta}\right\}$$

normwise relative perturbations in Q, $\frac{\|\Delta Q\|_F}{\|Q\|_F}$, from the perturbed linear equation

$$(F + \Delta F)(Q + \Delta Q) + D + \Delta D = 0$$

can be bounded to first order in ϵ as follows

Theorem 6 (Condition Number of Q via (64))

The relative perturbation in Q is bounded to first order in ϵ by

$$\frac{\|\Delta P\|_F}{\|P\|_F} \leq \sqrt{2}\Psi_{Q_1}(Q)\epsilon + \mathcal{O}(\epsilon^2)$$

where $\Psi_{Q_1}(Q)$, the condition number of Q, is given by

$$\Psi_{Q_1}(Q) = \frac{\left\|\left(I_{n_e} \otimes F\right)^{-1}\left[\phi Q' \otimes I_{n_y} \quad \delta I_{n_e n_y}\right]\right\|_2}{\|Q\|_F}$$

Proof. See the appendix here. □

This sharp bound can be weakened to

Corollary 3 (Bound of Condition Number of Q via (64))

The condition number of Q, $\Psi_{Q_1}(Q)$, can be bounded by

$$\Psi_{Q_1}(Q) \leq \left\|F^{-1}\right\|_2 \frac{\phi\|Q\|_F + \delta}{\|Q\|_F} = \Phi_{Q_1}(Q)$$

where

$$\left\|F^{-1}\right\|_2 = \sigma_{min}^{-1}(F)$$

where σ_{min} is the smallest singular value.

Proof. See the appendix here. □

As above, the weaker bound $\Phi_{Q_1}(Q)$ relates the condition number to the condition of the linear system in F. This metric, however, treats F as a primitive, which as $F \equiv AP + B$ is certainly is not. Hence this metric will generally miss sources of ill conditioning.

With Q defined as in (64) through $FQ = -D$ but taking perturbations in A, B, and P in the definition of $F \equiv AP + B$ into account via normwise perturbations

$$\epsilon = \max\left\{\frac{\|\Delta P\|_F}{\xi}, \frac{\|\Delta A\|_F}{\alpha}, \frac{\|\Delta B\|_F}{\beta}, \frac{\|\Delta D\|_F}{\delta}, \right\}$$

normwise relative perturbations in Q, $\frac{\|\Delta Q\|_F}{\|Q\|_F}$, from the perturbed linear equation

$$[(A+\Delta A)(P+\Delta P)+B+\Delta B](Q+\Delta Q)+D+\Delta D=0$$

can be bounded to first order in ϵ as follows

Theorem 7 (Condition Number of Q via (63))

The relative perturbation in Q is bounded to first order in ϵ by

$$\frac{\|\Delta P\|_F}{\|P\|_F} \leq \sqrt{4}\Psi_{Q_2}(Q)\epsilon+\mathcal{O}(\epsilon^2)$$

where $\Psi_{Q_2}(Q)$, the condition number of Q, is given by

$$\Psi_{Q_2}(Q)=\frac{\left\|\left(I_{n_e}\otimes(AP+B)\right)^{-1}\begin{bmatrix}\xi Q'\otimes A & \alpha(PQ)'\otimes I_{n_y} & \beta Q'\otimes I_{n_y} & \delta I_{n_e n_y}\end{bmatrix}\right\|_2}{\|Q\|_F}$$

Proof. See the appendix here. □

This sharp bound can be weakened to

Corollary 4 (Bound of Condition Number of Q via (63))

The condition number of Q, $\Psi_{Q_2}(Q)$, can be bounded by

$$\Psi_{Q_2}(Q)\leq\left\|(AP+B)^{-1}\right\|_2\frac{\xi\|Q\|_F\|A\|_F+\alpha\|P\|_F\|Q\|_F+\beta\|Q\|_F+\delta}{\|Q\|_F}=\Phi_{Q_2}(Q)$$

where

$$\left\|(AP+B)^{-1}\right\|_2=\sigma_{min}^{-1}(AP+B)$$

where σ_{min} is the smallest singular value.

Proof. See the appendix here. □

Comparing to the measures for Q_1 above, the leading smallest singular value contribution remains unchanged as $F\equiv AP+B$, so additional ill conditioning comes from the ill conditioning of P but not through solving for Q.

Taking that P itself is a function of A, B, and C into account and considering normwise perturbations accordingly

$$\epsilon=\max\left\{\frac{\|\Delta A\|_F}{\alpha},\frac{\|\Delta B\|_F}{\beta},\frac{\|\Delta C\|_F}{\gamma},\frac{\|\Delta D\|_F}{\delta}\right\}$$

relative normwise perturbations in Q, $\frac{\|\Delta Q\|_F}{\|Q\|_F}$, from the perturbed linear equation

$$[(A+\Delta A)(P+\Delta P)+B+\Delta B](Q+\Delta Q)+D+\Delta D=0$$

where ΔP satisfies

$$(A+\Delta A)(P+\Delta P)^2+(B+\Delta B)(P+\Delta P)+C+\Delta C=0$$

can be bounded to first order in ϵ as follows

Theorem 8 (Condition Number of Q via (63) with Dependency of P Taken into Account)
The relative perturbation in Q is bounded to first order in ϵ by

$$\frac{\|\Delta P\|_F}{\|P\|_F}\leq\sqrt{4}\Psi_{Q_3}(Q)\epsilon+\mathcal{O}(\epsilon^2)$$

where $\Psi_{Q_3}(Q)$, the condition number of Q, is given by

$$\Psi_{Q_3}(Q)=\Big\|\Big[\alpha\left(Q'\otimes I_{n_y}\right)V^{-1}\left(P'\otimes I_{n_y}\right)\quad\beta\left(Q'\otimes I_{n_y}\right)V^{-1}\quad\cdots$$
$$-\gamma\left(Q'\otimes(AP+B)^{-1}A\right)V^{-1}\quad\delta I_{n_e}\otimes(AP+B)^{-1}\Big]\Big\|_2/\|Q\|_F$$

Proof. See the appendix here. □

This sharp bound can be weakened to

Corollary 5 (Bound of Condition Number of Q via (63) with Dependency of P Taken into Account)
The condition number of Q, $\Psi_{Q_3}(Q)$, can be bounded by

$$\Psi_{Q_3}(Q)\leq\left\|V^{-1}\right\|_2\frac{\alpha\|Q\|_F\|P\|_F+\beta\|Q\|_F}{\|Q\|_F}+\left\|V^{-1}\right\|_2\left\|(AP+B)^{-1}\right\|_2\frac{\gamma\|Q\|_F\|A\|_F}{\|Q\|_F}$$
$$+\left\|(AP+B)^{-1}\right\|_2\frac{\delta}{\|Q\|_F}=\Phi_{Q_3}(Q)$$

Proof. See the appendix here. □

Notice now that in addition to the ill conditioning in P treated as a data matrix for the problem that can affect the conditioning of Q as captured above in Q_2, the problem in P also affects the solving for Q, with now the smallest singular value of V that entered in the conditioning analysis for P is also present here in the condition number of Q. This is quite natural, as P is not a primitive as is was treated in Q_2 above and ill conditioning of its solution will likely be translated to an ill conditioned solution of Q.

The weaker bounds for the condition numbers of Q can be ranked as follows

Corollary 6 (Rankings of the Bounds of Condition Numbers of Q)
The bounds on the condition numbers of Q can be ranked by

$$\Phi_{Q_1}(Q)\leq\Phi_{Q_2}(Q)\leq\Phi_{Q_3}(Q)$$

Proof. See the appendix here. □

This follows from the successive admission of F; then A, B, and P; and finally A, B, and C as primitives in the problem in Q.

Instead of examining P and Q individually, we ca, consider the conditioning of the entire solution of the linear DSGE model. Hence, considering P and Q jointly via the matrix $\begin{bmatrix} P & Q \end{bmatrix}$ as a function of A, B, C, and D and measuring perturbations normwise as

$$\epsilon = \max\left\{\frac{\|\Delta A\|_F}{\alpha}, \frac{\|\Delta B\|_F}{\beta}, \frac{\|\Delta C\|_F}{\gamma}, \frac{\|\Delta D\|_F}{\delta}\right\}$$

relative perturbations in P and Q, $\dfrac{\left\|\begin{bmatrix} \Delta P & \Delta Q \end{bmatrix}\right\|_F}{\left\|\begin{bmatrix} P & Q \end{bmatrix}\right\|_F}$, from the perturbed system

$$\begin{aligned}(A+\Delta A)(P+\Delta P)\begin{bmatrix} P+\Delta P & Q+\Delta Q \end{bmatrix} \\ +(B+\Delta B)\begin{bmatrix} P+\Delta P & Q+\Delta Q \end{bmatrix} \\ +\begin{bmatrix} C+\Delta C & D+\Delta D \end{bmatrix} = \begin{bmatrix} 0 & 0 \end{bmatrix}\end{aligned}$$

can be bounded to first order in ϵ as follows

Theorem 9 (Joint Condition Number of P and Q)
The relative perturbation in $\begin{bmatrix} P & Q \end{bmatrix}$ is bounded to first order in ϵ by

$$\frac{\left\|\begin{bmatrix} \Delta P & \Delta Q \end{bmatrix}\right\|_F}{\left\|\begin{bmatrix} P & Q \end{bmatrix}\right\|_F} \leq \sqrt{4}\Psi_{PQ}(P,Q)\epsilon + \mathcal{O}(\epsilon^2)$$

where $\Psi_{PQ}(P,Q)$, the condition number of $\begin{bmatrix} P & Q \end{bmatrix}$, is given by

$$\Psi_{PQ}(P,Q) = \frac{\|W^{-1}X\|_2}{\left\|\begin{bmatrix} P & Q \end{bmatrix}\right\|_F}$$

where

$$W = I_{n_y+n_e} \otimes (AP+B) + \begin{bmatrix} P' & 0_{n_y \times n_e} \\ Q' & 0_{n_e \times n_e} \end{bmatrix} \otimes A$$

and

$$X = \begin{bmatrix} \alpha \begin{bmatrix} P^{2\prime} \\ Q'P' \end{bmatrix} \otimes I_{n_y} & \beta \begin{bmatrix} P' \\ Q' \end{bmatrix} \otimes I_{n_y} & \gamma \begin{bmatrix} I_{n_y} \\ 0_{n_e \times n_y} \end{bmatrix} \otimes I_{n_y} & \delta \begin{bmatrix} 0_{n_y \times n_e} \\ I_{n_e} \end{bmatrix} \otimes I_{n_y} \end{bmatrix}$$

Proof. See the appendix here. □

This sharp bound can be weakened to

Corollary 7 (Bound of Joint Condition Number of P and Q)
The condition number of $\begin{bmatrix} P & Q \end{bmatrix}$, $\Psi_{PQ}(P,Q)$, *can be bounded by*

$$\Psi_{PQ}(P,Q) \leq \left\|W^{-1}\right\|_2 \frac{\alpha\left(\left\|P^2\right\|_2 + \|QP\|_2\right) + \beta\left(\left\|P^2\right\|_2 + \|Q\|_2\right) + \gamma + \delta}{\left\|\begin{bmatrix} P & Q \end{bmatrix}\right\|_F} = \Phi_{PQ}(P,Q)$$

where

$$\left\|W^{-1}\right\|_2 = \sigma_{min}^{-1}(W) = \sigma_{min}^{-1}(V) = \left\|V^{-1}\right\|_2 = \mathrm{Sep}^{-1}[(AP+B,A),(I,P)]$$

where σ_{min} *is the smallest singular value and* Sep *is the difference measure between the pencils* $(AP+B,A)$ *and* (I,P).

Proof. See the appendix here. □

The weaker bound shows that condition number of the entire problem rests on two components the conditioning of the underlying matrices A, B, C, and D, as summarized by the fraction in $\Phi_{PQ}(P,Q)$ and the inverse of the smallest singular value of W. The further development shows that this is the smallest singular value of V and hence the conditioning of the entire linear DSGE model hinges on the conditioning of the quadratic problem in P.

Table 2 contains an overview of all the condition numbers - the stronger bounds all differ from their respective weaker bounds in that they consider the right hand or data matrices jointly with the left hand matrix. That is, they acknowledge the Kronecker structure while solving the problem that can lead to a canceling or amelioration of some conditioning problems that will be overlooked when splitting the two sides of the problem. Again, the weaker bound is useful in theory and practice due to its diagnostic perspective (I have shown that the smallest singular value of W is identical to that of V) and is more easily computed numerically, especially advantageous for larger models.

3.3. **Practical Forward Error Bounds for Linear DSGE Model Solutions.** Of particular interest, especially to practitioners, is a measure of the accuracy of a calculated solution. That is, beyond measures and bounds of backward errors and condition numbers that reveal sources of numerical instabilities and errors in the problem being solved, we would like to know how accurate a given solution to the problem is. This is often called the a posteriori forward error bound, a posteriori in the sense of after having calculated a solution. I will derive such bounds for P, Q - from the different perspectives on the primitives examined also above, as well as a for $[PQ]$ joint that summarizes accuracy

	Sharp Bound	Weak Bound
P	$\left\|V^{-1}\left[\alpha\left(P^2\right)'\otimes I_{n_y} \quad \beta P'\otimes I_{n_y} \quad \gamma I_{n_{y^2}}\right]\right\|_2/\|P\|_F$	$\left\|V^{-1}\right\|_2\left(\alpha\left\|P^2\right\|_F+\beta\|P\|_F+\gamma\right)/\|P\|_F$
Q_1	$\left\|\left(I_{n_e}\otimes F\right)^{-1}\left[\phi Q'\otimes I_{n_y} \quad \delta I_{n_e n_y}\right]\right\|_2/\|Q\|_F$	$\left\|F^{-1}\right\|_2\left(\phi\|Q\|_F+\delta\right)/\|Q\|_F$
Q_2	$\left\|\left(I_{n_e}\otimes(AP+B)\right)^{-1}\left[\xi Q'\otimes A \quad \alpha(PQ)'\otimes I_{n_y} \quad \beta Q'\otimes I_{n_y} \quad \delta I_{n_e n_y}\right]\right\|_2/\|Q\|_F$	$\left\|(AP+B)^{-1}\right\|_2\left(\xi\|Q\|_F\|A\|_F+\alpha\|P\|_F\|Q\|_F+\beta\|Q\|_F+\delta\right)/\|Q\|_F$
Q_3	$\left\|\left[\alpha\left(Q'\otimes I_{n_y}\right)V^{-1}\left(P'\otimes I_{n_y}\right) \quad \beta\left(Q'\otimes I_{n_y}\right)V^{-1} \quad \cdots\right.\right.$ $\left.\left.-\gamma\left(Q'\otimes(AP+B)^{-1}A\right)V^{-1} \quad \delta I_{n_e}\otimes(AP+B)^{-1}\right]\right\|_2/\|Q\|_F$	$\left\|V^{-1}\right\|_2\frac{\alpha\|Q\|_F\|P\|_F+\beta\|Q\|_F}{\|Q\|_F}+\left\|V^{-1}\right\|_2\left\|(AP+B)^{-1}\right\|_2\frac{\gamma\|Q\|_F\|A\|_F}{\|Q\|_F} \quad \cdots$ $+\left\|(AP+B)^{-1}\right\|_2\frac{\delta}{\|Q\|_F}$
$\begin{bmatrix}P & Q\end{bmatrix}$	$\left\|W^{-1}X\right\|_2/\left\|\begin{bmatrix}P & Q\end{bmatrix}\right\|_F$	$\left\|W^{-1}\right\|_2\left[\alpha\left(\left\|P^2\right\|_2+\|QP\|_2\right)+\beta\left(\left\|P^2\right\|_2+\|Q\|_2\right)+\gamma+\delta\right]/\left\|\begin{bmatrix}P & Q\end{bmatrix}\right\|_F$

TABLE 2. Conditioning Number Bounds for Linear DSGE Model Solutions

- R_P and R_Q are the residuals of (5) and (6) respectively
- $V = I_{n_y}\otimes(AP+B)+P'\otimes A$
- $F = A\hat{P}+B$
- $W = I_{n_y+n_e}\otimes(AP+B)+\begin{bmatrix}P' & 0_{n_y\times n_e}\\ Q' & 0_{n_e\times n_e}\end{bmatrix}\otimes A$
- $X = \left[\alpha\begin{bmatrix}P^{2\prime}\\ Q'P'\end{bmatrix}\otimes I_{n_y} \quad \beta\begin{bmatrix}P'\\ Q'\end{bmatrix}\otimes I_{n_y} \quad \gamma\begin{bmatrix}I_{n_y}\\ 0_{n_e\times n_y}\end{bmatrix}\otimes I_{n_y} \quad \delta\begin{bmatrix}0_{n_y\times n_e}\\ I_{n_e}\end{bmatrix}\otimes I_{n_y}\right]$

in one measure. I will also provide two bounds, a tight bound and a weaker boud, with the weaker bound being useful for larger models where the tighter bound might be computationally prohibitive (see also the following section).

Beginning with the transition matrix P, we will take the residual to be the actual residual $A\hat{P}^2 + B\hat{P} + C$ of the solution $\hat{P}$ produced by some solution method.

Corollary 8 (A Posteriori Forward Error Bounds of P)

Let $\hat{P}$ be a computed solution to (5) with residual $R = A\hat{P}^2 + B\hat{P} + C$, the forward error of $\hat{P}$ can be bounded as

$$\frac{\|\Delta P\|_F}{\|\hat{P}\|_F} = \frac{\left\|V^{-1}\operatorname{vec}(R)\right\|_2}{\|\hat{P}\|_F} \leq \left\|V^{-1}\right\|_2 \frac{\|R\|_F}{\|\hat{P}\|_F} \tag{76}$$

where $V = I_{n_y} \otimes (AP + B) + P' \otimes A$.

Proof. Let $\Delta P = \hat{P} - P$, set $\Delta A = \Delta B = 0$ and $\Delta C = R = A\hat{P}^2 + B\hat{P} + C$, and invoke Theorem 5. □

This reiterates the point made often above: a small residual need not be associated with a small error and indeed the same culprit behind a large condition number $\left\|V^{-1}\right\|_2$ can potentiate a small residual into a large error in P.

Now turning to Q and starting with Q_1 that takes F as a primitive, the residual is set to the residual associated with an actual numerical solution $\hat{Q}$.

Corollary 9 (A Posteriori Forward Error Bounds of Q_1)

Let $\hat{Q}_1$ be a computed solution to (64) with residual $R = F\hat{Q}_1 + D$, the forward error of $\hat{Q}_1$ can be bounded as

$$\frac{\|\Delta Q\|_F}{\|\hat{Q}\|_F} = \frac{\left\|\left(I_{n_e} \otimes F^{-1}\right)\operatorname{vec}(R)\right\|_2}{\|\hat{Q}\|_F} \leq \left\|F^{-1}\right\|_2 \frac{\|R\|_F}{\|\hat{Q}\|_F} \tag{77}$$

Proof. Let $\Delta Q = Q - \hat{Q}_1$, set $\Delta F = 0$ and $\Delta D = R = F\hat{Q} + D$, and invoke Theorem 6. □

Here I relax the assumption that F is a primitive and consider F's dependency on A, B, and $\hat{P}$ when assessing $\hat{Q}$

Corollary 10 (A Posteriori Forward Error Bounds of Q_2)

Let $\hat{Q}_2$ be a computed solution to (63) with residual $R = \left(A\hat{P} + B\right)\hat{Q}_2 + D$, the forward error of $\hat{Q}_2$ can be bounded as

$$\frac{\|\Delta Q\|_F}{\|\hat{Q}\|_F} = \frac{\left\|\left(I_{n_e} \otimes \left(A\hat{P} + B\right)^{-1}\right)\operatorname{vec}(R)\right\|_2}{\|\hat{Q}\|_F} \leq \left\|\left(A\hat{P} + B\right)^{-1}\right\|_2 \frac{\|R\|_F}{\|\hat{Q}\|_F} \tag{78}$$

Proof. Let $\Delta Q = Q - \hat{Q}_2$, set $\Delta A = \Delta B = \Delta P = 0$ and $\Delta D = R = \left(A\hat{P} + B\right)\hat{Q} + D$, and invoke Theorem 7 . □

Notice that these first two bounds on $\hat{Q}$ are identical. As the $\hat{P}$ used in F and for the residual is taken at face value, the difference is in name only, as I sumamrize in the following

Corollary 11 (Equivalence of the A Posteriori Forward Error Bounds of Q_1 and Q_2)
The bounds in (77) and (78) are identical for $F = A\hat{P} + B$.

Proof. Inspection. □

Now I calculate forward error bounds on $\hat{Q}$ that takes the forward error of $\hat{P}$ used in the calculation of F into account

Corollary 12 (A Posteriori Forward Error Bounds of Q_3)
Let $\hat{Q}_3$ be a computed solution to (63) with residual $R_Q = \left(\left(A\hat{P} + B\right)\hat{Q} + D\right)\hat{Q}_3 + D$, with $\hat{P}$ a computed solution to (5) with residual $R_P = A\hat{P}^2 + B\hat{P} + C$, the forward error of $\hat{Q}_3$ can be bounded as

$$\frac{\|\Delta Q\|_F}{\|\hat{Q}\|_F} \leq \left\| \left(\hat{Q}' \otimes \left[\left(A\hat{P} + B\right)^{-1} A\right]\right) V^{-1} \operatorname{vec}(R_P) - \left(I_{n_e} \otimes \left(A\hat{P} + B\right)^{-1}\right) \operatorname{vec}(R_Q) \right\|_2 / \|\hat{Q}\|_F \tag{79}$$

$$\leq \left\| \left(A\hat{P} + B\right)^{-1} \right\|_2 \left(\frac{\|R_Q\|_F}{\|\hat{Q}\|_F} + \left\|V^{-1}\right\|_2 \|A\|_2 \|R_P\|_F \right) \tag{80}$$

where $V = I_{n_y} \otimes (AP + B) + P' \otimes A$.

Proof. Let $\Delta Q = Q - \hat{Q}_3$; set $\Delta A = \Delta B = 0$, $\Delta C = R_P = A\hat{P}^2 + B\hat{P} + C$, and $\Delta D = R_Q = \left(A\hat{P} + B\right)\hat{Q} + D$; and invoke Theorem 8. Details in the appendix here. □

As would be expected, the error from $\hat{P}$ affects the error of $\hat{Q}$ through A, as P enters F through AP. The error of $\hat{P}$ is governed by $\left\|V^{-1}\right\|_2$ as it is above.

Note that this can serve only to increase the looser of the bounds, as this bound considers the sources of errors separately

Corollary 13 (Comparison of the Upper A Posteriori Forward Error Bounds of Q_1, Q_2, and Q_3)
The upper bound in (79) is strictly larger than the upper bounds in (77) and (78) for $\hat{P}$ calculated with finite precision.

Proof. For $\left\|\hat{P}\right\| > 0$, $\left\|V^{-1}\right\|_2 \|A\|_2 \|R_P\|_F > 0$ and the result follows by inspection. □

Now I bound the forward error of the entire problem by considering forward error bounds on the joint problem of $\begin{bmatrix} P & Q \end{bmatrix}$

Corollary 14 (A Posteriori Forward Error Bounds of $\begin{bmatrix} P & Q \end{bmatrix}$)
Let $\begin{bmatrix} \hat{P} & \hat{Q} \end{bmatrix}$ *be a computed solution to (5) and (6) with residuals* $R_P = A\hat{P}^2 + B\hat{P} + C$ *and* $R_Q = \left(A\hat{P} + B\right)\hat{Q} + D$, *the forward error of* $\begin{bmatrix} \hat{P} & \hat{Q} \end{bmatrix}$ *can be bounded as*

$$\frac{\left\| \begin{bmatrix} \Delta P & \Delta Q \end{bmatrix} \right\|_F}{\left\| \begin{bmatrix} \hat{P} & \hat{Q} \end{bmatrix} \right\|_F} = \frac{\left\| W^{-1} \left[\begin{bmatrix} I \\ 0 \end{bmatrix} \otimes I \quad \begin{bmatrix} 0 \\ I \end{bmatrix} \otimes I \right] \operatorname{vec}\left(\begin{bmatrix} R_P & R_Q \end{bmatrix} \right) \right\|_2}{\left\| \begin{bmatrix} \hat{P} & \hat{Q} \end{bmatrix} \right\|_F} \tag{81}$$

$$\leq \left\| W^{-1} \right\|_2 \frac{\left\| \begin{bmatrix} R_P & R_Q \end{bmatrix} \right\|_F}{\left\| \begin{bmatrix} \hat{P} & \hat{Q} \end{bmatrix} \right\|_F} = \left\| V^{-1} \right\|_2 \frac{\left\| \begin{bmatrix} R_P & R_Q \end{bmatrix} \right\|_F}{\left\| \begin{bmatrix} \hat{P} & \hat{Q} \end{bmatrix} \right\|_F} \tag{82}$$

where $W = I_{n_y + n_e} \otimes (AP + B) + \begin{bmatrix} P' & 0_{n_y \times n_e} \\ Q' & 0_{n_e \times n_e} \end{bmatrix} \otimes A$ *and* $V = I_{n_y} \otimes (AP + B) + P' \otimes A$.

Proof. Let $\Delta P = P - \hat{P}$ and $\Delta Q = Q - \hat{Q}$; set $\Delta A = \Delta B = 0$ and $\begin{bmatrix} \Delta C & \Delta D \end{bmatrix} = \begin{bmatrix} R_P & R_Q \end{bmatrix}$; and invoke Theorem 9 and Corollary 7. □

As was the case above for the condition number, the looser bound on the joint problem is driven by the same potentiating factor as for the problem $\hat{P}$ alone, $\left\| V^{-1} \right\|_2$. This provides an explanation (at least out to the complete solution at first order, though the equations in all higher order parameters of a nonlinear perturbation solve linear equations analogous to that in Q here, see Lan and Meyer-Gohde (2014)) of the centrality of the accuracy of the quadratic problem in P for the entire solution as pointed out by Anderson, Levin, and Swanson (2006). The relation between the different measures can be more readily see in their juxtaposition in table 3. I now turn to the actual computation of these measures - recall the weaker bounds are useful in that they do not require the solution of the large Kronecker systems in V, F, or W, but only require the smallest singular value or underlying pencil separation.

3.4. **Numerical Considerations in Calculating the Error Bounds and Conditioning Numbers.** Given that the calculations above for determining the accuracy of the solution to (5) and (6) involve calculations such as solving systems in the square of the dimension (via the Kronecker product) of P and Q of singular values that are arguably as complicated as the calculations involved in the solution such as eigenvalues being evaluated, one might ask how accurate the evaluations of the accuracy themselves are. Demmel (1987) and Higham (1995) address this directly, establishing that the condition

	Sharp Bound	Weak Bound
P	$\left\|V^{-1}\operatorname{vec}(R_P)\right\|_2 / \left\|\hat{P}\right\|_F$	$\left\|V^{-1}\right\|_2 \left\|R_P\right\|_F / \left\|\hat{P}\right\|_F$
Q_1	$\left\|\left(I_{n_e} \otimes F^{-1}\right)\operatorname{vec}(R_Q)\right\|_2 / \left\|\hat{Q}\right\|_F$	$\left\|F^{-1}\right\|_2 \left\|R_Q\right\|_F / \left\|\hat{Q}\right\|_F$
Q_2	$\left\|\left(I_{n_e} \otimes \left(A\hat{P}+B\right)^{-1}\right)\operatorname{vec}(R_Q)\right\|_2 / \left\|\hat{Q}\right\|_F$	$\left\|\left(A\hat{P}+B\right)^{-1}\right\|_2 \left\|R_Q\right\|_F / \left\|\hat{Q}\right\|_F$
Q_3	$\left\|\left(\hat{Q}' \otimes \left[\left(A\hat{P}+B\right)^{-1}A\right]\right)V^{-1}\operatorname{vec}(R_P) - \left(I_{n_e} \otimes \left(A\hat{P}+B\right)^{-1}\right)\operatorname{vec}(R_Q)\right\|_2 / \left\|\hat{Q}\right\|_F$	$\left\|\left(A\hat{P}+B\right)^{-1}\right\|_2 \left(\frac{\left\|R_Q\right\|_F}{\left\|\hat{Q}\right\|_F} + \left\|V^{-1}\right\|_2 \left\|A\right\|_2 \left\|R_P\right\|_F\right)$
$\begin{bmatrix} P & Q \end{bmatrix}$	$\left\|W^{-1}\left[\begin{bmatrix} I \\ 0 \end{bmatrix} \otimes I \quad \begin{bmatrix} 0 \\ I \end{bmatrix} \otimes I\right]\operatorname{vec}\left(\begin{bmatrix} R_P & R_Q \end{bmatrix}\right)\right\|_2 / \left\|\begin{bmatrix} \hat{P} & \hat{Q} \end{bmatrix}\right\|_F$	$\left\|V^{-1}\right\|_2 \left\|\begin{bmatrix} R_P & R_Q \end{bmatrix}\right\|_F / \left\|\begin{bmatrix} \hat{P} & \hat{Q} \end{bmatrix}\right\|_F$

TABLE 3. Practical Forward Error Bounds for Linear DSGE Model Solutions

- R_P and R_Q are the residuals of (5) and (6) respectively
- $V = I_{n_y} \otimes (AP+B) + P' \otimes A$
- $F = A\hat{P} + B$
- $W = I_{n_y+n_e} \otimes (AP+B) + \begin{bmatrix} P' & 0_{n_y \times n_e} \\ Q' & 0_{n_e \times n_e} \end{bmatrix} \otimes A$

number of the conditioning number is the condition number - i.e. the sensitivity of the calculation of the condition number is of the same order as the condition number calculated. That is, while the condition number, like the numerical problem being addressed, is not calculated with infinite precision, the numerically calculated condition number accurately reflects the actual condition number in the sense that the numerical errors in its calculation are not of a greater magnitude than the calculated condition number.

For larger models, direct calculation of even the practical froward error bounds above might not be feasible as they involve solving linear systems with and calculating singular values of V which is an $n_y^2 \times n_y^2$ matrix.[17] Note that $V = I_{n_y} \otimes (AP+B) + P' \otimes A$ is a Sylvester operator and the quantity $V^{-1}\text{vec}(R)$ used above can be obtained as $\text{vec}(X)$ where X solves the following

$$(AP+B)X + AXP = R \tag{83}$$

As laid out above when deriving the condition numbers of P, the smallest singular value of a Sylvester operator is related to the associated pencils via $\sigma_{min}\left[\left(P' \otimes A\right) + (I \otimes (AP+B))\right] = \text{Sep}[(A, -(AP+B)), (I, P)]$. Hence, solving for X numerically and calculating the conditioning of the system will give the required quantities. I employ the algorithm of Gardiner, Laub, Amato, and Moler (1992), Gardiner, Wette, Laub, Amato, and Moler (1992), and Hopkins (2002) - ACM Algorithm 705 - which solves for X using a generalized Hessenberg-Schur algorithm directly on the Sylvester equation above.[18] However, emerging algorithms from Köhler (2021) and Köhler (2022) will likely replace this algorithm going forward.

3.5. **Numerical Insights into Generalized Schur or QZ Decompositions.** The generalized Schur or QZ based methods from section 2 are eigenvalue based methods, with the triangular structure of the factorizations revealing the eigenvalues of the underlying inflated matrix pencil. Studies concerning the numerical robustness of generalized eigenvalue problems date back at least to Stewart (1972) and Wilkinson (1979), who provided examples of essentially arbitrary results from the QZ algorithm in the presence of nearly singular pencils, i.e., violation of the regularity assumption above. The computation of eigenvalues numerically is likewise subject to finite precision, Hammarling, Munro, and

[17] And analogously for F, $I_{n_e} \otimes \left(A\hat{P} + B\right)$, and W.

[18] Alternatively, Kågström (1994) and Kågström and Poromaa (1996) solve a related simultaneous system representation of generalized Sylvester equations, see also Chu (1987), which is implemented in LAPACK as ZTGSYL, but this requires inflating the dimensions by doubling the size of the system.

Tisseur (2013) provide a comprehensive study on improving the accuracy of quadratic eigenvalue problems. Anguas, Bueno, and Dopico (2019) provides a comparison of different conditioning numbers for the eigenvalues of matrix polynomials and conditioning numbers of polynomial eiqenvalues can be obtained via eigenvalues for perturbations of the polynomial or pseudospectra (see Tisseur and Higham, 2001; Higham and Tisseur, 2002). Specifically, Tisseur and Higham (2001), Mengi and Overton (2005), and Michiels, Green, Wagenknecht, and Niculescu (2006) apply pseudospectra to stability radii in continuous-time applications.

While an obvious method to assess the accuracy of a numerical eigenvalue λ if a symbolic or analytic value is available would be

$$\max(|\Delta eig|) \equiv \max(|\lambda_{symbolic} - \lambda_{method}|) \tag{84}$$

or some other distance measure such as the chordal distance, the pseudospectrum can provide insight into the necessary count and unit-circle separation of eigenvalues for saddle-path stability in the absence of such symbolic or analytic value. Specifically, the pseudospectrum provides a perturbed analog to the spectrum or set of eigenvalues/latent roots of (9) and (10)

$$\rho_\epsilon(M) = \{\lambda \in \mathbb{C} : (M(\lambda) + \Delta M(\lambda))x = 0 \text{ for some } x \neq 0 \text{ and } \Delta M(\lambda) \tag{85}$$

$$\text{with } \|\Delta A\| \leq \epsilon\alpha_A, \|\Delta B\| \leq \epsilon\alpha_B, \|\Delta C\| \leq \epsilon\alpha_C\} \tag{86}$$

where $\Delta M(\lambda)$ represents the perturbation of the quadratic[19]

$$\Delta M(\lambda) \equiv \Delta A\, \lambda^2 + \Delta B\, \lambda + \Delta C \tag{87}$$

and the α_i's control the perturbation, which are set as $\alpha_X = |X|$ using the 2-norm following Tisseur (2000). As shown in Tisseur and Higham (2001), this 2-norm definition of the pseudopectrum corresponds to the backward errors of the eigenvalues.

As proven in Tisseur (2000), while the QZ or generalized Schur algorithm is numerically stable for the generalized eigenvalue problem (Stewart, 1972), this is not the case for the quadratic eigenvalue problem, as it does not respect the structure of the latter. To see this, first define the pseudospectrum of (13) analogous to above

$$\rho_\epsilon(P_{FG}) = \{\lambda \in \mathbb{C} : (P_{FG}(\lambda) + \Delta P_{FG}(\lambda))x = 0 \text{ for some } x \neq 0 \text{ and } \Delta P_{FG}(\lambda) \tag{88}$$

$$\text{with } \|\Delta F\| \leq \epsilon\alpha_F, \|\Delta G\| \leq \epsilon\alpha_G\} \tag{89}$$

[19] This is perhaps easier to see via the identity $M(\lambda) + \Delta M(\lambda) = (A + \Delta A)\lambda^2 + (B + \Delta B)\lambda + (C + \Delta C)$.

comparing the perturbations involved in (88) with (85)

$$\Delta P_{FG}(\lambda) \equiv \Delta F\lambda - \Delta G = \begin{bmatrix} \Delta F_{11} & \Delta F_{12} \\ \Delta F_{21} & \Delta F_{22} \end{bmatrix} \lambda - \begin{bmatrix} \Delta G_{11} & \Delta G_{12} \\ \Delta G_{21} & \Delta G_{22} \end{bmatrix} \tag{90}$$

$$\neq \left(\begin{bmatrix} I_{n_y} & 0_{n_y \times n_y} \\ 0_{n_y \times n_y} & \Delta A \end{bmatrix} \lambda - \begin{bmatrix} 0_{n_y \times n_y} & I_{n_y} \\ -\Delta C & -\Delta B \end{bmatrix} \right) \begin{bmatrix} I_{n_y} \\ I_{n_y}\lambda \end{bmatrix} = \begin{bmatrix} 0 \\ \Delta M(\lambda) \end{bmatrix} \tag{91}$$

Inspection underscores that, in general, perturbations of the QZ or generalized Schur of the companion linearization (12) do not respect the specific structure in the underlying matrix quadratic problem (5). That is, while the backward stability requires the admission of arbitrary (though bounded, see the definition of the pseudospectrum above) ΔF_{ij} and ΔG_{ij}, the companion linearization mandates that ΔF_{12}, ΔF_{21}, and ΔG_{11} must be zero (i.e., no perturbation is admissible here) and ΔF_{11} and ΔG_{12} must have unit (I_{n_y}) perturbations for the structure of the QZ problem to still represent the quadratic eigenvalue problem under numerical perturbations. This is obviously a source on tension between the QZ algorithm and the solution of the quadratic eigenvalue problem and Tisseur (2000) proves that this tension precisely leads to the QZ algorithm no longer being backward stable for the quadratic eigenvalue problem and hence, by extension as the solution of the matrix quadratic problem cast in terms of the QZ algorithm is numerically analogous to the quadratic eigenvalue problem following, e.g., Higham and Kim (2000), will not be backward stable for the matrix quadratic problem.

3.6. **Existing Error Checks.** The DSGE literature and existing linear implementations of course is not devoid of error and accuracy checks. These, however, are either residual based or eigenvalue based. Both are insufficient given the results above as I will now address.

As reviewed in section 2, the DSGE literature generally seeks solutions P and Q as a unique P with eigenvalues inside the closed unit circle. Hence of the $2n_y$ latent roots λ in (9), there are n_y inside (or on) and n_y outside the unit circle and the former are used to construct $P \in \mathbb{R}^{n_y \times n_y}$ such that $M(P) = 0$ and $|eig(P)| \leq 1$. The generalized Bézout theorem, e.g., Lan and Meyer-Gohde (2014), which states that a lambda-matrix divided on the right by a binomial in a matrix has as a remainder the matrix polynomial associated with the lambda-matrix evaluated at the matrix of the binomial.[20] For the matrix quadratic here

[20]As noted by Gantmacher (1959, Ch. 4) and repeated in **?**, Davis (1981), Higham and Kim (2000), and Higham and Kim (2001), if this matrix in the binomial is a solvent of the matrix polynomial, the division is without remainder, yielding a factorization of the matrix polynomial.

this yields the factorization

$$M(\lambda) = (A\lambda + AP + B)(I_{ny}\lambda - P) \tag{92}$$

If there is a unique stable solution, then the pencil (I_{ny}, P) contains eigenvalues only inside or on the unit circle and the pencil $(A, -(AP+B))$ contains only eigenvalues only outside the unit circle. Hence, the matrices/Sylvester operators V, F, and W that are the cornerstones of the condition numbers above are nonsingular, as the main theorem of Chu (1987) that requires the disjoint spectra (or sets of eigenvalues) for these pencils applies.

A natural assumption would be that the closer these spectra are, the closer V, F, and W are to being singular and, hence, the more ill conditioned a DSGE model is, the larger the practical forward error bounds are, and, finally, the less accurate the solution produced by a numerical algorithm is likely to be. Indeed those algorithms from above that implement some numerical check do exactly this (`qz_criterium` in Dynare, `TOL` in Uhlig's (1999) Toolkit and `div` in Sims's (2001) Gensys) if any unstable eigenvalue comes too close to the unit circle, implying that the smallest possible distance between the moduli of elements of the two spectra.

Unfortunately, this is insufficient as the distance between the two spectra is measured by the distance or separation between the two pencils as demonstrated by the results above and these two measures can differ arbitrarily, a result well established in the numerical literature. The bounds on the condition numbers are scaled by $\left\|V^{-1}\right\|_2$, $\left\|F^{-1}\right\|_2$, and $\left\|W^{-1}\right\|_2$, which correspond to the inverses of their lowest singular values. For $V = I_{n_y} \otimes (AP+B) + P' \otimes A$ this is $\mathrm{Sep}^{-1}[(A, AP+B), (I, P)]$. Note that the two pencils in the difference measure are precisely the two pencils whose spectra must be disjoint. A recent and succinct presentation of the difference between the pencil and spectra separation is given by Chen and Lv's (2018) Theorem 2.3

$$\mathrm{Sep}[(A,B),(C,D)] \leq \min_{i=1,2,\ldots,n_y; j=1,2,\ldots,n_y} \left|\alpha_i \delta_j - \beta_i \gamma_j\right| \tag{93}$$

where $\alpha_i = \lambda_i \beta_i$ and $\gamma_j = \mu_j \delta_j$ are the generalized (extended to "infinite" eigenvalues) of the pencils (A,B) and (C,D) respectively. This by itself is not conclusive, but equality only holds for definite matrix pairs, see Stewart and Sun (1990), with positive definite A and B will all eigenvalues real and semi simple, obviously not generic properties in DSGE models. Stewart (1973, pp. 754-755) demonstrates how disconnected the pencil and spectra separation can be: scaling A, B, C, and D all with $\sigma \neq 0$ leaves the eigenvalues

unchanged, $\sigma\alpha_i = \lambda_i \sigma\beta_i \Leftrightarrow \alpha_i = \lambda_i \beta_i$, but scales the pencil separation

$$\text{Sep}[(\sigma A, \sigma B),(\sigma C, \sigma D)] = \sigma \, \text{Sep}[(A,B),(C,D)] \tag{94}$$

Varah (1979) provides several examples how "incredibly small" this pencil separation can be and emphasize that "it is extremely important to realize that Sep can be very small even though the eigenvalues [...] are well separated." In sum, current implementations of (Moler and Stewart, 1973) QZ missed the contemporaneous work by one of the same authors, Stewart (1973), that would have pointed them away from the eigenvalue separation towards the pencil separation that shows up here. The consequences of this will be seen in the examples that follow.

The DSGE literature also has several measures to characterize the accuracy of a solution that are usually applied in the context of a nonlinear model. The two most prominent are Judd's (1998) Euler equation errors and the Haan and Marcet (1994) test and both of these statistics are residual based measures. Judd's (1998) Euler equation error statistic calculates the average or maximal one-step error in, say, the h'th equation of the nonlinear model (1) from using some approximation $\hat{y}_t = \hat{y}(y_{t-1}, \varepsilon_t)$ to the solution (2) over some range of the state space $y_{t-1} \times \varepsilon_t$[21]

$$NLEE_h(y_{t-1}, \varepsilon_t) = E_t[f_h(\hat{y}(\hat{y}(y_{t-1}, \varepsilon_t), \varepsilon_{t+1}), \hat{y}(y_{t-1}, \varepsilon_t), y_{t-1}, \varepsilon_t)] \tag{95}$$

Instead of this nonlinear measure, let us use the same method in the linear model (3) to assess the accuracy of different linear solutions $\hat{y}_t = \hat{P}\, y_{t-1} + \hat{Q}\, \varepsilon_t$

$$LEE(y_{t-1}, \varepsilon_t) = AE_t\left[\hat{P}\left(\hat{P}\, y_{t-1} + \hat{Q}\, \varepsilon_t\right) + \hat{Q}\, \varepsilon_{t+1}\right] + B\left(\hat{P}\, y_{t-1} + \hat{Q}\, \varepsilon_t\right) + C y_{t-1} + D\varepsilon_t \tag{96}$$

$$= A\hat{Q}E_t[\varepsilon_{t+1}] + \left(A\hat{P}^2 + B\hat{P} + C\right) y_{t-1} + \left(A\hat{P}\hat{Q} + B\hat{Q} + D\right) \varepsilon_t \tag{97}$$

As $E_t[\varepsilon_{t+1}]$, the measure, likewise for any row h, entirely reflects the residuals of $0 = AP^2 + BP + C$ for (5) and of $0 = (AP + B)Q + D$ in (6) and (63). As proven in Theorems 1 through 4 above, small residuals do *not* imply small backward errors (and, hence, forward errors). This is essential, as the relevant measure of accuracy is the forward error, which, in terms of the linear policy function, is

$$FE(y_{t-1}, \varepsilon_t) = \hat{y}_t - y_t = \left(\hat{P} - P\right) y_{t-1} + \left(\hat{Q} - Q\right) \varepsilon_t \tag{98}$$

[21] The calculation of the expectation is nontrivial and generally performed with quadrature or the like with respect to ε_{t1} and the error is usually measured in relative terms to consumption, perhaps $y_{k,t}$ the k-th variable in the vector of endogenous variables - but neither of these points are relevant here.

Although making small one-step ahead prediction mistakes with an approximated solution method is without question important, the direct measure of $\hat{y}_t - y_t$ is of even more obvious importance if such a measure is available. This measure is provided by the foregoing analysis (and are simply weighted by values in the state space $y_{t-1} \times \varepsilon_t$ above).

The Haan and Marcet (1994) test is similar in that it also is residual based and, hence, is subject to the same criticism. To see this, define the simulation residuals as

$$SNLR_{t+1} = f(\hat{y}_{t+1}, \hat{y}_t, \hat{y}_{t-1}, \varepsilon_t) \tag{99}$$

where the model is simulated using a sequence $\varepsilon_t{}_{t=1}^{T}$ of draws (also some initial value for the state y_0, but after an appropriate burn-in this should be irrelevant) and an approximated solution $\hat{y}_t = \hat{y}(y_{t-1}, \varepsilon_t)$. Choosing some n_z vector of simulated instruments, z_t, measurable with respect to the information set at t, then it must hold that $E[SR_{t+1} \otimes z_t] = 0$ which has simulated counterpart

$$SNLR_T = \frac{1}{T} \sum_{t=1}^{T} f(\hat{y}_{t+1}, \hat{y}_t, \hat{y}_{t-1}, \varepsilon_t) \otimes z_t \tag{100}$$

and an estimate $NL\Omega_T$ of the variance of $SNLR_T$, Haan and Marcet (1994) give the test statistic

$$J_T = T\, SNLR_T' NL\Omega_T^{-1} SNLR_T \tag{101}$$

which is asymptotically distributed χ^2 with $n_z n_y$ degrees of freedom. Consider now the linear counterpart

$$SLR_{t+1} = A\hat{Q}\varepsilon_{t+1} + \left(A\hat{P}^2 + B\hat{P} + C\right)\hat{y}_{t-1} + \left(A\hat{P}\hat{Q} + B\hat{Q} + D\right)\varepsilon_t \tag{102}$$

$$= A\hat{Q}\varepsilon_{t+1} + \left(A\hat{P}\hat{Q} + B\hat{Q} + D\right)\varepsilon_t + \left(A\hat{P}^2 + B\hat{P} + C\right)\left(\hat{Q}\varepsilon_{t-1} + \hat{P}\hat{Q}\varepsilon_{t-2} + \ldots\right) \tag{103}$$

Taking, without loss of generality, any ε_{t-j}, $j \geq 0$ as the instrument z_t, $E[SR_{t+1} \otimes z_t] = 0$ requires

$$E[SR_{t+1} \otimes z_t] = 0 \Rightarrow \begin{cases} \left(A\hat{P} + B\right)\hat{Q} + D = 0 & j = 0 \\ A\hat{P}^2 + B\hat{P} + C = 0 & \text{otherwise} \end{cases} \tag{104}$$

like the Euler equation errors above, this measure also operates on the residuals of $0 = AP^2 + BP + C$ for (5) and of $0 = (AP + B)Q + D$ in (6) and (63) and again the same criticism applies.

4. Applications

I now turn to two sets of applications to investigate the numerical stability of the different methods, QZ- and non-QZ-based, from section 2 in solving several DSGE models. The first comprises three specific models with specific parameterizations: two production-based asset pricing models, with the production side following standard real business cycle models (Kydland and Prescott, 1982; King and Rebelo, 1999) and habit formation (external and internal) on the part of households following, e.g., Constantinides (1990); Campbell and Cochrane (1999); Campbell (2003), and the medium-scale moentary model of Smets and Wouters (2007). The first model is particularly simple, with only external habit formation added in deviation from standard real business cycle analyses. This choice is made as symbolic solutions of the unknowns in the linearized model are available. The second is the model of Jermann (1998), which features internal habit formation and adjustment costs in the capital accumulation equation and the third is a policy relevant New Keynesian model featuring numerous shocks and frictions. The resulting linearized models for the latter two do not admit reliable symbolic solutions, so analyses of numerical solutions must rely on the numerical diagnostics developed above in section 3. I demonstrate errors of economic significance in all three models with existing QZ methods from the literature and demonstrate that my methods reliably detect these errors and provide the warning missing from the existing methods.

Having established that errors of economic significance can occur in standard solution methods of linear DSGE model, the second set takes a first pass at addressing how prevalent such errors in the literature might be. The first exercise implements the backward error and condition number measures from above in a database of roughly 100 different macroeconomic models from the literature, the suite of models in the Macroeconomic Model Data Base (MMB) (see Wieland, Cwik, Müller, Schmidt, and Wolters, 2012; Wieland, Afanasyeva, Kuete, and Yoo, 2016), a model comparison initiative at the Institute for Monetary and Financial Stability (IMFS), [22] chosen to assess the different methods' performance in as non-model specific an environment as possible. Then I turn the measures of accuracy over a set of draws from the posterior of the model of Smets and Wouters (2007). Fortunately, backward errors of economic significance were not found in either exercise, although differences in the accuracy of linear solution methods from the literature differ in some instances by multiple orders of magnitude.

[22] See http://www.macromodelbase.com

4.1. **A Simple Log Normal DSGE Asset Pricing Model.** The first model I examine is chosen specifically because it will admit a closed form solution for the coefficient matrices P and Q while still being capable of generating a macro variable of interest, consumption, and a non trivial financial variable, the risk premium via log normality. It is a toy production-based asset pricing model, based on a standard real business cycle model (Kydland and Prescott, 1982; King and Rebelo, 1999) with external habit formation and a power utility kernel. (Constantinides, 1990; Campbell and Cochrane, 1999; Campbell, 2003) The representative household seeks to maximize

$$E_0 \sum_{t=0}^{\infty} \beta^t u(c_t, X_t), \ 0 < \beta < 1 \tag{105}$$

where c_t is consumption and X_t the external habit stock, subject to

$$c_t + k_t = e^{z_t} k_{t-1}^{\alpha} + (1-\delta) k_{t-1}, \ 0 < \alpha, \delta < 1 \tag{106}$$

where k_t is the capital stock accumulated at time t and z_t is total factor productivity that follows the AR(1) process

$$z_t = \rho z_{t-1} + \omega \varepsilon_t, \ \varepsilon_t \overset{\text{i.i.d.}}{\sim} N(0,1), \ |\rho| < 1, \ 0 < \omega \tag{107}$$

The first order condition of the maximization problem is

$$1 = E_t \left[\underbrace{\beta \frac{u_c(c_{t+1}, X_{t+1})}{u_c(c_t, X_t)}}_{m_{t+1}} \underbrace{\left(\alpha e^{z_{t+1}} k_t^{\alpha - 1} + 1 - \delta \right)}_{R_{t+1}} \right] \tag{108}$$

where m_{t+1} is the stochastic discount factor or pricing kernel and R_{+1} is the (risky) return on capital. Assuming an external habit such that $X_t = c_{t-1}$ in equilibrium with h the degree of habit formation and power or CRRA utility with risk coefficient σ, marginal utility is $u_c(c_t, X_t) = (c_t - h c_{t-1})^{-\sigma}$. Equations (106)-(108) characterize a equilibrium for the stochastic sequences $\{c_t, k_t, z_t\}_{t=0}^{\infty}$ given a sequence of shocks $\{\varepsilon_t\}_{t=0}^{\infty}$ and initial conditions c_{-1}, k_{-1}, z_{-1}.

Defining the steady state, values $\overline{c}, \overline{k}, \overline{z}$ that solve (106)-(108) with $\varepsilon_t = 0 \, \forall t$, equations (106) and (108) can be log-linearized around these values to yield

$$0 = A E_t[y_{t+1}] + B y_t + C y_{t-1} + D z_t, \ y_t = \begin{bmatrix} \hat{c}_t & \hat{k}_t \end{bmatrix}' \tag{109}$$

$$z_t = \rho z_{t-1} + \omega \varepsilon_t, \ \varepsilon_t \overset{\text{i.i.d.}}{\sim} N(0,1) \tag{110}$$

a 2 by 2 system of equations linear in the log-deviations of the endogenous variables, c_t and k_t, from their steady states, $\hat{w}_t \equiv \log w_t - \log \overline{w}$, for $w \in c, k$.

Following Hansen and Singleton (1983); Campbell and Shiller (1988); Campbell (2003), risky (say, R_t from above) and risk-free (via no arbitrage, $1 = E_t[m_{t+1}]R_t^f$) assets can be priced under the implied joint log-normality of the approximation above via

$$1 = E_t\left[e^{\widehat{m_{t+1}}+\widehat{R_{t+1}}}\right] \text{ and } 1 = \overline{R^f}e^{R_t^f}E_t\left[e^{\widehat{m_{t+1}}}\right] \tag{111}$$

which gives the risk premium as $-cov_t(\widehat{m_{t+1}},\widehat{R_{t+1}})$, following, e.g., Lettau (2003), and can be expressed in terms of the variance of z_t (ω^2) as $\left[\frac{\sigma}{1-h}\alpha Q_{cz}\left(1+\beta(1-\delta)\right)\right]^2\omega^2$. Importantly, the coefficient Q_{cz}, the impact of technology on (log) consumption, must be solved for numerically even in this (log) linear case.

The model was chosen to be as simple as possible, in order to enable the symbolic solution of the underlying matrix quadratic problem; see Higham and Kim (2000) who argue that Matlab can successfully solve two-dimensional matrix quadratic problems reliably. I provide numerical results for two calibrations, see table 4, labeled standard and extreme. The standard calibration follows the RBC literature (see, e.g., King and Rebelo, 1999) with the degree of habit formation, h and curvature in the utility function, σ, elevated to match an equity premium of 7.8 in annual percentage points following Mehra (2003) for the post-war US and ω, the standard deviation of the technology shock, adjusted to deliver a standard deviation of consumption growth, $std(\log c_t)$, of 0.566 in quarterly percent, in line again with the post-war US experience. The extreme calibration is chosen to bring the eigenvalue separation between the stable and unstable pencils closer together, while maintaining the match of the symbolic solution to the equity premium and consumption growth volatility.

	h	β	δ	α	σ	ρ	ω
Standard	0.966	0.99	0.025	0.36	98.1	0.95	0.134
Extreme	1-3.907E-05	1-1.750E-10	0.6715	1-5.751E-05	9.151	1-5.184E-04	3.068E-03

TABLE 4. Calibrations

Besides assessing whether the different solution methods are able to recover the exact solutions for the two calibration targets, I examine the underlying causes of a degeneration in accuracy following the results of the previous sections. Namely the largest absolute deviation in the matrices for the linear solution or policy function (4), P and Q, and the largest absolute difference in the finite eigenvalues of the quadratic eigenvalue problem (10) relative to the symbolic solution, and the separation between the calculated stable and unstable eigenvalues along with the relative residuals, backward error bounds, pencil

separations, conditioning numbers and bounds on the forward error of the solutions, P and Q, produced by the various methods. Additionally, I provide plots of the pseudospectra of the matrix quadratic (85) and of the QZ companion linearization (88) used to solve for P. The results that are referred to as "symbolic" are solved symbolically and evaluated using Mathlab's VPA (variable precision arithmetic) with 100 digits of accuracy.

Table 5 contains the results for P under the standard calibration. The first line contains the equity premium predicted by the different methods and all of the methods successfully predict an equity premium of 7.8 annual percentage points, likewise the volatility of consumption growth, the third line, is identical across methods. Upon closer examination, the second line, the difference between the symbolic equity premium and that predicted by the varying methods differs across methods. The most accurate methods being those of Binder and Pesaran (1997) and the cyclic reduction method of Dynare, with all QZ-based methods apart from Dynare displaying degrees of accuracy several orders of magnitude lower. As laid out in Villemot (2011), Dynare reduces the problem solved with the QZ algorithm by, among others, eliminating zero column variables in the A and C matrices of the linear system (3); this is in line with one of the suggestions by Hammarling, Munro, and Tisseur (2013) to improve the accuracy of the quadratic eigenvalue problem. This is reflected in the fifth line of the table, where the largest error in the finite eigenvalues calculated by Dynare are in line with the non-QZ-based methods, those of the remaining QZ-based methods are several orders of magnitude larger, and that of Binder and Pesaran (1997) being the most accurate. The errors in the resulting matrix for the recursive component of the linear solution or policy function (4), P are roughly of the same order of magnitude as the eigenvalue errors. Despite the differences in the accuracy of calculating the eigenvalues, all of the methods yield the same eigenvalue and pencil separations and the conditioning numbers of the solvent P are likewise consistent across methods. Based on this standard calibration, the differences in the solutions generated by the different methods are of no economic consequence. Yet as indicated by the fourth line (alongside the differences in the risk premium in the second line), the methods differ in a numerically consequential and, more importantly, predictable manner. Note that the relative ordering of accuracy in the predicted equity premium, with Binder and Pesaran (1997) being the most down to Klein (2000) being the least accurate, is reflected in the accuracy of the solvent P in the fourth line as measured by the largest entrywise absolute deviation from the symbolic solution.

The symbolic solution will generally not be available, and the methods of the previous sections provide backward-forward error decompositions and argue theoretically they are superior to potential alternatives based on eigenvalue separations or residuals. Beginning with the residuals versus backward errors, the ordering of methods is well captured by the relative residuals although the variation in the growth or amplification factors, $\mu_P(\hat{P})$, indicate a variation in backward errors not captured by the relative residual. Neither the eigenvalue nor pencil separation produced by different methods are by themselves informative, as their values do not varying by the same orders of magnitude as the solution or moment differences. The agrement on the pencil separation contributes to the agrement on the condition number bounds. The order of the condition numbers corresponds roughly (see Judd (1998) or Higham (2002)) to a worst case loss of four significant digits in solutions, which corresponds cleanly to the differences in the orders of magnitude in the backward errors and the forward errors. Notice in particular that the tight, forward error bound 1 using the Frobenius norm corresponds in magnitude to the largest entrywise absolute difference in P relative to the symbolic solution. The forward bound 2 is a looser bound theoretically and indeed does not bound the errors in P as tightly as bound 1 - its advantage, however, is its calculation enables it to be applied to larger models as will be explored later. In summary, the forward error bounds that I provide here provide the same order of magnitude information about the accuracy of a numerically calculated solution that the presence of a symbolic solution would allow.

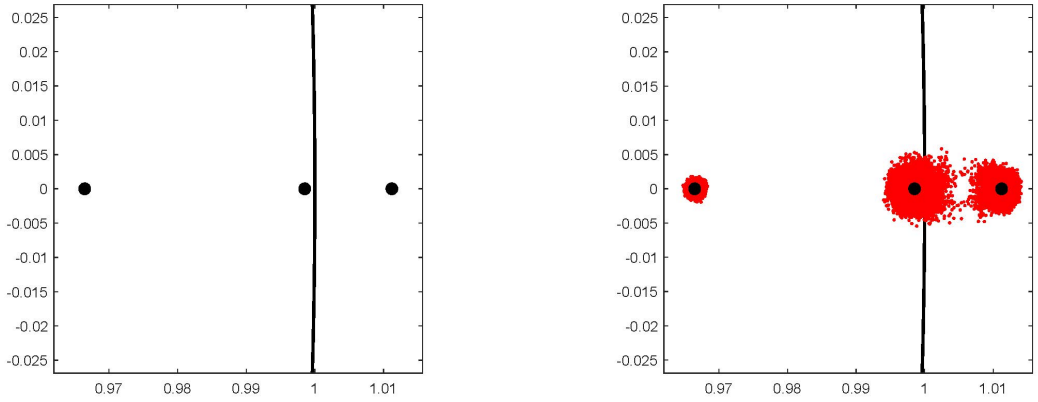
Figure 1 plots the pseudospectra for the extreme standard of the matrix quadratic (85) – in blue – and of the QZ algorithm (88) – in red – against the symbolic eigenvalues – in black – for two different sizes of perturbations. In the left panel, the pseudospectra are not visible, as they overlap with the symbolic results for perturbations of this size. For slightly larger perturbations (right panel), the pseudospectrum of the QZ algorithms encompasses the unit circle while that of the matrix quadratic remains invisible at this scale. This, following Tisseur and Higham (2001), indicates that the backward error in calculating the eigenvalues is not only larger than under the QZ algorithm than with the matrix quadratic, consistent with Tisseur (2000) and with the backward forward analysis of the solvents P in table 5, but also that the stable and unstable eigenvalues are potentially indistinguishable numerically.

Table 6 continues the results for the standard calibration, focussing now on the results that pertain to Q, the shock impact matrix. While the P matrix is most obviously subjected

Measure		QZ-Based Methods				Alternatives		
	Symbolic	Klein (2000)	Sims (2001)	Uhlig (1999)	Dynare QZ	Anderson (2010)	BP (1997)	Dynare CR
$E[rp]$	7.8	7.8	7.8	7.8	7.8	7.8	7.8	7.8
$\Delta E[rp]$		2.31e-08	-1.43e-09	2.27e-08	-3.08e-12	1.54e-11	1.71e-12	1.71e-12
$std(\Delta \log c_t)$	0.566	0.566	0.566	0.566	0.566	0.566	0.566	0.566
$\max(\lvert\Delta P\rvert)$		3.93e-11	2.36e-12	3.88e-11	7.17e-14	1.43e-14	1.51e-15	1.83e-14
$\max(\lvert\Delta eig\rvert)$		4.12e-11	2.47e-12	4.13e-11	1.32e-14	2.86e-14	3.11e-15	7.66e-15
Rel. Res.	2.65e-17	9.82e-14	5.88e-15	9.71e-14	2.7e-17	5.41e-17	2.65e-17	1.15e-18
BE Bound	6.84e-17	2.42e-13	1.45e-14	2.4e-13	6.96e-17	1.41e-16	6.83e-17	2.96e-18
μ_P	2.58	2.47	2.46	2.47	2.58	2.61	2.58	2.56
Eig. Sep.	0.0127	0.0127	0.0127	0.0127	0.0127	0.0127	0.0127	0.0127
Pencil Sep.	0.356	0.356	0.356	0.356	0.356	0.356	0.356	0.356
Ψ_P	1.17e+04	1.17e+04	1.17e+04	1.17e+04	1.17e+04	1.17e+04	1.17e+04	1.17e+04
Φ_P	2.86e+04	2.86e+04	2.86e+04	2.86e+04	2.86e+04	2.86e+04	2.86e+04	2.86e+04
FE Bound 1	2.43e-15	3.3e-11	2.01e-12	3.25e-11	4.26e-14	1.13e-14	1.96e-15	1.09e-14
FE Bound 2	7.57e-13	2.81e-09	1.68e-10	2.77e-09	7.71e-13	1.55e-12	7.57e-13	3.3e-14

TABLE 5. Results P: Standard Calibration

- For Dynare, refer to Adjemian, Bastani, Juillard, Mihoubi, Perendia, Ratto, and Villemot (2011). Dynare under QZ-Based Methods is documented in Villemot (2011) and under Alternatives is the cyclic reduction method. BP (1997) refers to Binder and Pesaran (1997).
- $E[rp]$ is expressed in annual %, $std(\log c_t)$ in quarterly %, and $*$ indicates a backward error less than machine precision, $2^{-52} = 2.2204E-16$.



(A) $1e-6$ (B) $2e-5$

FIGURE 1. Pseudospectrum: Standard Calibration x-axis: real component, y-axis: imaginary component, large black dots: eigenvalues, black curve: unit circle, small red dots: pseudospectrum QZ companion linearization (88), small blue dots: pseudospectrum matrix quadratic (85)

to numerical errors being a matrix quadratic problem, Q not only solves a linear problem that depends on P (hence inheriting numerical instabilities) but as it gives the impact effects of shocks on endogenous variables is as if not more important than the transition matrix P. Indeed the calculations above show that the expected risk premium is a function of the square of the element Q_{cz}. The first row of the table repeats the difference in the expected risk premium from the above and the second row gives the elementwise largest absolute difference in each method's Q relative to the symbolic solution. Again the ordering of accuracy in the predicted equity premium, with Binder and Pesaran (1997) being the most down to Klein (2000) being the least accurate, is reflected in the accuracy of the methods' Q.

In comparing the relative residuals and backward errors under Q_1 and Q_2, the difference between standard linear analyses exemplified by Q_1 where F in $F = AP + B$ is taken as the source of errors in calculations and the resulting larger potential errors when the interactions of errors in A, B, and P are taken into consideration - the growth or amplification factors μ increase beyond one in the latter case, reiterating the point from above that backward errors can differ arbitrarily from relative residuals for structured linear systems. While the different methods agree on the different condition numbers Ψ and Φ, the differences between the different condition numbers is instructive: from 1 to 3 with F, then $F = AP + B$ and finally $F = AP(A,B,C) + B$ taken as the coefficient matrix, the condition number becomes progressively larger. That is, the dependence

on the underlying quadratic problem leads to a less well conditioned linear problem than one would deduce by taking F or then P at face value. Finally, the forward error bounds for Q_2 are overly optimistic (and hence not to be trusted as upper bounds) with respect to the entrywise errors in Q in the second line - the forward bound 1 for Q_3 where the dependence of F on A, B, and P and of P on A, B, and C are taking explicitly into account, again provides the same order of magnitude information about the accuracy of a numerically calculated Q as one can obtain with a symbolic solution.

Table 7 continues the results for the standard calibration, focussing on the joint measure $[PQ]$. The first three rows repeat the moment results stated first above together with P for easy reference. Considering P and Q together gives one set of diagnostics for the entire problem and also enables the calculation of backward errors including those for Q that depend on P's dependence on A, B, and C - i.e., explicitly taking the underlying quadratic problem of P into account when considering Q's dependence on P. The results show that when considering P and Q together, the bounds are roughly though slightly less tight than the looser of the individual bounds on P or Q. That is, the accuracy of the combined results is at best as accurate as the less accurate of P and Q. The tight bound on the forward errors, bound 1, gives the same order of magnitude results on the accuracy of the solution as calculated with the symbolical solution available for the small scae problem here and the loose bound, FE bound 2, provides a computationally less demanding alternative.

Table 8 contains the results for the extreme calibration and the resulting predictions for the two calibration targets now differ significantly across methods.[23] While the non-QZ-based methods continue to maintain a significant match with the calibration targets, lines 1 and 3, the QZ-based methods including Dynare now mispredicts the equity premium by at least 75 annual basis points and as much as 3 annual percentage points, errors of genuine economic significance. The second line, containing the differences of the equity premium predicted by the different methods and the symbolic solution, now show the algorithm of Anderson (2010) as being more accurate than the method of Binder and Pesaran (1997) and the cyclic reduction method of Dynare being several orders of magnitude less accurate than either of the two non QZ-based alternatives. The largest

[23]The moments and difference $E[rp]$, $\Delta E[rp]$, and $std(\Delta \log c_t)$ are identical up to the digits shown in the table for Klein (2000) and Uhlig (1999), out 8 significant digits, these are (first Klein (2000) and then Uhlig (1999)) 7.0455204 and 7.0455406 for $E[rp]$; 0.7538038 and 0.75378359 for $\Delta E[rp]$; and 0.52910871and 0.52910956 for $std(\Delta \log c_t)$. Likewise the FE bound 1 is 6.9737898e-04 and 6.9736054e-04.

Measure		QZ-Based Methods				Alternatives		
	Symbolic	Klein (2000)	Sims (2001)	Uhlig (1999)	Dynare QZ	Anderson (2010)	BP (1997)	Dynare CR
$\Delta E[rp]$		2.31e-08	-1.43e-09	2.27e-08	-3.08e-12	1.54e-11	1.71e-12	1.71e-12
max($\lvert\Delta Q\rvert$)		3.21e-11	1.99e-12	3.14e-11	4.25e-15	2.13e-14	2.4e-15	2.39e-15
Rel. Res. Q_1	1.21e-18	1.13e-14	2.72e-15	1.21e-18	2.3e-17	4.84e-18	9.44e-21	1.21e-18
BE Bound Q_1	1.21e-18	1.13e-14	2.72e-15	1.21e-18	2.3e-17	4.84e-18	9.45e-21	1.21e-18
μ_{Q_1}	1	1	1	1	1	1	1	1
Rel. Res. Q_2	4.17e-19	3.89e-15	9.4e-16	4.17e-19	7.92e-18	1.67e-18	3.26e-21	4.17e-19
BE Bound Q_2	5.58e-19	5.21e-15	1.26e-15	5.58e-19	1.06e-17	2.23e-18	4.36e-21	5.58e-19
μ_{Q_2}	1.34	1.34	1.34	1.34	1.34	1.34	1.34	1.34
Pencil Sep.	0.953	0.953	0.953	0.953	0.953	0.953	0.953	0.953
Ψ_{Q_1}	3.07e+03	3.07e+03	3.07e+03	3.07e+03	3.07e+03	3.07e+03	3.07e+03	3.07e+03
Ψ_{Q_2}	6.65e+03	6.65e+03	6.65e+03	6.65e+03	6.65e+03	6.65e+03	6.65e+03	6.65e+03
Ψ_{Q_3}	1.24e+04	1.24e+04	1.24e+04	1.24e+04	1.24e+04	1.24e+04	1.24e+04	1.24e+04
Φ_{Q_1}	3.07e+03	3.07e+03	3.07e+03	3.07e+03	3.07e+03	3.07e+03	3.07e+03	3.07e+03
Φ_{Q_2}	1.4e+04	1.4e+04	1.4e+04	1.4e+04	1.4e+04	1.4e+04	1.4e+04	1.4e+04
Φ_{Q_3}	2.38e+07	2.38e+07	2.38e+07	2.38e+07	2.38e+07	2.38e+07	2.38e+07	2.38e+07
FE Bound 1 Q_2	1.21e-18	1.13e-14	3.39e-15	1.39e-17	7.13e-17	1.49e-17	2.76e-17	2.76e-17
FE Bound 1 Q_3	3.58e-15	3.2e-11	1.99e-12	3.14e-11	8.68e-15	1.5e-14	2.84e-15	2.38e-15
FE Bound 2 Q_2	3.71e-15	3.46e-11	8.36e-12	3.71e-15	7.05e-14	1.48e-14	2.9e-17	3.71e-15
FE Bound 2 Q_3	3.88e-09	1.44e-05	8.61e-07	1.42e-05	3.95e-09	7.92e-09	3.88e-09	1.69e-10

TABLE 6. Results Q: Standard Calibration

- For Dynare, refer to Adjemian, Bastani, Juillard, Mihoubi, Perendia, Ratto, and Villemot (2011). Dynare under QZ-Based Methods is documented in Villemot (2011) and under Alternatives is the cyclic reduction method. BP (1997) refers to Binder and Pesaran (1997).
- $E[rp]$ is expressed in annual %, $std(\log c_t)$ in quarterly %, and $*$ indicates a backward error less than machine precision, $2^{-52} = 2.2204E-16$.

Measure		QZ-Based Methods				Alternatives		
	Symbolic	Klein (2000)	Sims (2001)	Uhlig (1999)	Dynare QZ	Anderson (2010)	BP (1997)	Dynare CR
$E[rp]$	7.8	7.8	7.8	7.8	7.8	7.8	7.8	7.8
$\Delta E[rp]$		2.31e-08	-1.43e-09	2.27e-08	-3.08e-12	1.54e-11	1.71e-12	1.71e-12
$std(\Delta \log c_t)$	0.566	0.566	0.566	0.566	0.566	0.566	0.566	0.566
$\max(\lvert\Delta[PQ]\rvert)$		3.93e-11	2.36e-12	3.88e-11	7.17e-14	2.13e-14	2.4e-15	1.83e-14
$\max(\lvert\Delta eig\rvert)$		4.12e-11	2.47e-12	4.13e-11	1.32e-14	2.86e-14	3.11e-15	7.66e-15
Rel. Res.	2.33e-17	8.65e-14	5.2e-15	8.55e-14	2.4e-17	4.76e-17	2.33e-17	1.03e-18
BE Bound	6.84e-17	2.42e-13	1.45e-14	2.4e-13	7.03e-17	1.41e-16	6.83e-17	3.22e-18
$\mu_{[PQ]}$	2.93	2.8	2.79	2.8	2.93	2.96	2.93	3.13
Eig. Sep.	0.0127	0.0127	0.0127	0.0127	0.0127	0.0127	0.0127	0.0127
Pencil Sep.	0.302	0.302	0.302	0.302	0.302	0.302	0.302	0.302
$\Psi_{[PQ]}$	1.19e+04	1.19e+04	1.19e+04	1.19e+04	1.19e+04	1.19e+04	1.19e+04	1.19e+04
$\Phi_{[PQ]}$	2.79e+04	2.79e+04	2.79e+04	2.79e+04	2.79e+04	2.79e+04	2.79e+04	2.79e+04
FE Bound 1	2.78e-15	3.27e-11	2.01e-12	3.22e-11	3.66e-14	1.24e-14	2.22e-15	9.4e-15
FE Bound 2	6.51e-13	2.41e-09	1.45e-10	2.38e-09	6.69e-13	1.33e-12	6.5e-13	2.88e-14

TABLE 7. Results $[PQ]$: Standard Calibration

- For Dynare, refer to Adjemian, Bastani, Juillard, Mihoubi, Perendia, Ratto, and Villemot (2011). Dynare under QZ-Based Methods is documented in Villemot (2011) and under Alternatives is the cyclic reduction method. BP (1997) refers to Binder and Pesaran (1997).
- $E[rp]$ is expressed in annual %, $std(\log c_t)$ in quarterly %, and $*$ indicates a backward error less than machine precision, $2^{-52} = 2.2204E-16$.

entrywise absolute errors in P and in the eigenvalues are several orders of magnitude larger than under the baseline calibration. While the relative residuals have increased, the magnitude of these increases relative to the baseline case is inconsistent with the collapse of accuracy in the moments of the endogenous variables or that of P and the eigenvalues. This again casts further doubt on the appropriateness of residual based accuracy measures. Note that the reduction of the pencil separation is several orders of magnitude larger than the reduction in the eigenvalue separation, highlighting that focusing on the eigenvalue separation can be misleading as to the true depredation in the conditioning of the problem, which both measures and all solution methods agree is now ill conditioned. Turning to the forward error bounds, both bounds order the differ methods relative to one another consistently with the relative ordering based on the moments of endogenous variables or the largest entrywise absolute errors in P. Relative to these relative absolute errors that require an exact or symbolic solution and use a different measure than the normwise errors of the forward error bounds that do not require such an exact or symbolic solution, the forward error 1 bound is somewhat more pessimistic than under the baseline calibration. Yet both forward error bounds, 1 and the numerically less demanding 2, clearly provide an alarm that the results, especially of the QZ methods, are likely to be inaccurate.

Figure 2 plots the pseudospectra for the extreme calibration of the matrix quadratic (85) – in blue – and of the QZ algorithm (88) – in red – against the symbolic eigenvalues – in black – for two different sizes of perturbations. In contrast to the results for the standard calibration in figure 2, the finite eigenvalues are all much closer to the unit circle (see the scale on the x-axis) and dispersion away from the exact eigenvalues is visible with perturbations several orders of magnitude smaller. Again, the pseudospectrum of the QZ algorithm bleeds across the unit circle for smaller perturbations than does the matrix quadratic (right panel).

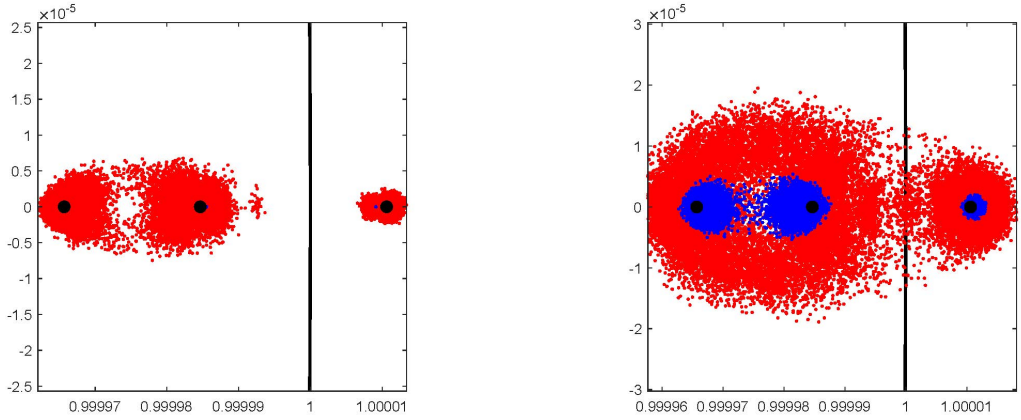
Table 9 provide the results of the extreme calibration for Q.[24] As above the different μ's for Q_1 and Q_2 reflect the difference between taking the dependence of F on A, B, and P into account or not. In general the different methods show relative residuals only a couple of orders of magnitude larger than under the baseline calibration - the results for Uhlig (1999) show that this method manages to solve the linear equation

[24]For FE Bound 2 Q_2, Klein (2000) and then Uhlig (1999) give 0.0010757802 and 0.0010757517 respectively out 8 significant digits.

Measure		QZ-Based Methods				Alternatives		
	Symbolic	Klein (2000)	Sims (2001)	Uhlig (1999)	Dynare QZ	Anderson (2010)	BP (1997)	Dynare CR
$E[rp]$	7.8	7.05	4.75	7.05	6.47	7.8	7.8	7.8
$\Delta E[rp]$		0.754	3.05	0.754	1.33	-8.31e-07	-3.06e-06	1.43e-04
$std(\log c_t)$	0.566	0.529	0.41	0.529	0.491	0.567	0.567	0.567
$\max(\lvert\Delta P\rvert)$		1.01e-06	4.07e-06	1.01e-06	1.74e-06	1.13e-12	4.01e-12	1.87e-10
$\max(\lvert\Delta eig\rvert)$		1.84e-06	6.51e-06	1.84e-06	5.73e-06	1.55e-12	1.57e-11	7.36e-10
Rel. Res.	1.8e-17	3.47e-12	2.07e-11	3.37e-12	2.36e-11	1.85e-17	7.2e-17	2.3e-15
BE Bound	5.08e-17	9.75e-12	5.19e-11	1.04e-11	6.65e-11	5.29e-17	2.03e-16	6.49e-15
μ_P	2.82	2.81	2.5	3.09	2.82	2.86	2.82	2.82
Eig. Sep.	2.82e-05	2.6e-05	2.09e-05	2.6e-05	2.55e-05	2.82e-05	2.82e-05	2.82e-05
Pencil Sep.	1.2e-06	1.28e-06	1.68e-06	1.28e-06	1.47e-06	1.2e-06	1.2e-06	1.2e-06
Ψ_P	2.56e+11	2.42e+11	1.84e+11	2.42e+11	2.11e+11	2.56e+11	2.56e+11	2.56e+11
Φ_P	7.23e+11	6.82e+11	5.17e+11	6.82e+11	5.94e+11	7.23e+11	7.23e+11	7.23e+11
FE Bound 1	1e-09	6.97e-04	0.0029	6-97e-04	0.0011	1.94e-09	4.01e-09	1.28e-07
FE Bound 2	1.3e-05	2.37	10.7	2.3	14	1.34e-05	5.2e-05	0.00167

TABLE 8. Results P: Extreme Calibration

- For Dynare, refer to Adjemian, Bastani, Juillard, Mihoubi, Perendia, Ratto, and Villemot (2011). Dynare under QZ-Based Methods is documented in Villemot (2011) and under Alternatives is the cyclic reduction method. BP (1997) refers to Binder and Pesaran (1997).
- $E[rp]$ is expressed in annual %, $std(\log c_t)$ in quarterly %, and $*$ indicates a backward error less than machine precision, $2^{-52} = 2.2204E-16$.



(A) $5e-11$ (B) $2e-10$

FIGURE 2. Pseudospectrum: Extreme Calibration x-axis: real component, y-axis: imaginary component, large black dots: eigenvalues, black curve: unit circle, small red dots: pseudospectrum QZ companion linearization (88), small blue dots: pseudospectrum matrix quadratic (85)

exactly, conditional on the coefficient matrix $F = AP + B$. An error detection based on residuals would erroneously conclude the solution of Uhlig (1999) for Q and hence the risk premium (which is a function of the square of the entry Q_{cz}) are beyond reproach and, similarly, users of other methods relying on residual errors would likely likewise conclude their results are reliable. The inaccuracies stem from P and are then transmitted to Q as indicated by the condition numbers associated with Q_3 that additionally takes the dependence of P on A, B, and C into account. The importance of this dependence can then be seen in the different forward error bounds. Whereas the bounds for Q_2 do not indicate any especially worrisome inaccuracy (and for Uhlig (1999) these bounds are exactly zero), the bounds for Q_3 correctly indicate potentially catastrophic errors, also for Uhlig (1999) despite the exactly zero residual, and both the relative ordering of the methods and the orders of magnitude of the forward error bound 1 align with the largest entrywise absolute errors in Q that require the availability of a symbolic or closed form solution.

Table 10 contains the results for $[PQ]$ taken jointly.[25] As under the baseline calibration, the combined bounds here take the more pessimistic of the results from P and Q individually. This has the advantage of providing a single error measure and the forward error bound 1 provides a measure of the error in $[PQ]$ of the same order of magnitude as

[25]For FE Bound 2 PQ, Klein (2000) and Uhlig (1999) give 8.2765608e-04 and 8.2763419e-04 respectively out 8 significant digits.

Measure		QZ-Based Methods				Alternatives		
	Symbolic	Klein (2000)	Sims (2001)	Uhlig (1999)	Dynare QZ	Anderson (2010)	BP (1997)	Dynare CR
$\Delta E[rp]$		0.754	3.05	0.754	1.33	-8.31e-07	-3.06e-06	0.000143
$\max(\lvert\Delta Q\rvert)$		0.00127	0.00515	0.00127	0.00224	1.4e-09	5.16e-09	2.41e-07
Rel. Res. Q_1	1.61e-18	4.64e-12	6.21e-11	0	7.41e-13	3.93e-22	8.06e-18	2.63e-16
BE Bound Q_1	1.61e-18	4.64e-12	6.21e-11	0	7.41e-13	3.93e-22	8.06e-18	2.63e-16
μ_{Q_1}	1	1	1		1	1	1	1
Rel. Res. Q_2	4.75e-19	1.37e-12	1.83e-11	0	2.19e-13	1.16e-22	2.38e-18	7.75e-17
BE Bound Q_2	6.62e-19	1.91e-12	2.55e-11	0	3.04e-13	1.62e-22	3.31e-18	1.08e-16
μ_{Q_2}	1.39	1.39	1.39		1.39	1.39	1.39	1.39
Pencil Sep.	0.719	0.719	0.719	0.719	0.719	0.719	0.719	0.719
Ψ_{Q_1}	3.26e+05	3.26e+05	3.26e+05	3.26e+05	3.26e+05	3.26e+05	3.26e+05	3.26e+05
Ψ_{Q_2}	7.93e+05	7.93e+05	7.93e+05	7.93e+05	7.93e+05	7.93e+05	7.93e+05	7.93e+05
Ψ_{Q_3}	3.96e+11	3.73e+11	2.83e+11	3.73e+11	3.25e+11	3.96e+11	3.96e+11	3.96e+11
Φ_{Q_1}	3.26e+05	3.26e+05	3.26e+05	3.26e+05	3.26e+05	3.26e+05	3.26e+05	3.26e+05
Φ_{Q_2}	1.71e+06	1.71e+06	1.71e+06	1.71e+06	1.71e+06	1.71e+06	1.71e+06	1.71e+06
Φ_{Q_3}	6.33e+16	5.98e+16	4.53e+16	5.98e+16	5.2e+16	6.33e+16	6.33e+16	6.33e+16
FE Bound 1 Q_2	1.61e-18	4.64e-12	6.21e-11	0	7.42e-13	9.22e-17	9.25e-17	6.71e-14
FE Bound 1 Q_3	1.55e-09	0.00108	0.00447	0.00108	0.0017	2.99e-09	6.18e-09	1.98e-07
FE Bound 2 Q_2	5.25e-13	1.51e-06	2.02e-05	0	2.41e-07	1.28e-16	2.62e-12	8.56e-11
FE Bound 2 Q_3	7.88	1.43e+06	6.49e+06	1.39e+06	8.48e+06	8.08	31.5	1.01e+03

TABLE 9. Results Q: Extreme Calibration

- For Dynare, refer to Adjemian, Bastani, Juillard, Mihoubi, Perendia, Ratto, and Villemot (2011). Dynare under QZ-Based Methods is documented in Villemot (2011) and under Alternatives is the cyclic reduction method. BP (1997) refers to Binder and Pesaran (1997).
- $E[rp]$ is expressed in annual %, $std(\log c_t)$ in quarterly %, and $*$ indicates a backward error less than machine precision, $2^{-52} = 2.2204E-16$.

the entrywise maximal absolute error relative to the symbolic solution. The preceding analysis of P and Q individually was useful diagnostically attributing the error to an ill conditioned problem in P which led to high forward errors in P and, through its dependance on P, then also in Q. The errors in the moments in the first and third rows are ordered relatively among the different methods consistent with the forward errors, but have fewer significant digits as both moments are second moments, i.e., involved products of elements of $[PQ]$ combining the errors of the elements individually, see Higham (2002). In sum, regardless of the method used, the resulting condition number indicates the problem at this calibration is ill conditioned; meaning that small backward errors, which here are several times larger than the relative residuals, can be potentiated into very large forward errors; and the resulting forward error bounds indicate a catastrophic loss of accuracy for the QZ methods and numerically large but economically insignificant (in terms of the first several significant digits for first and second order terms) errors to be concerned about. These diagnostic indicators are particularly useful as none of the methods produced any warning or error that would have alerted the user to the loss of accuracy - resulting in zero significant digits in the risk premium for two of the QZ methods.

Table 11 contains a summary of results from additional alternate calibrations (see the appendix, Table 34), in all calibrations, the parameters are chosen to match the annual equity premium of 7.8 and the quarterly standard deviation of consumption growth of 0.566%. Calibrations I and II are alternative "standard" calibrations, holding all parameters apart from h, σ and ω constant. Calibration I has a higher curvature in the utility function, σ, and a lower degree of habit formation, h, and calibration II vice versa than in the standard calibration above. As in the standard calibration, these first two calibrations are similarly conditioned, with II being less well conditioned by about an order of magnitude, which corresponds with the forward errors being similar though about an order of magnitude worse for calibration II than I and the differences (recall this moment is second order in the underlying solution matrices) in the expected risk premium about two orders of magnitude worse for II than I. All methods successfully recover the equity premium up to economically irrelevant numerical errors.

Calibration III is similar to the extreme calibration above, but with a slightly reduced degree of habit formation, h, and discount factor, β, compensated by an increased curvature in the utility function, σ. The pencil separation drops roughly six orders of magnitude

Measure		QZ-Based Methods				Alternatives		
	Symbolic	Klein (2000)	Sims (2001)	Uhlig (1999)	Dynare QZ	Anderson (2010)	BP (1997)	Dynare CR
$E[rp]$	7.8	7.05	4.75	7.05	6.47	7.8	7.8	7.8
$\Delta E[rp]$		0.754	3.05	0.754	1.33	-8.31e-07	-3.06e-06	1.43e-04
$std(\Delta \log c_t)$	0.566	0.529	0.41	0.529	0.491	0.567	0.567	0.567
$\max(\lvert\Delta[PQ]\rvert)$		0.00127	0.00515	0.00127	0.00224	1.4e-09	5.16e-09	2.41e-07
$\max(\lvert\Delta eig\rvert)$		1.84e-06	6.51e-06	1.84e-06	5.73e-06	1.55e-12	1.57e-11	7.36e-10
Rel. Res.	1.52e-17	3.01e-12	1.97e-11	2.85e-12	1.99e-11	1.56e-17	6.07e-17	1.94e-15
BE Bound	5.11e-17	9.79e-12	5.22e-11	1.06e-11	6.65e-11	5.29e-17	2.03e-16	6.49e-15
$\mu_{[PQ]}$	3.36	3.26	2.64	3.71	3.34	3.39	3.34	3.34
Eig. Sep.	2.82e-05	2.6e-05	2.09e-05	2.6e-05	2.55e-05	2.82e-05	2.82e-05	2.82e-05
Pencil Sep.	8.51e-07	9.02e-07	1.19e-06	9.02e-07	1.04e-06	8.51e-07	8.51e-07	8.51e-07
$\Psi_{[PQ]}$	3.04e+11	2.87e+11	2.18e+11	2.87e+11	2.5e+11	3.04e+11	3.04e+11	3.04e+11
$\Phi_{[PQ]}$	7.19e+11	6.78e+11	5.15e+11	6.78e+11	5.9e+11	7.19e+11	7.19e+11	7.19e+11
FE Bound 1	1.19e-09	8.28e-04	0.00344	8.28e-04	0.0013	2.3e-09	4.76e-09	1.52e-07
FE Bound 2	1.09e-05	2.04	10.2	1.93	11.8	1.12e-05	4.37e-05	0.0014

TABLE 10. Results $[PQ]$: Extreme Calibration

- For Dynare, refer to Adjemian, Bastani, Juillard, Mihoubi, Perendia, Ratto, and Villemot (2011). Dynare under QZ-Based Methods is documented in Villemot (2011) and under Alternatives is the cyclic reduction method. BP (1997) refers to Binder and Pesaran (1997).
- $E[rp]$ is expressed in annual %, $std(\log c_t)$ in quarterly %, and $*$ indicates a backward error less than machine precision, $2^{-52} = 2.2204E-16$.

relative to the first two calibrations and, accordingly, the condition numbers increase by roughly six orders of magnitude. The QZ methods demonstrate significant deviations in their predicted equity premia as above, though now some methods over and some methods under predict the premium. Calibrations IV-VI provide further examples of potentially arbitrary results from QZ methods. Note that the relative ordering of the accuracy of the QZ methods changes, likewise among the alternative methods. Hence, although one can conclude that the alternative methods outperform the QZ methods, there is not a uniformly better performing method among the two categories.

However, the forward error bounds systematically align with the relative ordering of the accuracy of the methods' expected risk premia. That is, for Dynare QZ, one could conclude that the results for calibrations IV-VI were likely to be accurate for economic purposes, whereas this could not be said for calibration II. Additionally, the pencil separation increases between calibrations III and V and again from V to IV, corresponding to decrease in the condition numbers from III to V and again from V to IV - the eigenvalue separation moves in the opposite direction, again calling its use into question. Again, none of the algorithms produced any warning as to the potential inaccuracy of their solutions.

The backward and forward error analysis of the previous sections has been shown to successfully diagnose numerical errors of economic significance reliably and predictably for this simple macro finance model with an available symbolic solution corroborating this positive conclusion. These diagnostics are especially useful as none of the methods from the literature used here produced any sort of warning for any of the different calibrations, even those with a catastrophic loss (i.e., all significant digits) in accuracy.

4.2. **Jermann's (1998) DSGE Macro-Finance Model.** Turning now to a more economically relevant but still small enough scale model for detailed analysis, I will now apply the backward forward error analysis of the previous sections and methods in the literature when a symbolic solution is not available. Jermann's (1998) macro finance model provides such a model and is able to successfully replicate key finance variables, such as the average equity premium, in a production based asset pricing framework - that is, with endogenous production and consumption. THis model can also be viewed as an extension of the model of the previous section that replaces the external habit with an internal one, adds friction to capital accumulation via adjustment costs, and adds constant growth to the environment. The first two changes increase the dimensionality of the model, precluding the use of a symbolic solution to solve the linearized model and

			QZ-Based Methods				Alternatives		
Calibration	Separation	Condition	Klein (2000)	Sims (2001)	Uhlig (1999)	Dynare QZ	Anderson (2010)	BP (1997)	Dynare CR
I	0.011	2.43e+04	-9.41e-14	4.85e-11	-3.62e-12	6.49e-10	-5.03e-13	5.42e-13	1.24e-12
	0.116	5.14e+04	1.39e-15	2.6e-13	1.64e-14	2.26e-12	4.59e-15	7.93e-16	1.76e-15
II	0.0106	1.84e+05	3.09e-10	1.49e-09	2.97e-10	-5.61e-10	-1.72e-12	1.02e-11	1.52e-11
	0.421	4.86e+05	2.18e-13	7.5e-12	2.25e-13	3.05e-12	4.89e-15	1.77e-16	5.58e-15
III	1.08e-04	4.6e+12	-0.505	-0.373	-0.505	1.3	3.02e-07	3.45e-07	9.96e-05
	1.24e-07	1.08e+13	0.00394	0.00296	0.00394	0.0124	1.34e-10	7.44e-10	8.14e-07
IV	2.24e-05	2.14e+09	-4.9	-0.176	-4.9	4.61e-05	-1.69e-05	-4.38e-06	-4.59e-06
	7.62e-05	5e+09	3.08e-04	3.16e-05	3.08e-04	2.84e-09	1.19e-09	5.34e-10	1.92e-10
V	3.69e-05	1.73e+10	0.0631	-3.91	0.0567	-9.19e-06	-5.46e-06	-4.5e-06	-2.23e-05
	1.56e-05	3.81e+10	1.45e-05	9.45e-04	1.29e-05	1.29e-09	8.65e-10	3.06e-10	4.54e-09
VI	3.34e-05	1.84e+10	0.812	-1.84	0.813	2.28e-05	4.24e-06	-5.78e-06	2.92e-05
	1.2e-05	4.3e+10	2.05e-04	4.54e-04	2.05e-04	5.98e-09	1.47e-09	4.85e-10	7.39e-09

TABLE 11. Results for additional calibrations I-VI (see the appendix), satisfying $E[rp] = 7.8$ and $std(\log c_t) = 0.566\%$ symbolically. For the column "Separation", the first row is Eig. Sep., the second Pencil Sep., for the column "Condition" the first column is $\Psi_{[PQ]}$, the secon $\Phi_{[PQ]}$. For all other columns, the first row is $\Delta E[rp]$, the second $FE_{PQ,1}$.

For Dynare, refer to Adjemian, Bastani, Juillard, Mihoubi, Perendia, Ratto, and Villemot (2011). Dynare under QZ-Based Methods is documented in Villemot (2011) and under Alternatives is the cyclic reduction method. BP (1997) refers to Binder and Pesaran (1997).

$*$ indicates a backward error less than machine precision, $2^{-52} = 2.2204E-16$.

necessitating the use of numerical diagnostics. I show that the baseline calibration of the model admits a well conditioned solution with reasonable backward errors by all the numerical methods from the literature examined here, albeit it again with QZ based methods performing relatively worse than the alternatives. However, nearby parameterizations produce models whose solution differs depending on the numerical algorithm from the literature chosen, again with moments in macroeconomic variables and asset pricing predictions that differ to an economically relevant degree. Furthermore, I show that tautological redefinition of the model's equation can result in different solutions from the same numerical method or render a particular method unable to solve the rearranged model.

In contrast to the model of the previous section, the habit stock, X_t, is internal - households internalize the effect of consumption today on the habit they will face tomorrow, altering marginal utility from consumption, λ_t in the pricing kernel $m_{t+1} \equiv \beta\lambda_{t+1}/\lambda_t$ as follows

$$\lambda_t \equiv \frac{\partial u(c_t, X_t) + \beta E_t[u(c_{t+1}, X_{t+1})]}{\partial c_t} \tag{112}$$

If habit formation is external, as in the previous section, this is simply $u_c(c_t, X_t)$, when the habit is internalized and is a function of the previous period's consumption, $X_t \equiv X(c_{t-1})$, this becomes

$$\lambda_t \equiv \frac{\partial u(c_t, X_t) + \beta E_t[u(c_{t+1}, X_{t+1})]}{\partial c_t} = u_c(c_t, X(c_{t-1})) + \beta E_t[u_X(c_{t+1}, X(c_t))X_c(c_t)] \tag{113}$$

with the period utility function $u(\cdot) = (\cdot)^{1-\tau}/(1-\tau)$ governed by the curvature τ and one-period linear habit $\cdot_t = c_t - bc_{t-1}$ by the degree of habit formation b.

Habit formation is not enough to match the equity premium in this model, households need not only to care about volatile consumption streams, but they need to be prevented from doing anything about it, as Jermann (1998) points out. Hence capital accumulation now faces adjustment costs

$$k_t = (1-\delta)k_{t-1} + \phi\left(\frac{i_t}{k_{t-1}}\right)k_{t-1} \tag{114}$$

where the capital adjustment cost function is given as

$$\phi\left(\frac{i_t}{k_{t-1}}\right) = \frac{\tilde{b}}{1-\xi}\left(\frac{i_t}{k_{t-1}}\right)^{1-\xi} + \tilde{c} \tag{115}$$

with $\tilde{b}$ and $\tilde{c}$ set such that the steady state is identical to the case without adjustment costs and $1/\xi$ is the elasticity of the investment-capital ratio with respect to Tobin's q.

	τ	b	$\beta^* = \beta a_{bar}^{1-\tau}$	ξ	σ_Z
Baseline	7.92	0.74	0.99	3.75	0.999%
Alternative	29.96	0.66	0.94	0.76	0.997%

Common Parameters

α	a_{bar}	δ	ρ_Z	N_{ss}	SS_{adjc}
0.36	0.005	0.025	0.99	1	1

TABLE 12. Calibrations for Jermann's (1998) DSGE Macro-Finance Model

The parameterizations can be found in table 12. The baseline calibration is very close to the calibration given by Jermann (1998) with only slight adjustments made to match the set of moments in the first six lines of table 13, the average risk premium, the risk free rate and the standard deviations of output, investment and consumption growth. The alternative parameterization is merely a close-by calibration with an increase in the curvature of the utility kernel - increasing households' unconditional sensitivity to volatile consumption streams - but increasing the elasticity in the adjustment costs - making them, however, more able to respond to this volatility. As there is no closed form or symbolic solution available, the consequences for the moments must be determined numerically and as, can be seen in 16, the different methods from the literature disagree.

Beginning with the results for P under the standard calibration in table 13, all of the methods agree on the macro and finance moments in the first six lines. The second line contains the difference in the risk premium calculated by the different methods to that predicted by the cyclic reduction method of Dynare - remember, we do not have a symbolic solution available for this model. I chose this solution as the "reference" solution simply because its forward error bound 1 was the lowest. This, of course, means we do not have exact or symbolic comparisons for P or the eigenvalues of the quadratic problem to use as a reference and must rely on a numerical nalysis of the quality of the solutions. Examining the relative residuals versus the backward error bounds, the potential for arbitrary differences between the two particularly across methods is apparent, while the residual of Anderson (2010) is an order of magnitude lower, the backward error is comparable to that of Uhlig (1999). As laid out above, the residual (and hence residual based methods) are insufficient or at least problematic as an error diagnostic. Note that all methods agree on the eigenvalue and pencil separation and place the eigenvalue separation at a value twice as high as for the baseline calibration of the previous model (see table 5). This is far from the case of the pencil separation, which is four orders of

Measure		QZ-Based Methods				Alternatives		
	Data	Klein (2000)	Sims (2001)	Uhlig (1999)	Dynare QZ	Anderson (2010)	BP (1997)	Dynare CR
$E[rp]$	6.18	6.18	6.18	6.18	6.18	6.18	6.18	6.18
$\Delta E[rp]$		-6.69e-12	2.36e-11	1.52e-12	7.47e-13	2.09e-13	-2.4e-14	-
$E\left[R_t^f\right]$	0.8	0.8	0.8	0.8	0.8	0.8	0.8	0.8
$std(\Delta \log y_t)$	0.01	0.01	0.01	0.01	0.01	0.01	0.01	0.01
$std(\Delta \log c_t)$	0.51	0.51	0.51	0.51	0.51	0.51	0.51	0.51
$std(\Delta \log i_t)$	2.65	2.65	2.65	2.65	2.65	2.65	2.65	2.65
Rel. Res.		3.92e-16	1.34e-16	1.43e-17	3.95e-15	3.26e-18	9.15e-19	2.33e-18
BE Bound		1.42e-14	2.27e-15	1.34e-16	1.25e-14	1.41e-16	2.91e-17	3.06e-17
μ_P		36.2	17	9.32	3.16	43.3	31.8	13.1
Eig. Sep.		0.0224	0.0224	0.0224	0.0224	0.0224	0.0224	0.0224
Pencil Sep.		1.36e-05	1.36e-05	1.36e-05	1.36e-05	1.36e-05	1.36e-05	1.36e-05
Ψ_P		9.4e+05	9.4e+05	9.4e+05	9.4e+05	9.4e+05	9.4e+05	9.4e+05
Φ_P		1.04e+08	1.04e+08	1.04e+08	1.04e+08	1.04e+08	1.04e+08	1.04e+08
FE Bound 1		1.38e-12	5.38e-12	3.47e-13	2.6e-13	7.04e-15	5.51e-15	2.23e-15
FE Bound 2		4.08e-08	1.39e-08	1.49e-09	4.11e-07	3.4e-10	9.53e-11	2.43e-10

TABLE 13. Results P: Baseline Calibration

- For Dynare, refer to Adjemian, Bastani, Juillard, Mihoubi, Perendia, Ratto, and Villemot (2011). Dynare under QZ-Based Methods is documented in Villemot (2011) and under Alternatives is the cyclic reduction method. BP (1997) refers to Binder and Pesaran (1997).
- $E[rp]$ is expressed in annual %, $std(\log c_t)$ in quarterly %, and $*$ indicates a backward error less than machine precision, $2^{-52} = 2.2204E-16$.

magnitude smaller than in the previous model. This is consistent with the condition number for this model's baseline calibration being up to four orders of magnitude higher than those associated with the model of the previous section. That is, this model at its baseline calibration is far less well conditioned than the previous model, a warning that would be missed by methods that focus on the eigenvalue separation instead of the pencil separation. Nonetheless, the economic consequences are irrelevant here, as one can also surmise from the forward error bounds, though note that relative ordering of the forward bounds align with the relative ordering of the deviations of the equity premium from that of Dynare's cyclic reduction method, chosen as the reference in the absence of a closed form or symbolic solution here.

The results for the shock impact matrix Q under the standard calibration are contained in table 14. As in the previous model, the relationship between the residuals and backward errors for Q_1 and Q_2 shows again that taking the underlying dependencies of F into account is important to understand the differences between these two measures - in the absence of a measure for Q_3 (due to its nonlinear nature), the dependence on the dependencies of P in the backward errors will only be uncovered in the joint assessment of $[PQ]$. For the condition numbers, however, we see that all methods agree on the values of these numbers and that the condition numbers deteriorate as dependencies of F on P and then those of P on A, B, and C are included. The large jump in Φ_{Q_3} is a consequence of the small pencil separation and large condition number Φ_P involved in solving for P. However, the economic consequences are in sum inconsequential as indicated by the forward errors bounds at the end of the table, with the loose bound for Q_3 - the forward error bound 2 - being very pessimistic for the same reasons that drive the large jump in Φ_{Q_3}.

Table 15 continues the results for the joint measure $[PQ]$ under the standard calibration. The unreliability of the residuals is highlighted here again with the backward errors being roughly one order of magnitude higher than the relative residuals for Sims (2001) but nearly three orders of magnitude for Binder and Pesaran (1997). As for P, the condition number is significantly larger for this model than the simple habit model of the previous section, which would be entirely missed by focussing on the eigenvalue separation or similar measure to assess the quality of a solution. Despite the poorer conditioning, the solutions do not suffer from numerical errors of economic consequence as indicated by the forward errors. Note this conclusion is suggested by the agreement among the different

Measure	QZ-Based Methods				Alternatives		
	Klein (2000)	Sims (2001)	Uhlig (1999)	Dynare QZ	Anderson (2010)	BP (1997)	Dynare CR
$\Delta E[rp]$	-6.69e-12	2.36e-11	1.52e-12	7.47e-13	2.09e-13	-2.4e-14	-
Rel. Res. Q_1	2.76e-16	2.67e-16	5.79e-19	4.98e-15	2.1e-19	4.64e-18	3.12e-19
BE Bound Q_1	2.76e-16	2.68e-16	5.79e-19	4.98e-15	2.1e-19	4.64e-18	3.12e-19
μ_{Q_1}	1	1	1	1	1	1	1
Rel. Res. Q_2	1.45e-16	1.41e-16	3.05e-19	2.62e-15	1.1e-19	2.44e-18	1.64e-19
BE Bound Q_2	1.88e-16	1.82e-16	3.95e-19	3.4e-15	1.43e-19	3.16e-18	2.13e-19
μ_{Q_2}	1.29	1.29	1.29	1.29	1.29	1.29	1.29
Pencil Sep.	0.00981	0.00981	0.00981	0.00981	0.00981	0.00981	0.00981
Ψ_{Q_1}	7.66e+04	7.66e+04	7.66e+04	7.66e+04	7.66e+04	7.66e+04	7.66e+04
Ψ_{Q_2}	1.12e+05	1.12e+05	1.12e+05	1.12e+05	1.12e+05	1.12e+05	1.12e+05
Ψ_{Q_3}	1.02e+06	1.02e+06	1.02e+06	1.02e+06	1.02e+06	1.02e+06	1.02e+06
Φ_{Q_1}	7.66e+04	7.66e+04	7.66e+04	7.66e+04	7.66e+04	7.66e+04	7.66e+04
Φ_{Q_2}	5.34e+06	5.34e+06	5.34e+06	5.34e+06	5.34e+06	5.34e+06	5.34e+06
Φ_{Q_3}	1.79e+12	1.79e+12	1.79e+12	1.79e+12	1.79e+12	1.79e+12	1.79e+12
FE Bound 1 Q_2	9.01e-14	4.52e-14	3.8e-17	3.11e-13	6.92e-16	1.67e-16	2.94e-16
FE Bound 1 Q_3	1.5e-12	5.54e-12	3.61e-13	4.07e-13	8.3e-16	5.88e-15	1.83e-15
FE Bound 2 Q_2	2.12e-11	2.05e-11	4.44e-14	3.82e-10	1.61e-14	3.56e-13	2.39e-14
FE Bound 2 Q_3	0.212	0.0725	0.00776	2.14	0.00177	0.000496	0.00126

TABLE 14. Results Q: Baseline Calibration

- For Dynare, refer to Adjemian, Bastani, Juillard, Mihoubi, Perendia, Ratto, and Villemot (2011). Dynare under QZ-Based Methods is documented in Villemot (2011) and under Alternatives is the cyclic reduction method. BP (1997) refers to Binder and Pesaran (1997).
- $E[rp]$ is expressed in annual %, $std(\log c_t)$ in quarterly %, and $*$ indicates a backward error less than machine precision, $2^{-52} = 2.2204E-16$.

Measure		QZ-Based Methods				Alternatives		
	Data	Klein (2000)	Sims (2001)	Uhlig (1999)	Dynare QZ	Anderson (2010)	BP (1997)	Dynare CR
$E[rp]$	6.18	6.18	6.18	6.18	6.18	6.18	6.18	6.18
$\Delta E[rp]$		-6.69e-12	2.36e-11	1.52e-12	7.47e-13	2.09e-13	-2.4e-14	-
$E\left[R_t^f\right]$	0.8	0.8	0.8	0.8	0.8	0.8	0.8	0.8
$std(\log y_t)$	0.01	0.01	0.01	0.01	0.01	0.01	0.01	0.01
$std(\log c_t)$	0.51	0.51	0.51	0.51	0.51	0.51	0.51	0.51
$std(\log i_t)$	2.65	2.65	2.65	2.65	2.65	2.65	2.65	2.65
Rel. Res.		3.92e-16	1.34e-16	1.43e-17	3.95e-15	3.26e-18	9.15e-19	2.33e-18
BE Bound		4.96e-14	2.29e-15	3.46e-15	1.78e-12	7e-16	5.57e-16	7.38e-16
μ_P		127	17.1	241	451	215	609	316
Eig. Sep.		0.0224	0.0224	0.0224	0.0224	0.0224	0.0224	0.0224
Pencil Sep.		1.36e-05	1.36e-05	1.36e-05	1.36e-05	1.36e-05	1.36e-05	1.36e-05
Ψ_P		9.4e+05	9.4e+05	9.4e+05	9.4e+05	9.4e+05	9.4e+05	9.4e+05
Φ_P		1.04e+08	1.04e+08	1.04e+08	1.04e+08	1.04e+08	1.04e+08	1.04e+08
FE Bound 1		1.38e-12	5.38e-12	3.47e-13	2.6e-13	7.04e-15	5.51e-15	2.23e-15
FE Bound 2		4.08e-08	1.39e-08	1.49e-09	4.11e-07	3.4e-10	9.53e-11	2.43e-10

TABLE 15. Results PQ: Baseline Calibration

- For Dynare, refer to Adjemian, Bastani, Juillard, Mihoubi, Perendia, Ratto, and Villemot (2011). Dynare under QZ-Based Methods is documented in Villemot (2011) and under Alternatives is the cyclic reduction method. BP (1997) refers to Binder and Pesaran (1997).
- $E[rp]$ is expressed in annual %, $std(\log c_t)$ in quarterly %, and $*$ indicates a backward error less than machine precision, $2^{-52} = 2.2204E-16$.

solution methods, but are confirmed individually by the forward errors. That is, there is no need to recalculated the solution with different methods as a user with forward error 1 bounds on the order of $e-12$ or lower can be confident that any numerical instability are inconsequential for the economic analysis.

Now turning to the alternative calibration and the results for P in table 16, there is disagreement on the moments as can be seen directly in the first row. Whereas Dynare's QZ method calculates the risk premium as below the target, the remaining methods suggest the risk premium has been overshoot at this calibration. Klein (2000) suggests a much higher overshot than the other methods and Sims (2001) suggests the risk free rate is much higher above its target. In the absence of a closed form or symbolic solution, it is unclear which method's moments we should consider correct. The consequences of this uncertainty also extends to estimates: while the results would suggest that the curvature in the utility function or habuit formation should be increased for some methods, the suggestion goes in the opposite direction for others. The danger is even greater: none of the methods here gave any warning or indication that their results might be problematic - practitioners in this situation would be entirely unaware of these discrepancies.

The differences between the relative residuals and the backward errors cannot be overlooked here with the growth or amplification factor ranging from 3.48 to 6.41e+06 - relying on relative residuals would lead to undue confidence in some of the results and in particular would grossly overstate their relative accuracies. The eigenvalue separation has improved from the baseline calibration and users relying on this measure to diagnose numerical inaccuracies would be mislead into being confident about their solution. The pencil separation shows a collapse of the distance between the stable and unstable pencils which then leads to very large condition numbers. Regardless of the method being used, the measures of this paper would warn the practitioner that the model at this calibration is very ill conditioned. The forward error bounds 1 show that the solution by Dynare QZ is least accurate, followed by Klein (2000) and then the remaining QZ methods - this is of course corroborated by comparing the moments with those of the other methods, in particular the alternative, non QZ methods whose forward error bounds 1 indicate that their results are reliable. The forward error bounds 2 are overly pessimistic here which follows directly from the small pencil separation and gives a worst case measure that neglects the structure of the solution to the problem. Hence, practitioners of every methods would be warned of the ill conditioning of the problem and QZ users could further

Measure		QZ-Based Methods				Alternatives		
	Data	Klein (2000)	Sims (2001)	Uhlig (1999)	Dynare QZ	Anderson (2010)	BP (1997)	Dynare CR
$E[rp]$	6.18	6.26	6.23	6.23	6.09	6.23	6.23	6.23
$\Delta E[rp]$		-0.0278	0.00174	9.06e-04	0.137	-	5.84e-10	3.97e-10
$E\left[R_t^f\right]$	0.8	0.823	0.899	0.824	0.696	0.824	0.824	0.824
$std(\log y_t)$	0.01	0.01	0.01	0.01	0.01	0.01	0.01	0.01
$std(\log c_t)$	0.51	0.517	0.519	0.519	0.584	0.519	0.519	0.519
$std(\log i_t)$	2.65	6.03	6.01	6.01	5.98	6.01	6.01	6.01
Rel. Res.		1.02e-10	1.72e-12	1.73e-12	1.85e-09	1.25e-23	1.78e-23	1.3e-22
BE Bound		3.37e-07	7.42e-11	1.62e-11	6.42e-09	5.26e-17	1.14e-16	1.6e-17
μ_P		3.31e+03	43.2	9.36	3.48	4.2e+06	6.41e+06	1.23e+05
Eig. Sep.		0.0765	0.0762	0.0765	0.0753	0.0764	0.0764	0.0764
Pencil Sep.		1.39e-15	1.39e-15	1.39e-15	1.4e-15	1.39e-15	1.39e-15	1.39e-15
Ψ_P		1.05e+05	1.06e+05	1.06e+05	1.22e+05	1.06e+05	1.06e+05	1.06e+05
Φ_P		7.17e+22	7.16e+22	7.15e+22	7.09e+22	7.16e+22	7.16e+22	7.16e+22
FE Bound 1		0.0101	1.71e-04	1.72e-04	0.183	1.24e-15	1.77e-15	1.29e-14
FE Bound 2		7.28e+12	1.23e+11	1.24e+11	1.31e+14	0.897	1.28	9.32

TABLE 16. Results P: Alternative Calibration

- For Dynare, refer to Adjemian, Bastani, Juillard, Mihoubi, Perendia, Ratto, and Villemot (2011). Dynare under QZ-Based Methods is documented in Villemot (2011) and under Alternatives is the cyclic reduction method. BP (1997) refers to Binder and Pesaran (1997).
- $E[rp]$ is expressed in annual %, $std(\log c_t)$ in quarterly %, and $*$ indicates a backward error less than machine precision, $2^{-52} = 2.2204E-16$.

be warned of the expected inaccuracy of their solution, warnings that again none of the methods from the literature provided here.

The results for the shock impact matrix Q under the alternative calibration can be found in table 17. Again note the large differences between the relative residuals and the accuracy of the moments - measured here by the difference to the risk premium predicted by Anderson (2010), the method with the lowest forward error 1. The residuals would not only suggest a reliability of the results from the individual methods, but also a relative order - with Uhlig's (1999) QZ method being tied as the most accurate. Access to condition numbers again indicate that this problem is ill conditioned with the looser bound deteriorating substantially as the dependencies on P and its dependencies are incorporated going to Φ_{Q_2} and then Φ_{Q_3}, consistent with the ill conditioning of the problem in P that serves to exacerbate the ill condition in the problem here in Q even when taking $F = AP+B$ at face value. The importance of incorporating these dependencies is highlighted by the differences between the forward error bounds for Q_2 and Q_3 using Uhlig (1999) and despite the ill conditioned nature of the problem, the tight forward error bounds 1 for Q_3 indicate the reliability of the solutions provided by the non QZ methods.

Table 18 continues the results for the joint measure $[PQ]$ under the alternative calibration. The diagnostics provided by examining P and Q jointly summarize the results from Q and P individually. Practitioners relying on residual or eigenvalue separation based measures or warnings of numerical instabilities would miss the problems associated with the model at this calibration: the backward errors differ by a factor of over 1e07 from the relative residuals for most methods and the pencil separation is smaller than the eigenvalue separation by a factor of 1e13. The condition numbers confirm that the problem is ill conditioned regardless of the method employed and the forward error bounds 1 show that the solution by Dynare QZ is the least accurate, followed by Klein (2000) and then the remaining QZ methods, with the alternative, non QZ methods displaying forward error bounds 1 that indicate the reliability of their results. Again the methods I propose would warn users of all methods of the ill conditioning of the problem and QZ users would be further warned of the expected inaccuracy of their solution, warnings that none of the methods from the literature provided here.

Just as different methods that are theoretically identical can have different numerical consequences, so too can different formulations of a model that are theoretically identical have different numerical consequences. Consider the application of the Lucas asset

Measure	QZ-Based Methods				Alternatives		
	Klein (2000)	Sims (2001)	Uhlig (1999)	Dynare QZ	Anderson (2010)	BP (1997)	Dynare CR
$\Delta E[rp]$	-0.0278	0.00174	0.000906	0.137	-	5.84e-10	3.97e-10
Rel. Res. Q_1	3.72e-11	3.21e-12	1.34e-29	2.42e-09	3.4e-24	3.4e-24	1.42e-29
BE Bound Q_1	3.72e-11	3.21e-12	1.34e-29	2.42e-09	3.4e-24	3.4e-24	1.42e-29
μ_{Q_1}	1	1	1	1	1	1	1
Rel. Res. Q_2	1.99e-11	1.73e-12	7.22e-30	1.32e-09	1.83e-24	1.83e-24	7.61e-30
BE Bound Q_2	2.61e-11	2.26e-12	9.45e-30	1.73e-09	2.4e-24	2.4e-24	9.97e-30
μ_{Q_2}	1.31	1.31	1.31	1.31	1.31	1.31	1.31
Pencil Sep.	1.19e-07	1.2e-07	1.19e-07	1.2e-07	1.19e-07	1.19e-07	1.19e-07
Ψ_{Q_1}	4.88e+08	4.88e+08	4.88e+08	5.01e+08	4.88e+08	4.88e+08	4.88e+08
Ψ_{Q_2}	6.94e+08	6.94e+08	6.94e+08	6.98e+08	6.94e+08	6.94e+08	6.94e+08
Ψ_{Q_3}	6.93e+08	6.93e+08	6.93e+08	6.98e+08	6.93e+08	6.93e+08	6.93e+08
Φ_{Q_1}	4.47e+14	4.47e+14	4.48e+14	4.54e+14	4.48e+14	4.48e+14	4.48e+14
Φ_{Q_2}	2.18e+21	2.17e+21	2.18e+21	2.18e+21	2.18e+21	2.18e+21	2.18e+21
Φ_{Q_3}	6.07e+36	6.05e+36	6.05e+36	5.99e+36	6.06e+36	6.06e+36	6.06e+36
FE Bound 1 Q_2	0.000222	4.86e-05	2.9e-16	0.00646	3.39e-16	4.29e-16	3.24e-16
FE Bound 1 Q_3	0.00192	0.000243	0.000855	0.00167	1.89e-16	2.77e-16	3.52e-14
FE Bound 2 Q_2	1.66e+04	1.44e+03	6.01e-15	1.1e+06	1.52e-09	1.52e-09	6.34e-15
FE Bound 2 Q_3	1.59e+34	2.67e+32	2.69e+32	2.85e+35	1.95e+21	2.78e+21	2.03e+22

TABLE 17. Results Q: Alternative Calibration

- For Dynare, refer to Adjemian, Bastani, Juillard, Mihoubi, Perendia, Ratto, and Villemot (2011). Dynare under QZ-Based Methods is documented in Villemot (2011) and under Alternatives is the cyclic reduction method. BP (1997) refers to Binder and Pesaran (1997).
- $E[rp]$ is expressed in annual %, $std(\log c_t)$ in quarterly %, and $*$ indicates a backward error less than machine precision, $2^{-52} = 2.2204E-16$.

Measure		QZ-Based Methods				Alternatives		
	Data	Klein (2000)	Sims (2001)	Uhlig (1999)	Dynare QZ	Anderson (2010)	BP (1997)	Dynare CR
$E[rp]$	6.18	6.26	6.23	6.23	6.09	6.23	6.23	6.23
$\Delta E[rp]$		-0.0278	0.00174	0.000906	0.137	-	5.84e-10	3.97e-10
$E\left[R_t^f\right]$	0.8	0.823	0.899	0.824	0.696	0.824	0.824	0.824
$std(\log y_t)$	0.01	0.01	0.01	0.01	0.01	0.01	0.01	0.01
$std(\log c_t)$	0.51	0.517	0.519	0.519	0.584	0.519	0.519	0.519
$std(\log i_t)$	2.65	6.03	6.01	6.01	5.98	6.01	6.01	6.01
Rel. Res.		1.02e-10	1.72e-12	1.73e-12	1.85e-09	1.25e-23	1.78e-23	1.3e-22
BE Bound		0.00294	7.45e-11	4.21e-05	0.0459	2.84e-16	3.97e-16	3.21e-15
μ_P		2.89e+07	43.4	2.44e+07	2.49e+07	2.26e+07	2.23e+07	2.46e+07
Eig. Sep.		0.0765	0.0762	0.0765	0.0753	0.0764	0.0764	0.0764
Pencil Sep.		1.39e-15	1.39e-15	1.39e-15	1.4e-15	1.39e-15	1.39e-15	1.39e-15
Ψ_P		1.05e+05	1.06e+05	1.06e+05	1.22e+05	1.06e+05	1.06e+05	1.06e+05
Φ_P		7.17e+22	7.16e+22	7.15e+22	7.09e+22	7.16e+22	7.16e+22	7.16e+22
FE Bound 1		0.0101	1.71e-04	1.72e-04	0.183	1.24e-15	1.77e-15	1.29e-14
FE Bound 2		7.28e+12	1.23e+11	1.24e+11	1.31e+14	0.897	1.28	9.32

TABLE 18. Results PQ: Alternative Calibration

- For Dynare, refer to Adjemian, Bastani, Juillard, Mihoubi, Perendia, Ratto, and Villemot (2011). Dynare under QZ-Based Methods is documented in Villemot (2011) and under Alternatives is the cyclic reduction method. BP (1997) refers to Binder and Pesaran (1997).
- $E[rp]$ is expressed in annual %, $std(\log c_t)$ in quarterly %, and $*$ indicates a backward error less than machine precision, $2^{-52} = 2.2204E-16$.

Measure		QZ-Based Methods				Alternatives		
	Data	Klein (2000)	Sims (2001)	Uhlig (1999)	Dynare QZ	Anderson (2010)	BP (1997)	Dynare CR
$E[rp]$	6.18	5.87	5.87	NaN	6.29	6.23	NaN	NaN
$\Delta E[rp]$		0.359	0.366	NaN	-0.0568	0	NaN	NaN
$E\left[R_t^f\right]$	0.8	5.91	5.11	NaN	-0.326	0.824	NaN	NaN
$std(\log y_t)$	0.01	0.01	0.00997	NaN	0.01	0.01	NaN	NaN
$std(\log c_t)$	0.51	0.469	0.499	NaN	0.531	0.519	NaN	NaN
$std(\log i_t)$	2.65	6.24	6.1	NaN	5.96	6.01	NaN	NaN
Rel. Res.		8.85e-13	4.48e-10	NaN	6.81e-12	1.73e-23	NaN	NaN
BE Bound		1.13e-09	1.91e-09	NaN	4.81e-11	6.01e-17	NaN	NaN
μ_P		1.27e+03	4.27	NaN	7.07	3.47e+06	NaN	NaN
Eig. Sep.		0.0696	0.0701	NaN	0.0759	0.0764	NaN	NaN
Pencil Sep.		1.57e-15	1.53e-15	NaN	1.32e-15	1.39e-15	NaN	NaN
Ψ_P		1.25e+05	1.25e+05	NaN	1.02e+05	1.05e+05	NaN	NaN
Φ_P		6.34e+22	6.5e+22	NaN	7.54e+22	7.17e+22	NaN	NaN
FE Bound 1		8.79e-05	0.0445	NaN	0.000676	1.72e-15	NaN	NaN
FE Bound 2		5.62e+10	2.91e+13	NaN	5.14e+11	1.24	NaN	NaN

TABLE 19. Results P: Alternative Calibration, Alternate Equations

- For Dynare, refer to Adjemian, Bastani, Juillard, Mihoubi, Perendia, Ratto, and Villemot (2011). Dynare under QZ-Based Methods is documented in Villemot (2011) and under Alternatives is the cyclic reduction method. BP (1997) refers to Binder and Pesaran (1997).
- $E[rp]$ is expressed in annual %, $std(\log c_t)$ in quarterly %, and $*$ indicates a backward error less than machine precision, $2^{-52} = 2.2204E-16$.

pricing equation to pricing economic capital

$$1 = E_t\left[e^{\hat{m}_{t+1}+\hat{R}_{t+1}}\right] \tag{116}$$

or equivalently

$$1 = \beta E_t\left[e^{-\tau a_{bar}+\hat{R}_{t+1}}\frac{\mu_{t+1}}{\mu_t}\right] \tag{117}$$

where

$$e^{\hat{m}_t} = \beta e^{-\tau a_{bar}}\frac{\mu_t}{\mu_{t-1}} \tag{118}$$

and μ_t is deterministically detrended λ_t from (112) above. In the results above, the formulation (117) and (**??**) was used, now I replace (117) with (116). This theoretically makes no difference, but numerically changes the problem significantly. To see this intuitively, notice that previously, there were entries in the B matrix (associated with contemporaneous variables) as well as in the A matrix (associated with future variables) in the row associated with this equation. Now with (116), there are only entries in the A matrix (associated with future variables) in the row associated with this equation.

The results for P under the alternative calibration and this alternate formulation can be found in table 19. Three methods, Uhlig (1999), Binder and Pesaran (1997), and Dynare's cyclic reduction, are now unable to solve the model as rank conditions on B required by these methods are violated. For the remaining algorithms, we find that now Dynare's QZ over estimates the risk premium, at least partially by calculating the risk free rate as negative. The remaining QZ algorithms do not fair much better, predicting very high risk free rates. The potentially much larger backward errors and ill conditioning from above is again found here and only for the method of Anderson (2010) would the forward error bound 1 indicate results that could be trusted. On the one hand, one might be tempted to call foul as the forward error 1 for Klein (2000) is significantly smaller than for Sims (2001) despite their similar predicted moments.

This is premature as the results here are only for P and the calculation of the moments also depends on Q, which are contained in table 20. As under the original formulation the equation in Q is ill conditioned even taking F at face value and focussing on the forward error bounds 1 for Q_3 (note that the forward error bounds 2 again potentiate the ill conditioning of P together with Q), it is now the error in Klein (2000) that is an order of magnitude larger than for Sims (2001). As for P, only Anderson (2010) has forward error bounds 1 that would lead the practitioner to trust the results. The results for P

Measure	QZ-Based Methods				Alternatives		
	Klein (2000)	Sims (2001)	Uhlig (1999)	Dynare QZ	Anderson (2010)	BP (1997)	Dynare CR
$\Delta E[rp]$	0.359	0.366	NaN	-0.0568	0	NaN	NaN
Rel. Res. Q_1	2.17e-10	8.4e-10	NaN	1.26e-11	6.81e-24	NaN	NaN
BE Bound Q_1	2.17e-10	8.4e-10	NaN	1.26e-11	6.81e-24	NaN	NaN
μ_{Q_1}	1	1	NaN	1	1	NaN	NaN
Rel. Res. Q_2	1.15e-10	4.5e-10	NaN	6.8e-12	3.66e-24	NaN	NaN
BE Bound Q_2	1.51e-10	5.9e-10	NaN	8.9e-12	4.79e-24	NaN	NaN
μ_{Q_2}	1.31	1.31	NaN	1.31	1.31	NaN	NaN
Pencil Sep.	1.32e-07	1.3e-07	NaN	1.17e-07	1.19e-07	NaN	NaN
Ψ_{Q_1}	4.74e+08	4.79e+08	NaN	4.91e+08	4.88e+08	NaN	NaN
Ψ_{Q_2}	6.8e+08	6.83e+08	NaN	6.97e+08	6.94e+08	NaN	NaN
Ψ_{Q_3}	6.79e+08	6.82e+08	NaN	6.96e+08	6.93e+08	NaN	NaN
Φ_{Q_1}	4.03e+14	4.11e+14	NaN	4.58e+14	4.48e+14	NaN	NaN
Φ_{Q_2}	1.78e+21	1.84e+21	NaN	2.27e+21	2.18e+21	NaN	NaN
Φ_{Q_3}	4.87e+36	5.07e+36	NaN	6.52e+36	6.07e+36	NaN	NaN
FE Bound 1 Q_2	0.00748	0.00427	NaN	3.37e-05	1.56e-16	NaN	NaN
FE Bound 1 Q_3	0.0366	0.00312	NaN	3.37e-05	1.32e-15	NaN	NaN
FE Bound 2 Q_2	8.72e+04	3.46e+05	NaN	5.78e+03	3.05e-09	NaN	NaN
FE Bound 2 Q_3	1e+32	5.37e+34	NaN	1.17e+33	2.7e+21	NaN	NaN

TABLE 20. Results Q: Alternative Calibration, Alternate Equations

- For Dynare, refer to Adjemian, Bastani, Juillard, Mihoubi, Perendia, Ratto, and Villemot (2011). Dynare under QZ-Based Methods is documented in Villemot (2011) and under Alternatives is the cyclic reduction method. BP (1997) refers to Binder and Pesaran (1997).
- $E[rp]$ is expressed in annual %, $std(\log c_t)$ in quarterly %, and $*$ indicates a backward error less than machine precision, $2^{-52} = 2.2204E-16$.

and Q individually are confirmed with the joint measure in $[PQ]$ in table 21. All methods that produced a solution agree that the problem is ill conditioned, which results from the small pencil separations despite the well spaced eigenvalues, the backward errors can differ by multiple orders of magnitude from the relative residuals, and the forward error bounds 1 would indicate to users of all methods but Anderson's (2010) that their results are likely to have errors of economic significance.

For the internal habit model of Jermann (1998) of production based asset pricing does not admit a closed form solution for the linearized problem. Nonetheless, the methods I proposed in the previous sections would warn users of the ill conditioning of the problem and of the expected inaccuracy of solution based on forward error bounds, warnings that none of the methods from the literature provide, even in alternative calibrations that lead to economically significant numerical errors.

4.3. **Smets and Wouters (2007) Medium Scale Macroeconomic Model.** Smets and Wouters (2007) analyze and estimate a DSGE model based on macroeconomic data from the US economy, providing a compact medium scale model that is arguably the benchmark for structural policy analyses. Their New Keynesian model features sticky prices and wages, inflation indexation, consumption habit formation as well as production frictions in investment and capital and fixed costs. The model leverages seven macroeconomic time series from the US economy to estimate the model parameters using Bayesian estimation. They show that the model matches the US macroeconomic data closely and that out-of-sample forecasting performance is comparable VAR and BVAR models.

I explore the permissable parameter space (as defined by the support of the authors' prior) and demonstrate again a numerically problematic parameterization that existing solution methods do not guard against. The resulting differences in the predictions of the moments of endogenous variables are of economic relevance. To begin, I examine the numerical stability of the solution at the posterior mode of Smets and Wouters (2007). In table 22 I present the results for the transition matrix P. The first three rows are the second moments of the three primary New Keynesian variables, inflation, output growth and the nominal interest rate. The different solution methods all produce the same results for these three moments. Notice that the moments are not perfectly matched (cf., Smets and Wouters's (2007, p. 604) table 6), as the posterior combines the prior as well as the likelihood that itself likely will not perfectly match these particular movements. The results are consistent with previous sections, backward errors can differ substantially

Measure		QZ-Based Methods				Alternatives		
	Data	Klein (2000)	Sims (2001)	Uhlig (1999)	Dynare QZ	Anderson (2010)	BP (1997)	Dynare CR
$E[rp]$	6.18	5.87	5.87	NaN	6.29	6.23	NaN	NaN
$\Delta E[rp]$		0.359	0.366	NaN	-0.0568	0	NaN	NaN
$E\left[R_t^f\right]$	0.8	5.91	5.11	NaN	-0.326	0.824	NaN	NaN
$std(\log y_t)$	0.01	0.01	0.00997	NaN	0.01	0.01	NaN	NaN
$std(\log c_t)$	0.51	0.469	0.499	NaN	0.531	0.519	NaN	NaN
$std(\log i_t)$	2.65	6.24	6.1	NaN	5.96	6.01	NaN	NaN
Rel. Res.		1.4e-12	4.48e-10	NaN	6.81e-12	1.73e-23	NaN	NaN
BE Bound		0.00101	1.92e-09	NaN	4.81e-11	2.67e-16	NaN	NaN
μ_P		7.16e+08	4.27	NaN	7.07	1.55e+07	NaN	NaN
Eig. Sep.		0.0696	0.0701	NaN	0.0759	0.0764	NaN	NaN
Pencil Sep.		1.56e-15	1.53e-15	NaN	1.32e-15	1.39e-15	NaN	NaN
Ψ_P		1.25e+05	1.25e+05	NaN	1.02e+05	1.05e+05	NaN	NaN
Φ_P		6.34e+22	6.5e+22	NaN	7.54e+22	7.17e+22	NaN	NaN
FE Bound 1		1.39e-04	0.0445	NaN	6.76e-04	1.72e-15	NaN	NaN
FE Bound 2		8.9e+10	2.91e+13	NaN	5.14e+11	1.24	NaN	NaN

TABLE 21. Results PQ: Alternative Calibration, Alternate Equations

- For Dynare, refer to Adjemian, Bastani, Juillard, Mihoubi, Perendia, Ratto, and Villemot (2011). Dynare under QZ-Based Methods is documented in Villemot (2011) and under Alternatives is the cyclic reduction method. BP (1997) refers to Binder and Pesaran (1997).
- $E[rp]$ is expressed in annual %, $std(\log c_t)$ in quarterly %, and $*$ indicates a backward error less than machine precision, $2^{-52} = 2.2204E-16$.

Measure		QZ-Based Methods				Alternatives		
	Data	Klein (2000)	Sims (2001)	Uhlig (1999)	Dynare QZ	Anderson (2010)	BP (1997)	Dynare CR
std(π)	0.615	0.608	0.608	0.608	0.608	0.608	0.608	0.608
std(Δy)	0.856	0.963	0.963	0.963	0.963	0.963	0.963	0.963
std(r)	0.830	0.656	0.656	0.656	0.656	0.656	0.656	0.656
Rel. Res.		1.32e-16	1.27e-16	1.96e-16	5.74e-17	1.02e-16	2.99e-17	2.06e-17
BE Bound		1e-15	1.28e-15	1.34e-15	1.19e-15	1.78e-15	4.11e-16	4.88e-16
μ_P		7.6	10.1	6.81	20.7	17.5	13.7	23.6
Eig. Sep.		0.0768	0.0768	0.0768	0.0768	0.0768	0.0768	0.0768
Pencil Sep.		4.68e-05	4.68e-05	4.68e-05	4.68e-05	4.68e-05	4.68e-05	4.68e-05
Ψ_P		1.26e+04	1.26e+04	1.26e+04	1.26e+04	1.26e+04	1.26e+04	1.26e+04
Φ_P		4.02e+05	4.02e+05	4.02e+05	4.02e+05	4.02e+05	4.02e+05	4.02e+05
FE Bound 1		1.19e-14	3.49e-14	8.44e-14	2.84e-14	5.82e-14	6.01e-15	1.55e-14
FE Bound 2		5.3e-11	5.11e-11	7.89e-11	2.31e-11	4.09e-11	1.2e-11	8.3e-12

TABLE 22. Results P: Posterior Mode

- For Dynare, refer to Adjemian, Bastani, Juillard, Mihoubi, Perendia, Ratto, and Villemot (2011). Dynare under QZ-Based Methods is documented in Villemot (2011) and under Alternatives is the cyclic reduction method. BP (1997) refers to Binder and Pesaran (1997).
- $E[rp]$ is expressed in annual %, $std(\log c_t)$ in quarterly %, and $*$ indicates a backward error less than machine precision, $2^{-52} = 2.2204E-16$.

from the relative residuals, the eigenvalue and pencil separations likewise differ and the later leads to the condition number that is while modest not ill conditioned considering that this model is medium scale with substantially more variables than the models in the previous sections. Looking at the forward error bounds 1, any numerical errors are going to be far too small to be of economic consequence.

These conclusions carry over to both the results for the impact matrix of shocks Q in table 23 and for the joint measure $[PQ]$ in 24. Relative residuals and eigenvalue separations differ arbitrarily from the backward errors and condition numbers that combine into the forward errors. The bounds on the forward errors 1 for both Q and $[PQ]$ indicate that any numerical errors are unlikely to be economically important.

Now I will turn to a problematic parameterization contained in Smets and Wouters's (2007) prior.[26] Examine the first three rows are the second moments of the three primary New Keynesian variables, inflation, output growth and the nominal interest rate, in table 25 with the results for the transition matrix P. The different methods solving the same linear model with the same parameters produce substantially different moments, some with inflation more and others with inflation less volatile than in the data. Simply looking at the relative residuals, a user might feel comforted as these numbers are of the order of near numerical insignificance. Yet the backward errors are much larger and in some cases the growth or amplification factor is on the order of 1e07. Methods that examine the eigenvalue separation likewise find well separated stable and unstable eigenvalues to assess numerically stability will miss the ill conditioning of this system as driven by the poor separation of the stable and unstable pencils.

The forward error bounds 1, both for P in table 25 and the joint measure $[PQ]$ in table 27 are not overwhelmingly indicative of solutions whose errors will be of economic consequence. This is not true, however, if the forward error bounds 1 for Q_3 in table 26 are examined. Looking at the forward errors for all three measures, the method of Binder and Pesaran (1997) is likely to be the most accurate, but its forward error bounds 1 for Q_3 of 2.5e-08 is still not beyond reproach and some iterative method, such as proposed by Meyer-Gohde and Saecker (2022) or Meyer-Gohde (2023), might be usefully applied to reduce further improve the solution. This also highlights how not one single measure will likely be sufficient in all cases. Certainly when one condition number of $[PQ]$ is on the order of 1e15, a cautious practitioner would calculate and look more closely at the

[26] Please see the code for the specific parameterization of the over 40 parameters.

Measure	QZ-Based Methods				Alternatives		
	Klein (2000)	Sims (2001)	Uhlig (1999)	Dynare QZ	Anderson (2010)	BP (1997)	Dynare CR
std(π)	0.608	0.608	0.608	0.608	0.608	0.608	0.608
Rel. Res. Q_1	4.99e-17	1.21e-16	8.23e-18	3.16e-17	1.46e-17	1.69e-17	1.19e-17
BE Bound Q_1	1.5e-16	4.54e-16	3.08e-17	1.31e-16	5.32e-17	7.64e-17	5.48e-17
μ_{Q_1}	3	3.75	3.74	4.16	3.64	4.51	4.6
Rel. Res. Q_2	7.1e-17	1.72e-16	1.17e-17	4.5e-17	2.08e-17	2.41e-17	1.7e-17
BE Bound Q_2	2.36e-16	7.15e-16	4.85e-17	2.06e-16	8.36e-17	1.21e-16	8.65e-17
μ_{Q_2}	3.32	4.15	4.14	4.58	4.02	5	5.09
Pencil Sep.	0.0491	0.0491	0.0491	0.0491	0.0491	0.0491	0.0491
Ψ_{Q_1}	384	384	384	384	384	384	384
Ψ_{Q_2}	240	240	240	240	240	240	240
Ψ_{Q_3}	2.59e+03	2.59e+03	2.59e+03	2.59e+03	2.59e+03	2.59e+03	2.59e+03
Φ_{Q_1}	502	502	502	502	502	502	502
Φ_{Q_2}	1.66e+03	1.66e+03	1.66e+03	1.66e+03	1.66e+03	1.66e+03	1.66e+03
Φ_{Q_3}	5.48e+06	5.48e+06	5.48e+06	5.48e+06	5.48e+06	5.48e+06	5.48e+06
FE Bound 1 Q_2	1.88e-15	2.42e-15	3.49e-16	1.11e-15	4.62e-16	3.19e-16	2.77e-16
FE Bound 1 Q_3	6.45e-15	1.09e-14	2.01e-14	6.58e-15	2.41e-14	1.82e-15	2.81e-15
FE Bound 2 Q_2	2.5e-14	6.07e-14	4.13e-15	1.59e-14	7.33e-15	8.5e-15	5.98e-15
FE Bound 2 Q_3	4.8e-08	4.63e-08	7.15e-08	2.09e-08	3.71e-08	1.09e-08	7.52e-09

TABLE 23. Results Q: Posterior Mode

- For Dynare, refer to Adjemian, Bastani, Juillard, Mihoubi, Perendia, Ratto, and Villemot (2011). Dynare under QZ-Based Methods is documented in Villemot (2011) and under Alternatives is the cyclic reduction method. BP (1997) refers to Binder and Pesaran (1997).
- $E[rp]$ is expressed in annual %, $std(\log c_t)$ in quarterly %, and $*$ indicates a backward error less than machine precision, $2^{-52} = 2.2204E-16$.

Measure		QZ-Based Methods				Alternatives		
	Data	Klein (2000)	Sims (2001)	Uhlig (1999)	Dynare QZ	Anderson (2010)	BP (1997)	Dynare CR
std(π)	0.615	0.608	0.608	0.608	0.608	0.608	0.608	0.608
std(Δy)	0.856	0.963	0.963	0.963	0.963	0.963	0.963	0.963
std(r)	0.830	0.656	0.656	0.656	0.656	0.656	0.656	0.656
Rel. Res.		1.19e-16	1.4e-16	1.68e-16	5.42e-17	8.79e-17	2.84e-17	1.97e-17
BE Bound		4.45e-15	1.69e-15	3.95e-15	2.13e-15	4.96e-15	6.08e-16	6.42e-16
μ_P		37.5	12.1	23.5	39.3	56.4	21.4	32.6
Eig. Sep.		0.0768	0.0768	0.0768	0.0768	0.0768	0.0768	0.0768
Pencil Sep.		4.64e-05	4.64e-05	4.64e-05	4.64e-05	4.64e-05	4.64e-05	4.64e-05
Ψ_P		1.07e+04	1.07e+04	1.07e+04	1.07e+04	1.07e+04	1.07e+04	1.07e+04
Φ_P		3.95e+05	3.95e+05	3.95e+05	3.95e+05	3.95e+05	3.95e+05	3.95e+05
FE Bound 1		1.01e-14	2.98e-14	7.15e-14	2.41e-14	4.93e-14	5.09e-15	1.31e-14
FE Bound 2		4.68e-11	5.51e-11	6.64e-11	2.14e-11	3.47e-11	1.12e-11	7.77e-12

TABLE 24. Results PQ: Posterior Mode

- For Dynare, refer to Adjemian, Bastani, Juillard, Mihoubi, Perendia, Ratto, and Villemot (2011). Dynare under QZ-Based Methods is documented in Villemot (2011) and under Alternatives is the cyclic reduction method. BP (1997) refers to Binder and Pesaran (1997).
- $E[rp]$ is expressed in annual %, $std(\log c_t)$ in quarterly %, and $*$ indicates a backward error less than machine precision, $2^{-52} = 2.2204E-16$.

Measure		QZ-Based Methods				Alternatives		
	Data	Klein (2000)	Sims (2001)	Uhlig (1999)	Dynare QZ	Anderson (2010)	BP (1997)	Dynare CR
std(π)	0.615	0.564	0.572	0.642	0.532	0.624	0.626	0.578
std(Δy)	0.856	0.557	0.557	0.557	0.557	0.557	0.557	0.557
std(r)	0.830	0.563	0.569	0.623	0.538	0.609	0.61	0.573
Rel. Res.		9.13e-17	3.29e-14	8.67e-14	1.6e-15	7.17e-16	1.03e-18	1.63e-15
BE Bound		4.37e-11	9.56e-12	3.19e-11	7.26e-11	4.11e-11	3.1e-11	2.89e-11
μ_P		4.79e+05	290	367	4.53e+04	5.73e+04	3.01e+07	1.77e+04
Eig. Sep.		0.11	0.11	0.11	0.11	0.11	0.11	0.11
Pencil Sep.		7.66e-14	7.66e-14	7.66e-14	7.66e-14	7.66e-14	7.66e-14	7.66e-14
Ψ_P		2.51e+04	2.51e+04	2.51e+04	2.51e+04	2.51e+04	2.51e+04	2.51e+04
Φ_P		2.84e+15	2.84e+15	2.84e+15	2.84e+15	2.84e+15	2.84e+15	2.84e+15
FE Bound 1		1.99e-14	7.17e-12	1.89e-11	3.49e-13	1.56e-13	2.14e-16	3.55e-13
FE Bound 2		0.26	93.6	247	4.55	2.04	0.00293	4.64

TABLE 25. Results P: Problematic Calibration

- For Dynare, refer to Adjemian, Bastani, Juillard, Mihoubi, Perendia, Ratto, and Villemot (2011). Dynare under QZ-Based Methods is documented in Villemot (2011) and under Alternatives is the cyclic reduction method. BP (1997) refers to Binder and Pesaran (1997).
- $E[rp]$ is expressed in annual %, $std(\log c_t)$ in quarterly %, and $*$ indicates a backward error less than machine precision, $2^{-52} = 2.2204E-16$.

backward and forward errors and condition numbers for P, Q, as well as $[PQ]$. Again the methods I propose would warn users of all methods of the ill conditioning of the problem and further warned of expected inaccuracies in their solution, warnings that none of the methods from the literature provided here.

4.4. **MMB Suite Comparison.** Having demonstrated the relevance of numerical inaccuracies in a controlled environment where a symbolic solution is available and for nearby calibrations of an economically relevant macro-finance model and a widely used policy model, I turn to the question of how prevalent ill-conditioned models or those with large backward errors in the literature as a whole might be. Examining every reasonable cailibration (or parameterization examined during iterative analysis such as posterior sampling via MCMC) of every model in the literature is obviously an impossible task. A useful first step in this direction is provided by the Macroeconomic Model Data Base (MMB) (see Wieland, Cwik, Müller, Schmidt, and Wolters, 2012; Wieland, Afanasyeva, Kuete, and Yoo, 2016), a model comparison initiative at the Institute for Monetary and Financial Stability (IMFS)[27]. While this platform was originally envisioned as a means to compare policy recommendations across a broad set of macroeconomic models from the literature, providing a venue for model robust policy recommendations, it can also be used as a database of models from the literature to compare solution methods with. Version 3.1 contains 151 different models, ranging from small scale, pedagogical models to large scale, estimated models of the US, EU, multi-country economies. Taking the model equations and parameterizations in the database as given, I examine the numerical stability using the methods developed above for the set of solution methods also presented above. I apply the methods of this paper to the set of models appropriate for reproduction,[28] the varying sizes of which are summarized in figure 3. Reiterating this point, this is the same suite of models used in Meyer-Gohde and Saecker (2022) and Meyer-Gohde (2023), which facilitates the comparison of the methods.

To asses the numerical stability within the model database, I will present the worst results for each measure and solution method in a table as well as a density approximation over all the models graphically. Beginning with the worst case measure and method wise

[27] See http://www.macromodelbase.com.

[28] Currently, this is 99 models, ranging from small scale DSGE models to models from policy institutions containing hundreds of variables. Some of the models in the database are deterministic and/or use nonlinear or non-rational (e.g., adaptive) expectations and, hence, are not appropriate for our comparison here.

Measure	QZ-Based Methods				Alternatives		
	Klein (2000)	Sims (2001)	Uhlig (1999)	Dynare QZ	Anderson (2010)	BP (1997)	Dynare CR
std(π)	0.564	0.572	0.642	0.532	0.624	0.626	0.578
Rel. Res. Q_1	7.32e-19	1.58e-17	7.55e-22	1.76e-18	3.01e-21	1.83e-21	2.66e-22
BE Bound Q_1	3.05e-17	1.1e-16	4.92e-21	6.48e-17	3.24e-20	3.23e-20	9.69e-21
μ_{Q_1}	41.6	7.01	6.51	36.9	10.8	17.7	36.4
Rel. Res. Q_2	1.57e-15	3.39e-14	1.62e-18	3.77e-15	6.46e-18	3.93e-18	5.72e-19
BE Bound Q_2	5.48e-14	2.23e-13	1.01e-17	1.4e-13	6.92e-17	6.86e-17	1.78e-17
μ_{Q_2}	34.8	6.57	6.23	37	10.7	17.4	31.2
Pencil Sep.	1.2e-06	1.2e-06	1.2e-06	1.2e-06	1.2e-06	1.2e-06	1.2e-06
Ψ_{Q_1}	3.88e+11	3.88e+11	3.88e+11	3.88e+11	3.88e+11	3.88e+11	3.88e+11
Ψ_{Q_2}	1.79e+08	1.79e+08	1.79e+08	1.79e+08	1.79e+08	1.79e+08	1.79e+08
Ψ_{Q_3}	1.12e+07	1.12e+07	1.12e+07	1.12e+07	1.12e+07	1.12e+07	1.12e+07
Φ_{Q_1}	3.89e+11	3.89e+11	3.89e+11	3.89e+11	3.89e+11	3.89e+11	3.89e+11
Φ_{Q_2}	1.86e+12	1.86e+12	1.86e+12	1.86e+12	1.86e+12	1.86e+12	1.86e+12
Φ_{Q_3}	9.24e+19	9.24e+19	9.24e+19	9.24e+19	9.24e+19	9.24e+19	9.24e+19
FE Bound 1 Q_2	6.63e-13	2.9e-12	7.55e-16	1.18e-12	3.04e-15	3.62e-15	2.86e-15
FE Bound 1 Q_3	1e-05	1.18e-05	1.28e-05	1.99e-07	4.17e-07	2.5e-08	3.49e-06
FE Bound 2 Q_2	2.85e-07	6.13e-06	2.94e-10	6.83e-07	1.17e-09	7.11e-10	1.03e-10
FE Bound 2 Q_3	3.21e+11	1.16e+14	3.04e+14	5.62e+12	2.52e+12	3.61e+09	5.73e+12

TABLE 26. Results Q: Problematic Calibration

- For Dynare, refer to Adjemian, Bastani, Juillard, Mihoubi, Perendia, Ratto, and Villemot (2011). Dynare under QZ-Based Methods is documented in Villemot (2011) and under Alternatives is the cyclic reduction method. BP (1997) refers to Binder and Pesaran (1997).
- $E[rp]$ is expressed in annual %, $std(\log c_t)$ in quarterly %, and $*$ indicates a backward error less than machine precision, $2^{-52} = 2.2204E-16$.

Measure		QZ-Based Methods				Alternatives		
	Data	Klein (2000)	Sims (2001)	Uhlig (1999)	Dynare QZ	Anderson (2010)	BP (1997)	Dynare CR
std(π)	0.615	0.564	0.572	0.642	0.532	0.624	0.626	0.578
std(Δy)	0.856	0.557	0.557	0.557	0.557	0.557	0.557	0.557
std(r)	0.830	0.563	0.569	0.623	0.538	0.609	0.61	0.573
Rel. Res.		9.13e-17	3.29e-14	8.67e-14	1.6e-15	7.17e-16	1.03e-18	1.63e-15
BE Bound		5.48e-11	9.56e-12	1.87e-09	1.13e-10	4.4e-11	3.1e-11	4.52e-11
μ_P		6e+05	290	2.15e+04	7.06e+04	6.14e+04	3.01e+07	2.77e+04
Eig. Sep.		0.11	0.11	0.11	0.11	0.11	0.11	0.11
Pencil Sep.		7.66e-14	7.66e-14	7.66e-14	7.66e-14	7.66e-14	7.66e-14	7.66e-14
Ψ_P		2.51e+04	2.51e+04	2.51e+04	2.51e+04	2.51e+04	2.51e+04	2.51e+04
Φ_P		2.84e+15	2.84e+15	2.84e+15	2.84e+15	2.84e+15	2.84e+15	2.84e+15
FE Bound 1		1.99e-14	7.17e-12	1.89e-11	3.49e-13	1.56e-13	2.14e-16	3.55e-13
FE Bound 2		0.26	93.6	247	4.55	2.04	0.00293	4.64

TABLE 27. Results PQ: Problematic Calibration

- For Dynare, refer to Adjemian, Bastani, Juillard, Mihoubi, Perendia, Ratto, and Villemot (2011). Dynare under QZ-Based Methods is documented in Villemot (2011) and under Alternatives is the cyclic reduction method. BP (1997) refers to Binder and Pesaran (1997).
- $E[rp]$ is expressed in annual %, $std(\log c_t)$ in quarterly %, and $*$ indicates a backward error less than machine precision, $2^{-52} = 2.2204E-16$.

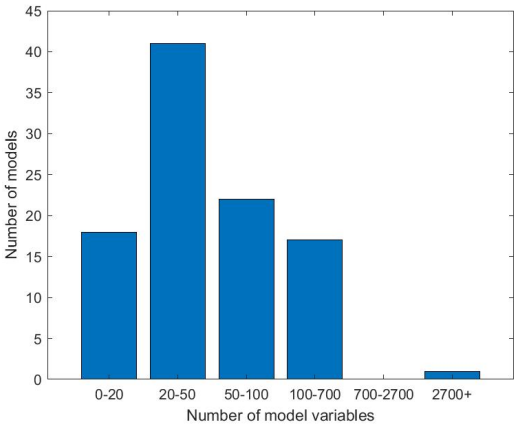


FIGURE 3. Histogram over the number of variables for the 99 MMB models

Figure 3 plots the number of model variables over the amount of MMB models. Currently the total amount of models considered is 99.

for the transition matrix P, see table 28.[29] Note first that the worst case amplification or growth factors that show how backward errors can differ from relative residuals are on the order of 1e06, 1e07 across all methods. With the exception of the cyclic reduction method of Dynare, the relative residuals and also the backward errors are even at their largest comfortingly small. There are models in the database that are ill conditioned at their baseline calibrations as both the condition numbers indicate, following from small separation. Brought together, there are models in the database whose transition matrix forward error bounds cannot rule out numerical instabilities.

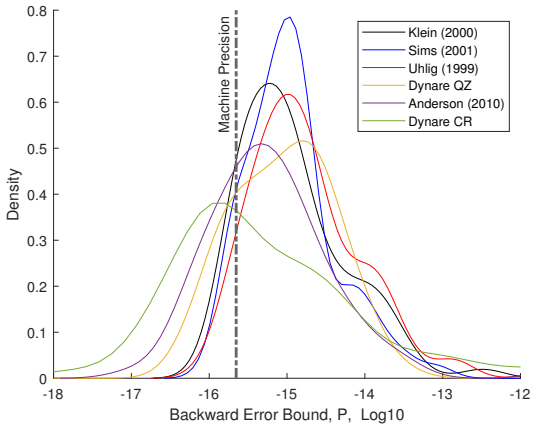
Figure 4 summarizes the backward errors, condition numbers and forward errors across all models graphically. Starting in the upper left panel, the backward errors are generally within a couple orders of magnitude of machine precisions for all methods. Even though the largest backward errors from table 28 are associated with the cyclic reduction method, this method actually is the most accurate on average with its mode even inside machine precision. The upper right shows both measures of the condition numbers are note that this number can be calculated regardless of the solution method and is an assessment of the problem and not the solution. With this depicted on a log10 scale, there are a number of models that are moderately ill conditioned. The lower left panel provides the two forward error bounds and the general conclusion to be taken from here is that all of the methods generally provide solutions that are numerically stable in the sense that

[29] Binder and Pesaran's (1997) method requires specifying a maximum iteration N before solving. Either a computationally prohibitively large N could have been chosen for all parameter draws or a smaller N could have been chosen that might have proven potentially insufficient for some parameter draws, biasing the accuracy measures. I chose at this stage to forgo including the method in the comparison.

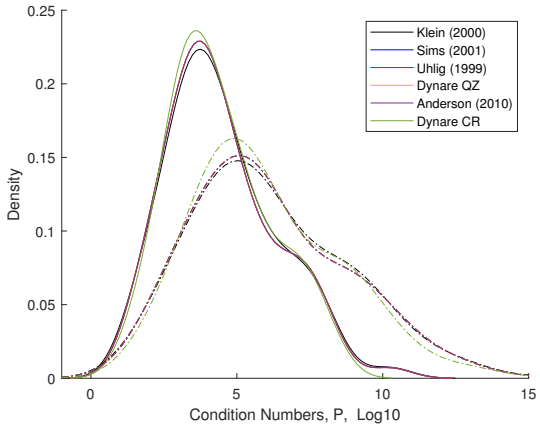
Measure	QZ-Based Methods				Alternatives	
	Klein (2000)	Sims (2001)	Uhlig (1999)	Dynare QZ	Anderson (2010)	Dynare CR
max(Rel. Res.)	1.19e-14	3.23e-11	1.69e-14	1.15e-14	1.25e-12	1.46e-05
max(BE Bound)	3.25e-13	1.24e-10	2.52e-12	2.95e-14	1.85e-10	3.17e-05
max(μ_P)	3.33e+07	3.33e+07	2.13e+06	3.33e+07	3.33e+07	3.33e+07
min(Eig. Sep.)	9.99e-16	6.66e-16	1.33e-15	6.66e-16	1.33e-15	3.42e-07
min(Pencil Sep.)	5.45e-16	5.45e-16	5.45e-16	5.45e-16	5.45e-16	5.45e-16
max(Ψ_P)	1.79e+10	1.79e+10	1.79e+10	1.79e+10	1.79e+10	1.59e+08
max(Φ_P)	1.81e+17	1.81e+17	1.81e+17	1.81e+17	1.81e+17	1.81e+17
max(FE Bound 1)	5.46e-09	1.9e-07	3.06e-08	9.93e-10	1.75e-04	0.00202
max(FE Bound 2)	1.68e-04	1.52e-04	1.05e-04	2.5e-05	0.247	0.868

TABLE 28. Results P: Density over MMB Models

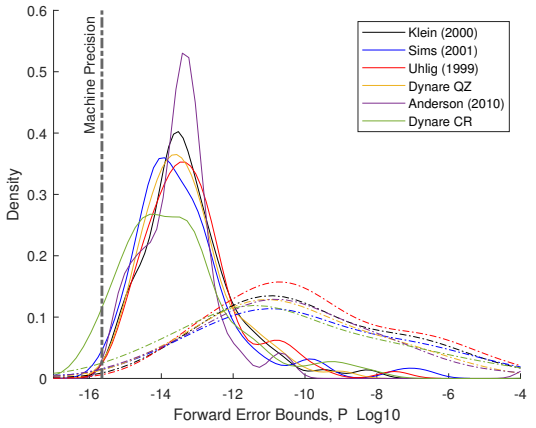
- For Dynare, refer to Adjemian, Bastani, Juillard, Mihoubi, Perendia, Ratto, and Villemot (2011). Dynare under QZ-Based Methods is documented in Villemot (2011) and under Alternatives is the cyclic reduction method. BP (1997) refers to Binder and Pesaran (1997).
- $E[rp]$ is expressed in annual %, $std(\log c_t)$ in quarterly %, and $*$ indicates a backward error less than machine precision, $2^{-52} = 2.2204E-16$.



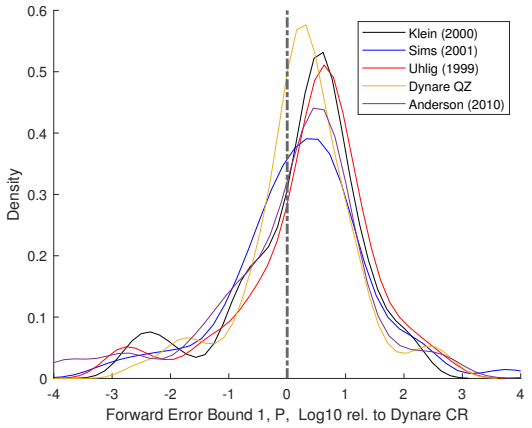
(A) Backward Errors



(B) Condition Numbers Ψ_P and Φ_P



(C) Forward Errors Bounds 1 and 2



(D) Forward Errors Bounds 1, Relative to Dynare CR

FIGURE 4. Backward Error, Condition Numbers, and Forward Error Bounds, P, for MMB Models

their numerical errors in P are unlikely to be of economic consequence. Expressed relative to Dynare's cyclic reduction, the forward error bounds are stably idstributed, meaing that generally the solutions gain and lose accuracy together, i.e. the less accurate solutions in the database are driven more by the problem than the solution method.

Table 29 provides the worst case results from the model database for the impact matrix Q. Again, the worst case amplification or growth factors that show how backward errors can differ from relative residuals are high and on the order of 1e05 to 1e06 across all methods. With the exception of Sims (2001), the relative residuals and also the backward errors are even at their largest comfortingly small. There are models in the database that are ill conditioned at their baseline calibrations as both the condition numbers indicate,

for Q_1, Q_2, and Q_3 indicating that ill conditioning within the database occurs natively in the problem in Q as well as being inherited from P. However, the forward errors seem to indicate that although numerical instabilities of economic consequences in the impact matrices cannot be ruled out, these worst case scenarios for the forward error bound 1 are not overwhelmingly large.

For the impact matrix, figure 4 summarizes the backward errors, condition numbers and forward errors across all models graphically. The results here are for Q_2 so $F = AP+B$ is used with errors in A, B, and P taken into account (but not the dependency of P and A, B, and C). Starting at the upper left, all methods except Klein (2000) and Sims (2001) produce backward errors comfortably inside machine precision, but even for these two methods, the errors are all within two orders of magnitude of machine precision. The condition numbers are in the upper right and they are in general small to modest, though there is a mass of model with significantly ill condition problems in Q. For the forward error bounds, most of the methods for most of the models are very close to machine precision, with the looser bounds driven bay the condition numbers at face value are often very pessimistic. The methods of Klein (2000) and Sims (2001), however, are noticeably less precise with Dynare's QZ joining them as an intermediate case in both the lower panels. Hence I can conclude that the QZ methods are consistently less accurate in their solutions for Q than the remaining QZ method, that of Uhlig (1999) (see the models above, this result is a recurring theme).

Finally the joint measure $[PQ]$ is contained in table 30. Here the differences in the worst case relative residuals and backward errors are apparent, highlighting again the danger of solely residual based error diagnostics, also underscored by the growth or amplification factor running at about 1e07. The worst case separation, both eigenvalue and pencil, are very small (with the exception being the eigenvalue separation for the cyclic reduction method of Dynare, hinting that practitioners using the eigenvalue separation might miss the most ill conditioned model) leading to a worst case very ill donditioned problem by noth measures. Now for the joint measures, economically significant numerical errors cannot be ruled out for the worst case model, espeically and interestingly for the non QZ methods.

Figure 6 summarizes the backward errors, condition numbers and forward errors across all models graphically for the joint measure $[PQ]$. The majority of the mass for all methods is close to machine precision for the backward errors and concentrated around

Measure	QZ-Based Methods				Alternatives	
	Klein (2000)	Sims (2001)	Uhlig (1999)	Dynare QZ	Anderson (2010)	Dynare CR
max(Rel. Res. Q_1)	2.5e-15	1.02e-10	5.53e-17	1.37e-15	4.21e-16	4.72e-17
max(BE Bound Q_1)	4.14e-14	1.02e-10	3.52e-16	4.13e-15	8.36e-16	2.07e-16
max(μ_{Q_1})	7.41e+05	4.45e+06	1.21e+06	4.45e+06	4.45e+06	7.52e+05
max(Rel. Res. Q_2)	2.24e-15	6.33e-11	4.48e-17	3.05e-15	1.34e-15	5.34e-17
max(BE Bound Q_2)	1.06e-13	8.74e-11	6.67e-16	2e-15	1e-15	8.14e-16
max(μ_{Q_2})	1.72e+06	6.06e+06	2.8e+06	6.06e+06	6.06e+06	1.74e+06
min(Pencil Sep.)	3.08e-07	3.08e-07	3.08e-07	3.08e-07	3.08e-07	3.08e-07
max(Ψ_{Q_1})	4.73e+09	4.73e+09	4.73e+09	4.73e+09	4.73e+09	7.08e+08
max(Ψ_{Q_2})	4.73e+09	4.73e+09	4.73e+09	4.73e+09	4.73e+09	7.08e+08
max(Ψ_{Q_3})	2.77e+10	2.77e+10	2.77e+10	2.77e+10	2.77e+10	1.59e+08
max(Φ_{Q_1})	4.8e+09	4.8e+09	4.8e+09	4.8e+09	4.8e+09	7.08e+08
max(Φ_{Q_2})	3.91e+14	3.91e+14	3.91e+14	3.91e+14	3.91e+14	3.91e+14
max(Φ_{Q_3})	7.95e+23	7.95e+23	7.95e+23	7.95e+23	7.95e+23	7.95e+23
max(FE Bound 1 Q_1)	4.19e-09	1.07e-07	4.46e-14	9.52e-13	3.05e-14	9.13e-13
max(FE Bound 1 Q_2)	5.04e-09	2.84e-07	7.78e-09	9.96e-10	3.32e-04	4.01e-04
max(FE Bound 2 Q_1)	3e-07	1.18e-06	6.25e-11	1.22e-10	2.26e-10	1.68e-11
max(FE Bound 2 Q_2)	4.79e+10	3.04e+10	4.11e+10	9.78e+09	1.53e+10	8.12e+09

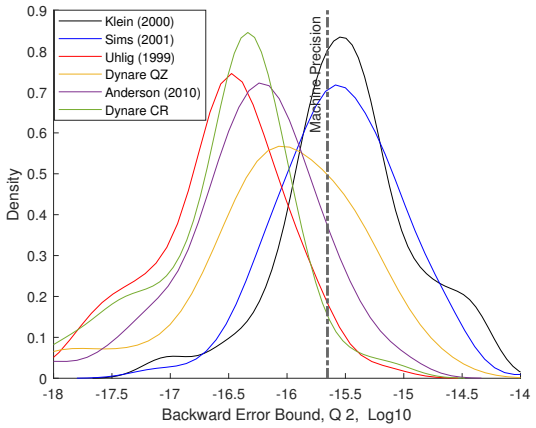
TABLE 29. Results Q: Density over MMB Models

- For Dynare, refer to Adjemian, Bastani, Juillard, Mihoubi, Perendia, Ratto, and Villemot (2011). Dynare under QZ-Based Methods is documented in Villemot (2011) and under Alternatives is the cyclic reduction method. BP (1997) refers to Binder and Pesaran (1997).
- $E[rp]$ is expressed in annual %, $std(\log c_t)$ in quarterly %, and $*$ indicates a backward error less than machine precision, $2^{-52} = 2.2204E-16$.

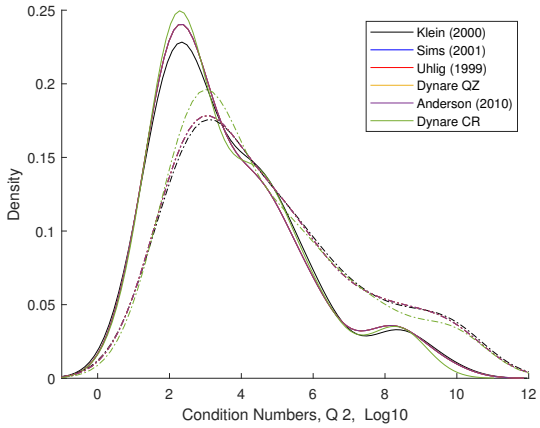
Measure	QZ-Based Methods				Alternatives	
	Klein (2000)	Sims (2001)	Uhlig (1999)	Dynare QZ	Anderson (2010)	Dynare CR
max(Rel. Res.)	1.59e-15	4.08e-11	1.3e-14	1.5e-15	1.02e-12	8.14e-06
max(BE Bound)	5.43e-09	1.24e-10	1.32e-10	1.91e-12	2.08e-08	0.00109
max(μ_P)	9.64e+07	1.73e+07	1e+08	9.36e+07	8.96e+07	8.87e+07
min(Eig. Sep.)	9.99e-16	6.66e-16	1.33e-15	6.66e-16	1.33e-15	3.42e-07
min(Pencil Sep.)	3.65e-16	3.65e-16	3.65e-16	3.65e-16	3.65e-16	3.65e-16
max(Ψ_P)	1.99e+10	1.99e+10	1.99e+10	1.99e+10	1.99e+10	1.59e+08
max(Φ_P)	1.81e+17	1.81e+17	1.81e+17	1.81e+17	1.81e+17	1.81e+17
max(FE Bound 1)	5.31e-09	2.09e-07	2.8e-08	9.94e-10	2.1e-04	0.00108
max(FE Bound 2)	1.56e-04	1.94e-04	7.05e-05	1.72e-05	0.224	0.594

TABLE 30. Results PQ: Density over MMB Models

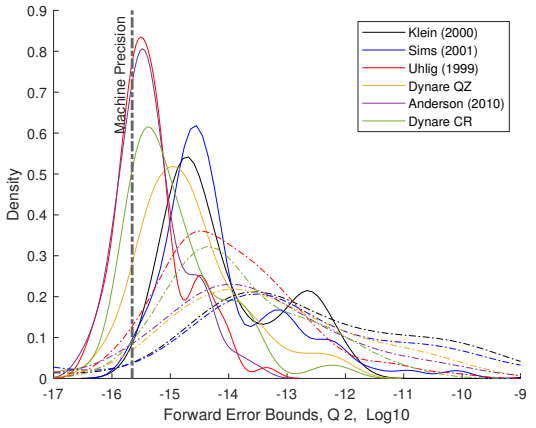
- For Dynare, refer to Adjemian, Bastani, Juillard, Mihoubi, Perendia, Ratto, and Villemot (2011). Dynare under QZ-Based Methods is documented in Villemot (2011) and under Alternatives is the cyclic reduction method. BP (1997) refers to Binder and Pesaran (1997).
- $E[rp]$ is expressed in annual %, $std(\log c_t)$ in quarterly %, and $*$ indicates a backward error less than machine precision, $2^{-52} = 2.2204E-16$.



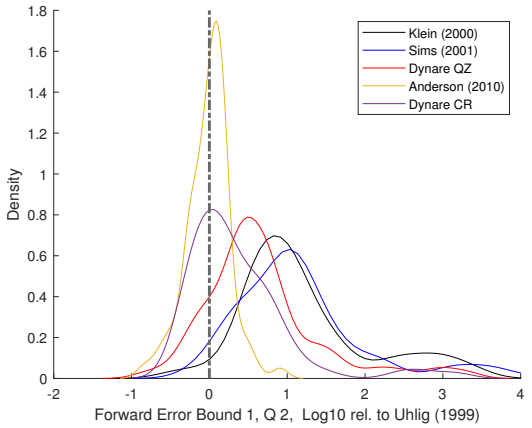
(A) Backward Errors



(B) Condition Numbers $\Psi_{Q,2}$ and $\Phi_{Q,2}$



(C) Forward Errors Bounds 1 and 2

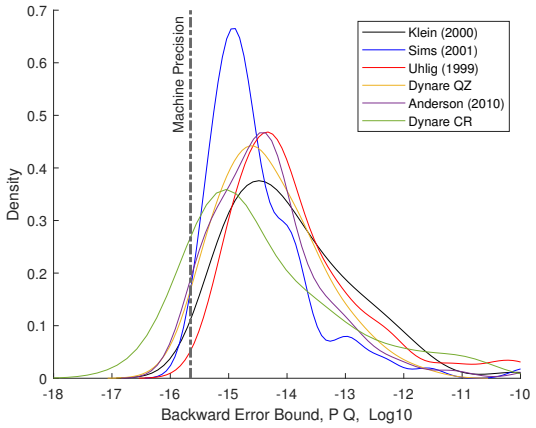


(D) Forward Errors Bounds 1, Relative to Uhlig (1999)

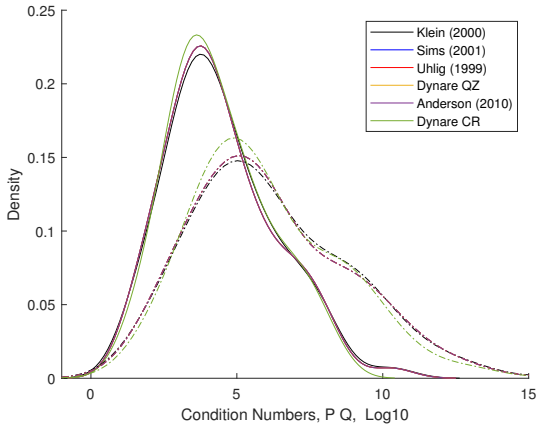
FIGURE 5. Backward Error, Condition Numbers, and Forward Error Bounds, Q, for MMB Models

well to slightly ill conditioned problems. The most accurate method might be judged to be the cyclic reduction method with its left shift of mass in the lower left panel, the forward error bound 1 (importantly recall its poor worst case performance above). Plotting the forward error bounds 1 relative to that of the cyclic reduction method, we see that most of the variation in accuracy is driven by the different problems in the database with the cyclic reduction method generally being about a bit less than an order of magnitude more accurate than the other methods.

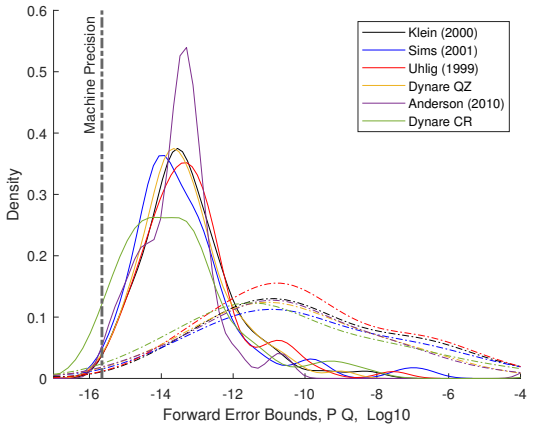
In sum, examining the models in the Macroeconomic Model Data Base (MMB) (see Wieland, Cwik, Müller, Schmidt, and Wolters, 2012; Wieland, Afanasyeva, Kuete, and Yoo, 2016), I do not find evidence for pervasive numerical inaccuracies in models from the



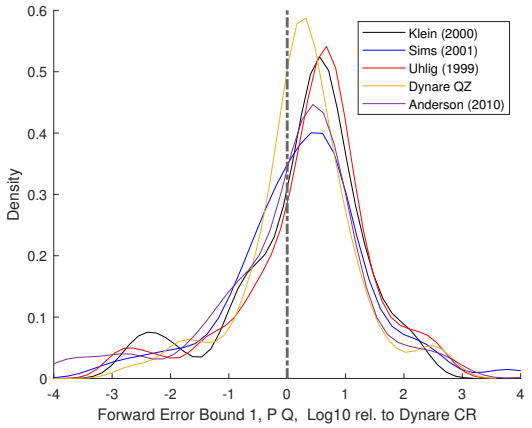
(A) Backward Errors



(B) Condition Numbers Ψ_{PQ} and Φ_{PQ}



(C) Forward Errors Bounds 1 and 2



(D) Forward Errors Bounds 1, Relative to Dynare CR

FIGURE 6. Backward Error, Condition Numbers, and Forward Error Bounds, PQ, for MMB Models

literature. That is not to say that the literature is free from the problems I examine here, with some worst case measures indicating potentially economically relevant numerical errors and many of the problems examined in the literature are ill conditioned. While the non QZ based methods generally perform better than the QZ methods, the method of Uhlig (1999) consistently solves for Q better than the other methods and the worst of the worst case results for the joint measure are produced by the non QZ methods. Hence, there is no systematically superior algorithm and each method may perform better or worse than others depending on the circumstance of the model. This makes the incorporation of diagnostics like those provided in this paper to be all the more important.

4.5. **Smets and Wouters (2007) Posterior.** I return to the Smets and Wouters (2007) model and generate 100,000 draws from their posterior. While the posterior mode was free from economically relevant errors, I also presented a parameterization within the prior that did have such aberrancies. I conclude that numerical instabilities, though present, are not pervasive in the estimates of Smets and Wouters (2007).

I will again present the worst results for each measure and solution method in a table as well as the posterior density of the measures graphically. Beginning with the worst case measure and method wise for the transition matrix P, see table 31, I find that the largest relative residuals and backward errors in the posterior are comfortably small, though again the latter can be multiple orders of magnitude larger than the former. The smallest eigenvalue separation is not of any concern, whereas the pencil separation points to some parameter draws leading to a reduction in conditioning. This is confirmed to some degree as the looser of the two condition numbers is somewhat large for in the worst case parameter constellation. In combination, however, the forward errors and in particular the first bound suggest that all the results in the posterior are reliable, at least from an economic perspective.

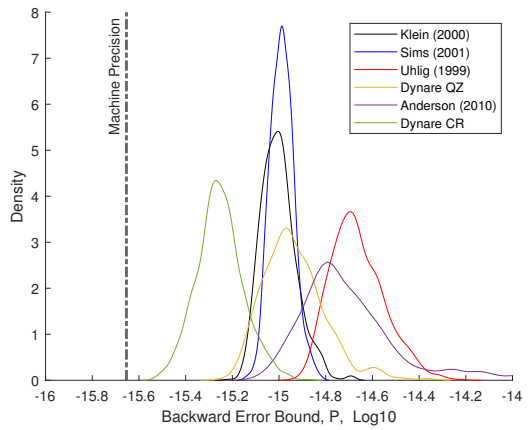
Figure 7 summarizes the backward errors, condition numbers and forward errors across the posterior graphically. Starting in the upper left panel, the backward errors are all within a couple of orders of magnitude of machine precision with the cyclic reduction method leading to the lowest and Uhlig (1999) the highest backward errors. The two condition numbers are tightly distributed through the posterior and agreed upon by all the methods as is to be expected - I do not find convincing evidence of ill conditioning with respect to P in the posterior. Finally referring to the forward errors, the higher accuracy of the cyclic reduction method is maintained - see the lower right panel, it is generally on order of magnitude more accurate - but the entire posterior for all methods is within several orders of magnitude of machine precision. I conclude that the calculations of P in the posterior of Smets and Wouters (2007) for all the methods examined here are free from errors of economic consequence.

Table 29 provides the worst case results from the posterior for the impact matrix Q. Although the backward errors and relative residuals can differ, they are generally very close and around machine precision. The condition numbers are small apart from when the dependency of F on P and P on A, B, and C are taken into account, with the loose bound rising up to 1e07. Nonetheless, the forward error bounds 1 that take into account

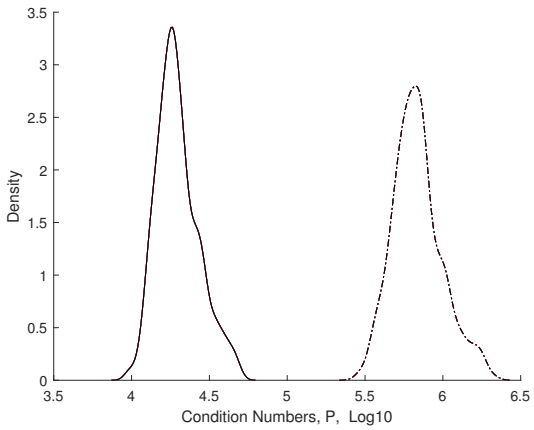
Measure	QZ-Based Methods				Alternatives	
	Klein (2000)	Sims (2001)	Uhlig (1999)	Dynare QZ	Anderson (2010)	Dynare CR
max(Rel. Res.)	1.57e-16	3.84e-16	4.81e-16	1.53e-16	1.33e-15	1.77e-16
max(BE Bound)	2.04e-15	1.5e-15	5.97e-15	4.35e-15	8.27e-14	1.32e-15
max(μ_P)	26.8	18.1	38.5	55.1	71.4	84.3
min(Eig. Sep.)	0.0567	0.0567	0.0567	0.0567	0.0567	0.0567
min(Pencil Sep.)	9.12e-06	9.12e-06	9.12e-06	9.12e-06	9.12e-06	9.12e-06
max(Ψ_P)	5.01e+04	5.01e+04	5.01e+04	5.01e+04	5.01e+04	5.01e+04
max(Φ_P)	2.12e+06	2.12e+06	2.12e+06	2.12e+06	2.12e+06	2.12e+06
max(FE Bound 1)	3.4e-13	2.69e-13	5.77e-13	2.24e-13	3.44e-13	7.19e-14
max(FE Bound 2)	2.18e-10	3.62e-10	8.2e-10	2.17e-10	8.54e-10	1.33e-10

TABLE 31. Results P: Smets and Wouters (2007) Posterior

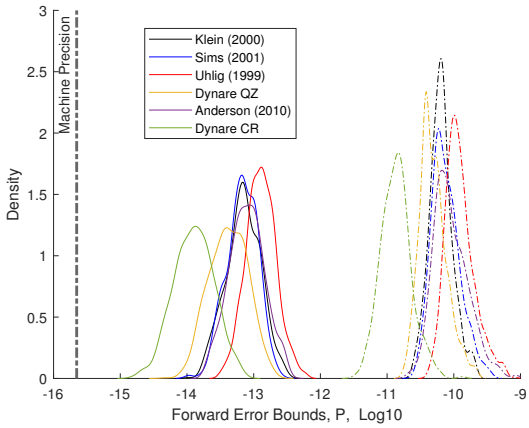
- For Dynare, refer to Adjemian, Bastani, Juillard, Mihoubi, Perendia, Ratto, and Villemot (2011). Dynare under QZ-Based Methods is documented in Villemot (2011) and under Alternatives is the cyclic reduction method. BP (1997) refers to Binder and Pesaran (1997).
- $E[rp]$ is expressed in annual %, $std(\log c_t)$ in quarterly %, and $*$ indicates a backward error less than machine precision, $2^{-52} = 2.2204E-16$.



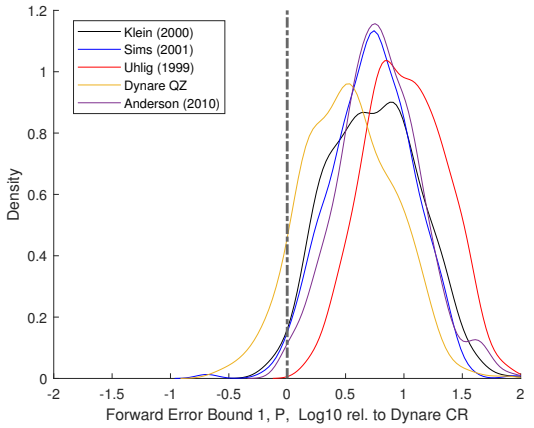
(A) Backward Errors



(B) Condition Numbers Ψ_P and Φ_P



(C) Forward Errors Bounds 1 and 2



(D) Forward Errors Bounds 1, Relative to Dynare CR

FIGURE 7. Backward Error, Condition Numbers, and Forward Error Bounds, P, for Smets and Wouters (2007) Posterior

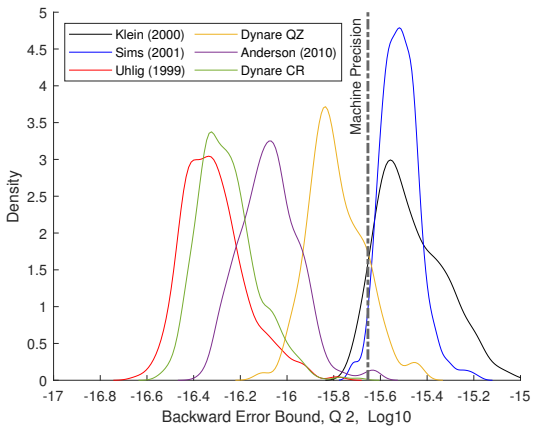
the interaction during the solution places the worst case forward errors only a few orders of magnitude away from machine precision.

For the impact matrix, figure 4 summarizes the backward errors, condition numbers and forward errors across the posterior graphically. The most accurate method for Q is again Uhlig (1999) both in terms of backward and forward errors, though all methods produce low backward errors. The problem across the posterior is well conditioned and in terms of the forward error, Uhlig (1999) and the non QZ methods are roughly equally accurate and the remaining QZ methods roughly comparable and about one order of magnitude less accurate. Nonetheless, there is no evidence of any economically significant numerical errors in the calculation of Q.

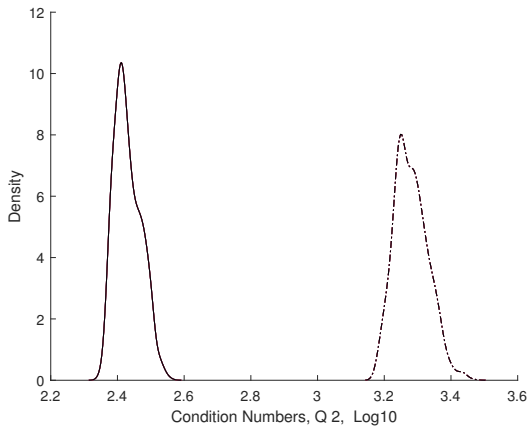
Measure	QZ-Based Methods				Alternatives	
	Klein (2000)	Sims (2001)	Uhlig (1999)	Dynare QZ	Anderson (2010)	Dynare CR
max(Rel. Res. Q_1)	1.21e-16	1.81e-16	1.63e-17	5.13e-17	2.44e-17	1.94e-17
max(BE Bound Q_1)	5.43e-16	4.62e-16	7.81e-17	2.13e-16	1.11e-16	1.05e-16
max(μ_{Q_1})	5.95	4.54	6.19	6.19	6.86	6.22
max(Rel. Res. Q_2)	1.82e-16	2.36e-16	2.72e-17	8.59e-17	4.04e-17	3.4e-17
max(BE Bound Q_2)	9.23e-16	6.58e-16	1.66e-16	3.77e-16	2.39e-16	2.03e-16
max(μ_{Q_2})	6.62	5.01	6.73	6.9	7.63	6.85
min(Pencil Sep.)	0.0365	0.0365	0.0365	0.0365	0.0365	0.0365
max(Ψ_{Q_1})	358	358	358	358	358	358
max(Ψ_{Q_2})	362	362	362	362	362	362
max(Ψ_{Q_3})	6e+03	6e+03	6e+03	6e+03	6e+03	6e+03
max(Φ_{Q_1})	1.07e+03	1.07e+03	1.07e+03	1.07e+03	1.07e+03	1.07e+03
max(Φ_{Q_2})	2.91e+03	2.91e+03	2.91e+03	2.91e+03	2.91e+03	2.91e+03
max(Φ_{Q_3})	3.37e+07	3.37e+07	3.37e+07	3.37e+07	3.37e+07	3.37e+07
max(FE Bound 1 Q_2)	4.93e-15	6.82e-15	7.45e-16	3.35e-15	6.77e-16	7.28e-16
max(FE Bound 1 Q_3)	6.67e-14	5.18e-14	1.17e-13	3.38e-14	1.05e-13	1.64e-14
max(FE Bound 2 Q_2)	6.73e-14	9.24e-14	1.09e-14	3.45e-14	1.63e-14	1.2e-14
max(FE Bound 2 Q_3)	3.02e-07	4.87e-07	1.27e-06	3.39e-07	9.5e-07	1.5e-07

TABLE 32. Results Q: Smets and Wouters (2007) Posterior

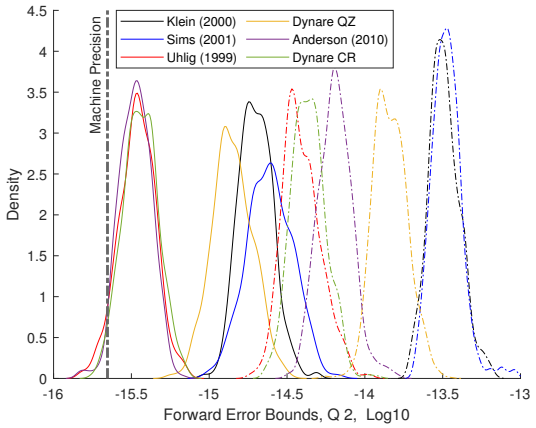
- For Dynare, refer to Adjemian, Bastani, Juillard, Mihoubi, Perendia, Ratto, and Villemot (2011). Dynare under QZ-Based Methods is documented in Villemot (2011) and under Alternatives is the cyclic reduction method. BP (1997) refers to Binder and Pesaran (1997).
- $E[rp]$ is expressed in annual %, $std(\log c_t)$ in quarterly %, and $*$ indicates a backward error less than machine precision, $2^{-52} = 2.2204E-16$.



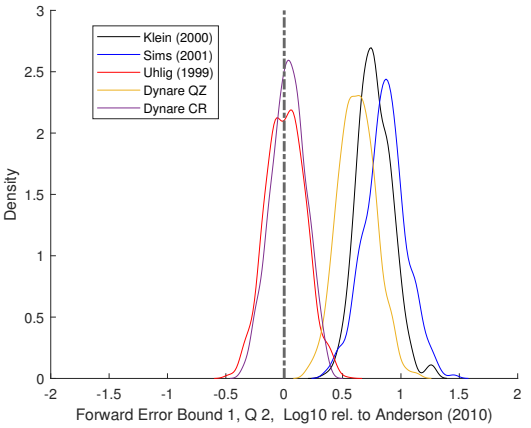
(A) Backward Errors



(B) Condition Numbers $\Psi_{Q,2}$ and $\Phi_{Q,2}$



(C) Forward Errors Bounds 1 and 2



(D) Forward Errors Bounds 1, Relative to Uhlig (1999)

FIGURE 8. Backward Error, Condition Numbers, and Forward Error Bounds, Q, forSmets and Wouters (2007) Posterior

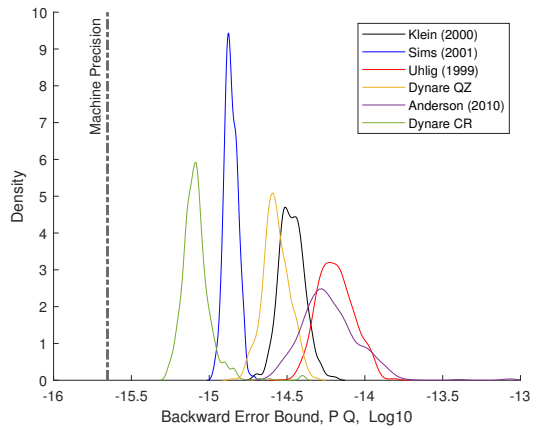
Finally turning to the joint measure $[PQ]$ is table 33. Backward errors and relative residuals differ by about and order of magnitude or two in the worst case and are within a couple orders of magnitude of the machine precision. Eigenvalue and pencil separation results are in line with the results for P and even in the worst case, there is no strong evidence of ill conditioning. An the forward error bounds even at their highest in the posterior are consistent across all solution methods with results free from economically consequential numerical errors.

Figure 6 plots the posterior of the backward errors, condition numbers and forward errors for the joint measure $[PQ]$. The backward errors are all within a few orders of magnitude of machine precision although the cyclic reduction method and Sims (2001) are noticeably more accurate than the remaining methods. The problem is not ill conditioned

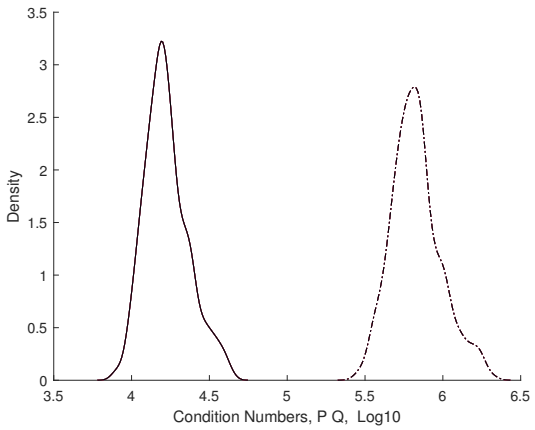
Measure	QZ-Based Methods				Alternatives	
	Klein (2000)	Sims (2001)	Uhlig (1999)	Dynare QZ	Anderson (2010)	Dynare CR
max(Rel. Res.)	1.49e-16	3.46e-16	4.15e-16	1.35e-16	1.15e-15	1.52e-16
max(BE Bound)	6.45e-15	1.94e-15	1.55e-14	4.84e-15	8.53e-14	3.98e-15
max(μ_P)	73.4	23	88.6	78.6	133	95.3
min(Eig. Sep.)	0.0567	0.0567	0.0567	0.0567	0.0567	0.0567
min(Pencil Sep.)	9.1e-06	9.1e-06	9.1e-06	9.1e-06	9.1e-06	9.1e-06
max(Ψ_P)	4.42e+04	4.42e+04	4.42e+04	4.42e+04	4.42e+04	4.42e+04
max(Φ_P)	2.1e+06	2.1e+06	2.1e+06	2.1e+06	2.1e+06	2.1e+06
max(FE Bound 1)	2.92e-13	2.26e-13	5.01e-13	1.91e-13	3e-13	6.2e-14
max(FE Bound 2)	2.07e-10	3.27e-10	7.24e-10	1.96e-10	7.32e-10	1.13e-10

TABLE 33. Results PQ: Smets and Wouters (2007) Posterior

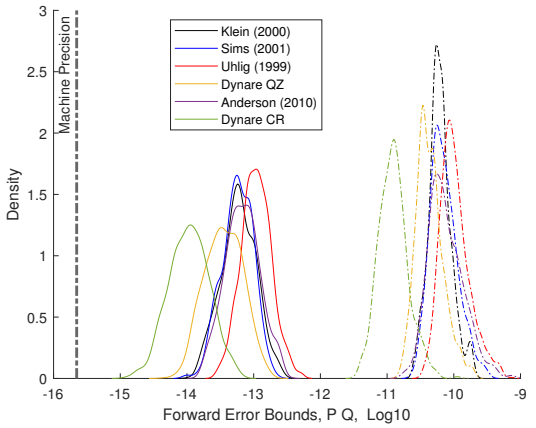
- For Dynare, refer to Adjemian, Bastani, Juillard, Mihoubi, Perendia, Ratto, and Villemot (2011). Dynare under QZ-Based Methods is documented in Villemot (2011) and under Alternatives is the cyclic reduction method. BP (1997) refers to Binder and Pesaran (1997).
- $E[rp]$ is expressed in annual %, $std(\log c_t)$ in quarterly %, and $*$ indicates a backward error less than machine precision, $2^{-52} = 2.2204E-16$.



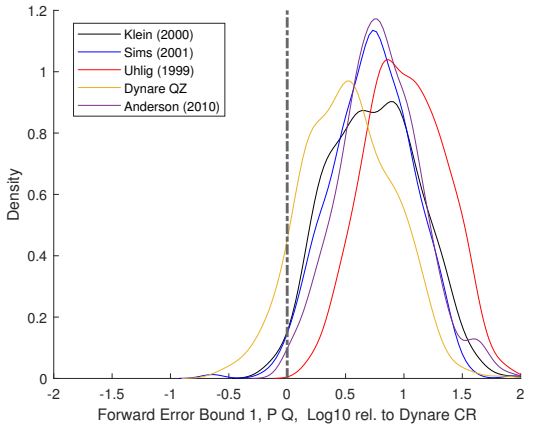
(A) Backward Errors



(B) Condition Numbers Ψ_{PQ} and Φ_{PQ}



(C) Forward Errors Bounds 1 and 2



(D) Forward Errors Bounds 1, Relative to Dynare CR

FIGURE 9. Backward Error, Condition Numbers, and Forward Error Bounds, PQ, for Smets and Wouters (2007) Posterior

as indicated by all methods and all methods produce forward error bounds 1 within a few orders of magnitude of machine precision, with all methods roughly equal apart from the cyclic reduction method which again is about an order of magnitude more accurate that the remaining methods.

All told, I do not find evidence for pervasive numerical inaccuracies Smets and Wouters's (2007) posterior. For the posterior of this particular model, the diagnostics seem to indicate that the cyclic reduction method is the most accurate, generally around one order of magnitude more accurate than the other methods. Yet I find no evidence that use of any of the other methods from the literature examined here would lead to numerical inaccuracies of economic consequence.

5. Conclusion

This paper has provided a complete backward error and condition number analysis of the canonical linear DSGE model. The measures derived here serve to provide practitioners with a warning of ill conditioning and unreliable numerical results from their chosen solution method from the literature with the practical forward error bounds having proven particularly useful. The analysis shows that residual based error diagnostics or error checks based on separation of stable and unstable eigenvalues are theoretically incomplete, as backward errors can be arbitrarily larger than the relative residuals and the appropriate metric for the condition number is the separation between the stable and unstable pencils. The majority of the numerical implantations in the literature rely on the QZ/generalized Schur algorithm is likely to be particularly susceptible to numerical problems as its "stacking" or companion linearization breaks the backward stability of the original algorithm as has been demonstrated previously for the quadratic eigenvalue problem.

I assess the theoretical bounds in five different experiments. In two concrete macro-finance models, calibrated to macroeconomic and financial data, as well as the policy relevant New Keynesian model of Smets and Wouters (2007), I demonstrate economically significant numerical errors especially from standard QZ-based methods. None of the linear solution methods from the literature examined here - Dynare (Adjemian, Bastani, Juillard, Mihoubi, Perendia, Ratto, and Villemot, 2011), Gensys (Sims, 2001), AIM (Anderson and Moore, 1985; Anderson, Levin, and Swanson, 2006), Binder and Pesaran (1997), Uhlig's Toolkit (Uhlig, 1999) and Solab (Klein, 2000) - produced any warnings despite moments predicted by the different methods differing in all significant digits. For both the set of models from the Macroeconomic Model Data Base (MMB) (see Wieland, Cwik, Müller, Schmidt, and Wolters, 2012; Wieland, Afanasyeva, Kuete, and Yoo, 2016) and the posterior of Smets and Wouters (2007), I find that numerical instabilities are not pervasive although the set of models in the MMB is not free from ill conditioned models and potentially large forward errors.

This paper should serve as a cautionary tale that the ubiquitous linear DSGE model's accuracy has been taken for granted and provides a full set of numerical accuracy metrics. These metrics were derived using the same approach as underlies standard error safeguards for nearly singular matrices undoubtedly familiar to all economists. In a set of follow up studies, Meyer-Gohde and Saecker (2022), Meyer-Gohde (2023), Binder and

Meyer-Gohde (2023), and Huber, Meyer-Gohde, and Saecker (2023) all examine different iterative methods introduced in the applied mathematics literature since Moler and Stewart's (1973) QZ algorithm to improve on a solution deemed inaccurate be the methods here.

REFERENCES

ADJEMIAN, S., H. BASTANI, M. JUILLARD, F. MIHOUBI, G. PERENDIA, M. RATTO, AND S. VILLEMOT (2011): "Dynare: Reference Manual, Version 4," Dynare Working Papers 1, CEPREMAP.

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

——— (2010): "A Reliable and Computationally Efficient Algorithm for Imposing the Saddle Point Property in Dynamic Models," *Journal of Economic Dynamics and Control*, 34(3), 472–489.

ANDERSON, G. S., A. LEVIN, AND E. SWANSON (2006): "Higher-Order Pertubation Solutions to Dynamic Discrete-Time Rational Expectations Models," Discussion Paper 2006-01, Federal Reserve Bank of San Francisco Working Paper Series.

ANDERSON, G. S., AND G. MOORE (1985): "A Linear Algebraic Procedure for Solving Linear Perfect Foresight Models," *Economics Letters*, 17(3), 247–252.

ANGUAS, L. M., M. I. BUENO, AND F. M. DOPICO (2019): "A comparison of eigenvalue condition numbers for matrix polynomials," *Linear Algebra and its Applications*, 564, 170 – 200.

BINDER, M., AND A. MEYER-GOHDE (2023): "Revisiting the Fully Recursive Computation of Multivariate Linear Rational Expectations Models," mimeo, Goethe University Frankfurt, Institute for Monetary and Financial Stability (IMFS).

BINDER, M., AND M. H. PESARAN (1997): "Multivariate Linear Rational Expectations Models: Characterization of the Nature of the Solutions and Their Fully Recursive Computation," *Econometric Theory*, 13(6), 877–88.

BINI, D. A., G. LATOUCHE, AND B. MEINI (2002): "Solving matrix polynomial equations arising in queueing problems," *Linear Algebra and its Applications*, 340(1), 225–244.

BLANCHARD, O. J. (1979): "Backward and Forward Solutions for Economies with Rational Expectations," *The American Economic Review*, 69(2), 114–118.

BLANCHARD, O. J., AND C. M. KAHN (1980): "The Solution of Linear Difference Models under Rational Expectations," *Econometrica*, 48(5), 1305–1311.

CAMPBELL, J. Y. (2003): "Consumption-based asset pricing," in *Financial Markets and Asset Pricing*, vol. 1 of *Handbook of the Economics of Finance*, chap. 13, pp. 803–887. Elsevier.

CAMPBELL, J. Y., AND J. H. COCHRANE (1999): "By Force of Habit: A Consumption-Based Explanation of Aggregate Stock Market Behavior," *Journal of Political Economy*, 107(2), 205–251.

CAMPBELL, J. Y., AND R. J. SHILLER (1988): "The Dividend-Price Ratio and Expectations of Future Dividends and Discount Factors," *The Review of Financial Studies*, 1(3), 195–228.

CHEN, X. S., AND P. LV (2018): "On estimating the separation between (A,B) and (C,D) associated with the generalized Sylvester equation $AXD - BXC = E$," *Journal of Computational and Applied Mathematics*, 330, 128–140.

CHU, K.-W. E. (1987): "The Solution of the Matrix Equations $AXB - CXD = E$ and $(YA - DZ, YC - BZ) = (E,F)$," *Linear Algebra and its Applications*, 93, 93–105.

COCHRANE, J. H. (2008): "Financial Markets and the Real Economy," in *Handbook of the Equity Risk Premium*, ed. by R. Mehra, Handbooks in Finance, pp. 237 – 325. Elsevier, San Diego.

CONSTANTINIDES, G. M. (1990): "Habit Formation: A Resolution of the Equity Premium Puzzle," *Journal of Political Economy*, 98(3), 519–543.

DAVIS, G. J. (1981): "Numerical Solution of a Quadratic Matrix Equation," *SIAM Journal on Scientific and Statistical Computing*, 2(2), 164–175.

DEMMEL, J. W. (1987): "On Condition Numbers and the Distance to the Nearest Ill-posed Problem.," *Numerische Mathematik*, 51, 251–290.

DEMMEL, J. W., AND B. KÅGSTRÖM (1987): "Computing Stable Eigendecompositions of Matrix Pencils," *Linear Algebra and its Applications*, (88/89), 139–186.

DENNIS, JR., J. E., J. F. TRAUB, AND R. P. WEBER (1976): "The Algebraic Theory of Matrix Polynomials," *SIAM Journal on Numerical Analysis*, 13(6), 831–845.

FARRAR, D. E., AND R. R. GLAUBER (1967): "Multicollinearity in Regression Analysis: The Problem Revisited," *The Review of Economics and Statistics*, 49(1), 92–107.

FERNÁNDEZ-VILLAVERDE, J., J. RUBIO-RAMÍREZ, AND F. SCHORFHEIDE (2016): "Solution and Estimation Methods for DSGE Models," in *Handbook of Macroeconomics, Volume 2A*, ed. by J. B. Taylor, and H. Uhlig, Handbooks in Economics, chap. 9, pp. 527–724. North-Holland/Elsevier.

GANTMACHER, F. R. (1959): *The Theory of Matrices*, vol. I&II. Chelsea Publishing Company, New York, NY.

GARDINER, J. D., A. J. LAUB, J. J. AMATO, AND C. B. MOLER (1992): "Solution of the Sylvester Matrix Equation AXBT + CXDT = E," *ACM Trans. Math. Softw.*, 18(2), 223–231.

GARDINER, J. D., M. R. WETTE, A. J. LAUB, J. J. AMATO, AND C. B. MOLER (1992): "Algorithm 705; a FORTRAN-77 Software Package for Solving the Sylvester Matrix Equation AXBT + CXDT = E," *ACM Trans. Math. Softw.*, 18(2), 232–238.

GHAVIMI, A. R., AND A. J. LAUB (1995): "Backward error, sensitivity, and refinement of computed solutions of algebraic Riccati equations," *Numerical Linear Algebra with Applications*, 2(1), 29–49.

GOLUB, G. H., AND C. F. VAN LOAN (2013): *Matrix Computations*. The Johns Hopkins University Press, fourth edn.

HAAN, W. J. D., AND A. MARCET (1994): "Accuracy in Simulations," *Review of Economic Studies*, 61(1), 3–17.

HAMMARLING, S., C. J. MUNRO, AND F. TISSEUR (2013): "An Algorithm for the Complete Solution of Quadratic Eigenvalue Problems," *ACM Transactions On Mathematical Software*, 39(3), 18:1–18:19.

HANSEN, L. P., AND K. J. SINGLETON (1983): "Stochastic Consumption, Risk Aversion, and the Temporal Behavior of Asset Returns," *Journal of Political Economy*, 91(2), 249–265.

HEILBERGER, C., T. KLARL, AND A. MAUSSNER (2015): "On the uniqueness of solutions to rational expectations models," *Economics Letter*, 128, 14–16.

HIGHAM, D. J. (1995): "Condition numbers and their condition numbers," *Linear Algebra and its Applications*, 214, 193–213.

HIGHAM, D. J., AND N. J. HIGHAM (1992a): "Backward error and condition of structured linear systems," *SIAM Journal on Matrix Analysis and Applications*, 13(1), 162–175.

HIGHAM, D. J., AND N. J. HIGHAM (1992b): "Componentwise perturbation theory for linear systems with multiple right-hand sides," *Linear Algebra and its Applications*, 174, 111–129.

——— (1998): "Structured Backward Error and Condition of Generalized Eigenvalue Problems," *SIAM Journal on Matrix Analysis and Applications*, 20(2), 493–512.

HIGHAM, N., AND S. HAMMARLING (2005): "Early Numerical Linear Algebra in the UK," SIAM Annual Meeting, New Orleans, July 2005.

HIGHAM, N. J. (1993): "Perturbation Theory and Backward Error for $AX - XB = C$," *BIT*, 33, 124–136.

——— (2002): *Accuracy and Stability of Numerical Algorithms*. Society for Industrial and Applied Mathematics, second edn.

HIGHAM, N. J., AND H.-M. KIM (2000): "Numerical Analysis of a Quadratic Matrix Equation," *IMA Journal of Numerical Analysis*, 20, 499–519.

——— (2001): "Solving a Quadratic Matrix Equation by Newton's Method with Exact Line Searches," *SIAM Journal on Matrix Analysis and Applications*, 23(2), 499–519.

HIGHAM, N. J., D. S. MACKEY, AND F. TISSEUR (2006): "The Conditioning of Linearizations of Matrix Polynomials," *SIAM Journal on Matrix Analysis and Applications*, 28(4), 1005–1028.

HIGHAM, N. J., D. S. MACKEY, F. TISSEUR, AND S. D. GARVEY (2008): "Scaling, sensitivity and stability in the numerical solution of quadratic eigenvalue problems," *International Journal for Numerical Methods in Engineering*, 73(3), 344–360.

HIGHAM, N. J., AND F. TISSEUR (2002): "More on pseudospectra for polynomial eigenvalue problems and applications in control theory," *Linear Algebra and its Applications*, 351-352, 435 – 453, Fourth Special Issue on Linear Systems and Control.

HOPKINS, T. (2002): "Remark on Algorithm 705: A Fortran-77 Software Package for Solving the Sylvester Matrix Equation AXBT + CXDT = E," *ACM Trans. Math. Softw.*, 28(3), 372–375.

HORN, R. A., AND C. R. JOHNSON (1994): *Topics in Matrix Analysis*. Cambridge University Press, Cambridge; New York.

——— (2013): *Matrix Analysis*. Cambridge University Press, Cambridge; New York, 2nd edn.

HUBER, J., A. MEYER-GOHDE, AND J. SAECKER (2023): "Solving Linear DSGE Models with Structure Preserving Doubling Methods," mimeo, Goethe University Frankfurt, Institute for Monetary and Financial Stability (IMFS).

JERMANN, U. J. (1998): "Asset Pricing in Production Economies," *Journal of Monetary Economics*, 41(2), 257–275.

JIN, H.-H., AND K. L. JUDD (2002): "Pertubation Methods for General Dynamic Stochastic Models," Mimeo December, Stanford University.

JUDD, K. L. (1998): *Numerical Methods in Economics*. MIT Press, Cambridge, MA.

JUDD, K. L., L. MALIAR, AND S. MALIAR (2017): "Lower Bounds on Approximation Errors to Numerical Solutions of Dynamic Economic Models," *Econometrica*, 85(3), 991–1012.

KÅGSTRÖM, B., AND P. POROMAA (1996): "LAPACK-Style Algorithms and Software for Solving the Generalized Sylvester Equation and Estimating the Separation between Regular Matrix Pairs," *ACM Transactions on Mathematical Software*, 22(1), 78–103.

KING, R. G., AND S. T. REBELO (1999): "Resuscitating real business cycles," vol. 1 of *Handbook of Macroeconomics*, chap. 14, pp. 927–1007. Elsevier.

KLEIN, P. (2000): "Using the Generalized Schur Form to Solve a Multivariate Linear Rational Expectations Model," *Journal of Economic Dynamics and Control*, 24(10), 1405–1423.

KÖHLER, M. (2021): *Approximate Solution of Non-Symmetric Generalized Eigenvalue Problems and Linear Matrix Equation on HPC Platforms*. Logos Verlag Berlin, Magdeburg, Germany.

——— (2022): "Matrix Equations PACKage – A Fortran library for the solution of Sylvester-like Matrix equations," Doi: 10.5281/zenodo.10016456, Zendo.

KÅGSTRÖM, B. (1994): "A Perturbation Analysis of the Generalized Sylvester Equation $(AR-LB, DR-LE) = (C, F)$," *SIAM Journal on Matrix Analysis and Applications*, 15(4), 1045–1060.

KYDLAND, F. E., AND E. C. PRESCOTT (1982): "Time to Build and Aggregate Fluctuations," *Econometrica*, 50(6), 1345–1370.

LAN, H., AND A. MEYER-GOHDE (2014): "Solvability of Perturbation Solutions in DSGE Models," *Journal of Economic Dynamics and Control*, 45, 366–388.

LETTAU, M. (2003): "Inspecting the Mechanism: Closed-Form Solutions for Asset Prices in Real Business Cycle Models," *The Economic Journal*, 113(489), 550–575.

LUENBERGER, D. G. (1978): "Time-invariant descriptor systems," *Automatica*, 14(5), 473–480.

MEHRA, R. (2003): "The Equity Premium: Why Is It a Puzzle?," *Financial Analysts Journal*, 59(1), 54–69.

MEHRA, R., AND E. C. PRESCOTT (1985): "The Equity Premium: A Puzzle," *Journal of Monetary Economics*, 15(2), 145–161.

MENGI, E., AND M. L. OVERTON (2005): "Algorithms for the computation of the pseudospectral radius and the numerical radius of a matrix," *IMA Journal of Numerical*

Analysis, 25(4), 648–669.

MEYER-GOHDE, A. (2023): "Solving Linear DSGE Models with Bernoulli Methods," IMFS Working Paper Series 182, Goethe University Frankfurt, Institute for Monetary and Financial Stability (IMFS).

MEYER-GOHDE, A., AND J. SAECKER (2022): "Solving Linear DSGE Models with Newton Methods," IMFS Working Paper Series 174, Goethe University Frankfurt, Institute for Monetary and Financial Stability (IMFS).

MICHIELS, W., K. GREEN, T. WAGENKNECHT, AND S.-I. NICULESCU (2006): "Pseudospectra and stability radii for analytic matrix functions with application to time-delay systems," *Linear Algebra and its Applications*, 418(1), 315 – 335.

MOLER, C. B., AND G. W. STEWART (1973): "An Algorithm for Generalized Matrix Eigenvalue Problems," *SIAM Journal on Numerical Analysis*, 10(2), 241–256.

PESARAN, M. H. (2015): *Time series and panel data econometrics*. Oxford University Press.

RIGAL, J. L., AND J. GACHES (1967): "On the Compatibility of a Given Solution With the Data of a Linear System," *J. ACM*, 14(3), 543–548.

SIMS, C. A. (2001): "Solving Linear Rational Expectations Models," *Computational Economics*, 20(1-2), 1–20.

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

SPANOS, A., AND A. MCGUIRK (2002): "The problem of near-multicollinearity revisited: erratic vs systematic volatility," *Journal of Econometrics*, 108(2), 365–393.

STEWART, G. W. (1972): "On the Sensitivity of the Eigenvalue Problem $Ax = \lambda Bx$," *SIAM Journal on Numerical Analysis*, 9(4), 669–686.

STEWART, G. W. (1973): "Error and perturbation bounds for subspaces associated with certain eigenvalue problems," *SIAM review*, 15(4), 727–764.

STEWART, G. W., AND J. SUN (1990): *Matrix perturbation theory*. Academic Press.

TISSEUR, F. (2000): "Backward error and condition of polynomial eigenvalue problems," *Linear Algebra and its Applications*, 309(1), 339–361.

TISSEUR, F., AND N. J. HIGHAM (2001): "Structured Pseudospectra for Polynomial Eigenvalue Problems, with Applications," *SIAM Journal on Matrix Analysis and Applications*, 23(1), 187–208.

TISSEUR, F., AND K. MEERBERGEN (2001): "The Quadratic Eigenvalue Problem," *SIAM Review*, 43(2), 235–286.

TURING, A. M. (1948): "Rounding-Off Errors in Matrix Processes," *The Quarterly Journal of Mechanics and Applied Mathematics*, 1(1), 287–308.

UHLIG, H. (1999): "A Toolkit for Analysing Nonlinear Dynamic Stochastic Models Easily," in *Computational Methods for the Study of Dynamic Economies*, ed. by R. Marimon, and A. Scott, chap. 3, pp. 30–61. Oxford University Press.

VARAH, J. M. (1979): "On the Separation of Two Matrices," *SIAM Journal on Numerical Analysis*, 16(2), 216–222.

VILLEMOT, S. (2011): "Solving Rational Expectations Models at First Order: What Dynare Does," Dynare Working Papers 2, CEPREMAP.

WIELAND, V., E. AFANASYEVA, M. KUETE, AND J. YOO (2016): "New Methods for Macro-Financial Model Comparison and Policy Analysis," in *Handbook of Macroeconomics*, ed. by J. B. Taylor, and H. Uhlig, vol. 2 of *Handbook of Macroeconomics*, pp. 1241–1319. Elsevier.

WIELAND, V., T. CWIK, G. J. MÜLLER, S. SCHMIDT, AND M. WOLTERS (2012): "A new comparative approach to macroeconomic modeling and policy analysis," *Journal of Economic Behavior & Organization*, 83(3), 523–541.

WILKINSON, J. (1979): "Kronecker's canonical form and the QZ algorithm," *Linear Algebra and its Applications*, 28, 285 – 303.

APPENDIX

	h	β	δ	α	σ	ρ	ω
I	0.8617	0.99	0.025	0.36	324.3	0.95	8.355E-02
II	1-9.857E-05	0.99	0.025	0.36	6.109	0.95	6.175E-02
III	1-1.008E-04	1-8.991E-06	0.6402	1-5.680E-04	51.53	1-6.066E-05	7.742E-04
IV	1-6.829E-06	1-5.863E-08	0.6562	1-2.652E-05	1+2.591E-08	1-3.437E-03	1.594E-02
V	1-4.294E-06	1-1.012E-12	0.4727	1-9.990E-05	1+7.590E-08	1-9.628E-04	7.898E-03
VI	1-5.070E-06	1-4.259E-08	0.6539	1-5.715E-05	1+4.755E-05	1-1.221E-03	7.102E-03

TABLE 34. Additional Calibrations I-VI

5.1. **Multivariate pivot derivation of the linear solution using the generalized Schur decomposition.** While this derivation contains nothing substantially new compared with, say Klein (2000), its formulation commensurate with (3) enables a straightforward application of Blanchard's (1979) forward method, making the derivations potentially more transparent and accessible than existing expositions in the literature.

Rearranging the model (3) into the companion linearization yields

$$F\begin{bmatrix} y_t \\ E_t[y_{t+1}] \end{bmatrix} = G\begin{bmatrix} y_{t-1} \\ y_t \end{bmatrix} + \begin{bmatrix} 0_{n_y \times n_\varepsilon} \\ D \end{bmatrix} \varepsilon_t, \quad F \equiv \begin{bmatrix} I_{n_y} & 0_{n_y \times n_y} \\ 0_{n_y \times n_y} & A \end{bmatrix}, G \equiv \begin{bmatrix} 0_{n_y \times n_y} & I_{n_y} \\ -C & -B \end{bmatrix} \tag{A1}$$

where I_{n_y} is an $n_y \times n_y$ identity matrix and $0_{n_y \times n_y}$ is an $n_y \times n_y$ zero matrix.

The generalized Schur decomposition (unitary Q and Z and upper triangular S and T with $Q^*FZ = S$ and $Q^*GZ = T$) of the matrix pencil $P_{FG}(z) = Fz - G$, can be ordered arbitrarily to form

$$\begin{bmatrix} S_{11} & S_{12} \\ 0 & S_{22} \end{bmatrix}\begin{bmatrix} E_t[w^s_{t+1}] \\ E_t[w^u_{t+1}] \end{bmatrix} = \begin{bmatrix} T_{11} & T_{12} \\ 0 & T_{22} \end{bmatrix}\begin{bmatrix} w^s_t \\ w^u_t \end{bmatrix} + Q^*\begin{bmatrix} 0_{n_y \times n_\varepsilon} \\ D \end{bmatrix}\epsilon_t \tag{A2}$$

with the definition $Z\begin{bmatrix} w^{s\prime}_t & w^{u\prime}_t \end{bmatrix}' = \begin{bmatrix} y'_{t-1} & y'_t \end{bmatrix}'$. With any generalized Schur decomposition of $P_{DE}(z)$, the spectrum or set of eigenvalues of the pencil $P_{DE}(z)$ is determined by the diagonal entries of S and T

$$\rho(P_{DE}) = \{t_{ii}/s_{ii}, \text{ if } s_{ii} \neq 0; \infty, \text{ if } s_{ii} = 0; \emptyset, \text{ if } s_{ii} = t_{ii} = 0; i = 1, \ldots, 2n_y\} \tag{A3}$$

where s_{ii} and t_{ii} denote the i'th row and i'th column of S and T respectively. Ordering the decomposition so that the unstable eigenvalues are in the lower right blocks of S and T (hence S_{22} and T_{22}), this lower block can be solved forward following Blanchard (1979) to yield

$$w^u_t = \lim_{j \to \infty} \left(T^{-1}_{22}S_{22}\right)^j E_t\left[w^u_{t+j}\right] - T^{-1}_{22}\underbrace{\begin{bmatrix} \{Q^*\}_{21} & \{Q^*\}_{22} \end{bmatrix}}_{\equiv \{Q^{-1}\}_{2\bullet}}\underbrace{\begin{bmatrix} 0'_{n_y \times n_\varepsilon} & D' \end{bmatrix}'}_{\equiv \hat{D}}\varepsilon_t = -T^{-1}_{22}\{Q^*\}_{2\bullet}\hat{D}\varepsilon_t \tag{A4}$$

where the invertibility of T_{22} and convergence of $\lim_{j\to\infty}\left(T_{22}^{-1}S_{22}\right)^j$ follow directly from the ordering above. Using the definition $Z\left[w_t^{s\prime} \quad w_t^{u\prime}\right]' = \left[y_{t-1}' \quad y_t'\right]'$ from above delivers

$$w_t^u = \left[Z_{21}^* \quad Z_{22}^*\right]\left[y_{t-1}' \quad y_t'\right]' = -T_{22}^{-1}\{Q^*\}_2 . \hat{D}\varepsilon_t \tag{A5}$$

where * indicates the complex conjugation of Z that delivers its inverse by virtue of it being a unitary matrix. If the necessary and sufficient assumptions for a unique stable solution for y_t of (3) from the main text hold, the unique stable solution for y_t is given by

$$y_t = Z_{21}Z_{11}^{-1}y_{t-1} - \left(Z_{22} - Z_{21}Z_{11}^{-1}Z_{12}\right)T_{22}^{-1}\{Q^*\}_2 . \hat{D}\varepsilon_t \tag{A6}$$

$$= Q_{11}S_{11}^{-1}T_{11}Q_{11}^{-1}y_{t-1} - \left(Z_{22} - Z_{21}Z_{11}^{-1}Z_{12}\right)T_{22}^{-1}\{Q^*\}_2 . \hat{D}\varepsilon_t \tag{A7}$$

where ${Z_{22}^*}^{-1} = Z_{22} - Z_{21}Z_{11}^{-1}Z_{12}$ and ${Z_{22}^*}^{-1}Z_{21}^* = -Z_{21}Z_{11}^{-1}$ follow from the properties of unitary matrices and $Z_{21}Z_{11}^{-1} = Q_{11}S_{11}^{-1}T_{11}Q_{11}^{-1}$ from the first block rows of F and G in (12) and upper triangularity of S and T. From $Q_{11}S_{11}^{-1}T_{11}Q_{11}^{-1}$, it follows that the recursion in y_t is stable from the ordering of the eigenvalues above, i.e. the eigenvalues of the upper left block of the generalized Schur decomposition, $det(S_{11}\lambda - T_{11}) = 0$, are inside the unit circle.

5.2. **Norms, eigenvalues, singular values.** These results will be used repeatedly.

The 2-norm of X, $\|X\|_2$ is given by

$$\|X\|_2 = \sigma_{max}(X) = \left(\lambda_{max}\left(XX^*\right)\right)^{1/2} = \left(\lambda_{max}\left(X^*X\right)\right)^{1/2} \tag{A8}$$

where σ_{max} indicates the largest singular value and λ_{max} the largest eigenvalue.

The Frobineus-norm of X, $\|X\|_F$ is given by

$$\|X\|_2 = \left(\text{trace}\left(A^*A\right)\right)^{1/2} = \left(\sum_i \sigma_i(X)^2\right)^{1/2} \tag{A9}$$

As $\|X\|_F^2 = \|X\|_2^2 + \sum_{i\neq\max}\sigma_i(X)^2$ and $\sum_i \sigma_i(X)^2 \leq n\sigma_{\max}(X)^2$, it follows that

$$\|X\|_2 \leq \|X\|_F \leq n^{1/2}\|X\|_2 \tag{A10}$$

$\|Q\|_F^2 = \text{trace}(Q*Q) = \sum_{i=1}^{\min(n_y,n_\varepsilon)}\sigma_i(Q) \geq \sigma_{max}(Q)$.

The eigenvalues of $A\otimes B$ are the eigenvalues of A times the eigenvalues of B, (Horn and Johnson, 1994, Theorem 4.2.12).

Eigenvalue inequalities used in the proofs follow from results on AA^* being Hermitian, (Horn and Johnson, 2013, p. 226) and the (Courant-Fischer-)Weyl Theorem, (Horn and Johnson, 2013, p. 239).

5.3. **Proof of Theorem 1 - Backward error: Matrix quadratic (P).** For an approximate solution $\hat{P}$ to $AP^2 + BP + C = 0$.

$$\eta_P(\hat{P}) = \min\Big\{\epsilon : (A+\Delta A)\hat{P}^2 + (B+\Delta B)\hat{P} + C + \Delta C = 0,$$
$$\|\Delta A\|_F \leq \epsilon\alpha, \|\Delta B\|_F \leq \epsilon\beta, \|\Delta C\|_F \leq \epsilon\gamma\Big\}$$

the constraint can be written as

$$\Delta A\hat{P}^2 + \Delta B\hat{P} + \Delta C = -R \qquad \text{where } R = A\hat{P}^2 + B\hat{P} + C$$

Hence

$$\|R\|_F = \left\|\Delta A\hat{P}^2 + \Delta B\hat{P} + \Delta C\right\|_F \le \|\Delta A\|_F \left\|\hat{P}^2\right\|_F + \|\Delta B\|_F \left\|\hat{P}\right\|_F + \|\Delta C\|_F$$
$$\le \left(\alpha \left\|\hat{P}^2\right\|_F + \beta \left\|\hat{P}\right\|_F + \gamma\right) \eta_P(\hat{P})$$

so

$$\frac{\|R\|_F}{\alpha \left\|\hat{P}^2\right\|_F + \beta \left\|\hat{P}\right\|_F + \gamma} \le \eta_P(\hat{P})$$

Define

$$z = \begin{bmatrix} \alpha^{-1}\,\mathrm{vec}(\Delta A) \\ \beta^{-1}\,\mathrm{vec}(\Delta B) \\ \gamma^{-1}\,\mathrm{vec}(\Delta C) \end{bmatrix}$$

Then using the property that $\|\mathrm{vec}(X)\|_2 = \|X\|_F$ and the property that $\|X\|_F = \sqrt{\mathrm{tr}(X^*X)}$ (and hence $\left\|\begin{bmatrix} X & Y \end{bmatrix}\right\|_F = \sqrt{\|\mathrm{tr}(X)\| + \|\mathrm{tr}(Y)\|}$) gives

$$\|z\|_2^2 = \alpha^{-2}\|\Delta A\|_F^2 + \beta^{-2}\|\Delta B\|_F^2 + \gamma^{-2}\|\Delta C\|_F^2 \le 3\eta_P(\hat{P})^2$$
$$\text{and } \|z\|_2^2 \ge \eta_P(\hat{P})^2$$

So

$$\frac{1}{\sqrt{3}}\|z\|_2 \le \eta_P(\hat{P}) \le \|z\|_2$$

using the Kronecker / vectorized representation ($\mathrm{vec}(XYZ) = (Z' \otimes X)\,\mathrm{vec}(Y)$)

$$\underbrace{\begin{bmatrix} \alpha\hat{P}^{2\prime} \otimes I_{n_y} & \beta\hat{P}' \otimes I_{n_y} & \gamma I_{n_{y^2}} \end{bmatrix}}_{\equiv H} \cdot z = \underbrace{-\,\mathrm{vec}(R)}_{\equiv r}$$

where H has dimensions $n_{y^2} \times 3n_{y^2}$ and $H \cdot z = r$ is an underdetermined system in z with the minimum 2-norm solution

$$z = H^+ r$$

So $\|z\|_2 = \left\|H^+ \cdot r\right\|_2$ and $\eta_P(\hat{P}) \le \left\|H^+ \cdot r\right\|_2 \le \left\|H^+\right\|_2 \cdot \|r\|_2 = \left\|H^+\right\|_2 \cdot \|R\|_F$

Hence the backward error is bounded with respect to the relative residual, $RR(\hat{P}) = \frac{\|R\|_F}{\alpha\left\|\hat{P}^2\right\|_F + \beta\left\|\hat{P}\right\|_F + \gamma}$

$$RR(\hat{P}) \le \eta_P(\hat{P}) \le \left\|H^+\right\|_2 \cdot \left(\alpha \left\|\hat{P}^2\right\|_F + \beta \left\|\hat{P}\right\|_F + \gamma\right) \cdot RR(\hat{P})$$

as $\left\|X^+\right\|_2 = \sigma_{min}(X)^{-1}$ and $\sigma_{min}(X)^2 = \lambda_{min}(XX^*)$

$$\left\|H^+\right\|_2^{-2} = \sigma_{min}^2(H) = \lambda_{min}(HH^*) = \lambda_{min}\left(\alpha^2(\hat{P}^{2\prime}\overline{\hat{P}^2}) \otimes I_{n_y} + \beta^2(\hat{P}'\overline{\hat{P}}) \otimes I_{n_y} + \gamma^2 I_{n_{y^2}}\right)$$
$$\ge \alpha^2\sigma_{min}^2(\hat{P}^2) + \beta^2\sigma_{min}^2(\hat{P}) + \gamma^2$$

$$\mu_P(\hat{P}) = \frac{\alpha \left\|\hat{P}^2\right\|_F + \beta \left\|\hat{P}\right\|_F + \gamma}{\left(\alpha^2\sigma_{min}^2(\hat{P}^2) + \beta\sigma_{min}^2(\hat{P}) + \gamma^2\right)^{1/2}} \qquad \text{and} \qquad RR(\hat{P}) \le \eta_P(\hat{P}) \le \mu_P(\hat{P})RR(\hat{P})$$

5.4. **Proof of Theorem 2 - Backward error: Linear equation (Q) 1.**

$$\eta_{Q_1}(\hat{Q}) = \min\Big\{\epsilon : (F+\Delta F)\hat{Q} = -D - \Delta D, \quad \|\Delta F\|_F \le \epsilon\phi, \quad \|\Delta D\|_F \le \epsilon\delta\Big\}$$

The constraint can be written as

$$\Delta F\hat{Q} + \Delta D = -R, \qquad \text{where } R = F\hat{Q} + D$$

Hence

$$\|R\|_F = \left\|\Delta F\hat{Q} + \Delta D\right\|_F \le \|\Delta F\|_F \left\|\hat{Q}\right\|_F + \|\Delta D\|_F \le \left(\phi\left\|\hat{Q}\right\|_F + \delta\right)\eta_{Q_1}(\hat{Q})$$

So the relative residual $RR_{Q_1}(\hat{Q}) = \frac{\|R\|_F}{\phi\|\hat{Q}\|_F+\delta}$ is bounded by the backward error

$$RR_{Q_1}(\hat{Q}) \le \eta_{Q_1}(\hat{Q})$$

Define $z = \begin{bmatrix}\phi^{-1}\operatorname{vec}(\Delta F)\\ \delta^{-1}\operatorname{vec}(\Delta D)\end{bmatrix}$ then $\|z\|_2^2 \ge \eta_{Q_1}(\hat{Q})^2$ and $\|z\|_2^2 \le 2\eta_{Q_1}(\hat{Q})^2$ so $\frac{1}{\sqrt{2}}\|z\|_2 \le \eta_{Q_1}(\hat{Q}) \le \|z\|_2$ using the Kronecker / vectorized representation

$$\underbrace{\left[\phi\hat{Q}' \otimes I_{n_e} \quad \delta I_{n_y\cdot n_e}\right]}_{\equiv H}\cdot z = \underbrace{-\operatorname{vec}(R)}_{\equiv r}$$

where H has dimensions $n_y n_e \times 2n_y n_e$ and $H\cdot z = r$ is an underdetermined system in z with the minimum 2-norm solution

$$z = H^+ r$$

So $\|z\|_2 = \left\|H^+\cdot r\right\|_2$ and $\eta_{Q_1}(\hat{Q}) \le \left\|H^+\cdot r\right\|_2 \le \left\|H^+\right\|_2 \cdot \|R\|_F$
as $\left\|X^+\right\|_2 = \sigma_{min}(X)^{-1}$ and $\sigma_{min}(X)^2 = \lambda_{min}(XX^*)$

$$\begin{aligned}\left\|H^+\right\|_2^{-2} = \sigma_{min}(H)^2 = \lambda_{min}(HH^*) &= \lambda_{min}\left(\phi^2(\hat{Q}'\overline{\hat{Q}}) \otimes I_{n_e} + \delta^2 I_{n_y\cdot n_e}\right)\\ &= \lambda_{min}\left(\phi^2(\hat{Q}^+\overline{\hat{Q}}) \otimes I_{n_e} + \delta^2 I_{n_y\cdot n_e}\right)\\ &\ge \phi^2\sigma_{min}(\hat{Q}) + \delta^2\end{aligned}$$

So

$$RR_{Q_1}(\hat{Q}) \le \eta_{Q_1}(\hat{Q}) \le \mu_{Q_1}(\hat{Q})RR_{Q_1}(\hat{Q})$$

where $\mu_{Q_1} = \frac{\|\hat{Q}\|_F+\delta}{\left(\phi^2\sigma_{min}^2(\hat{Q})+\delta^2\right)^{1/2}}$

5.5. **Proof of Theorem 3 - Backward error: Linear equation (Q) 2.**

$$\eta_{Q_2}(\hat{P},\hat{Q}) = \min\Big\{\epsilon : \big((A+\Delta A)\hat{P} + B + \Delta B\big)\hat{Q} = -D - \Delta D, \\ \|\Delta A\|_F \le \epsilon\alpha, \quad \|\Delta B\|_F \le \epsilon\beta, \quad \|\Delta D\|_F \le \epsilon\delta\Big\}$$

The constraint can be written as

$$\Delta A\hat{P}\hat{Q} + \Delta B\hat{Q} + \Delta D = -R, \qquad \text{where } R = A\hat{P}\hat{Q} + B\hat{Q} + D$$

$$\|R\|_F = \left\|\Delta A\hat{P}\hat{Q} + \Delta B\hat{Q} + \Delta D\right\|_F \le \|\Delta A\|_F \left\|\hat{P}\hat{Q}\right\|_F + \|\Delta B\|_F \left\|\hat{Q}\right\|_F + \|\Delta D\|_F \\ \le \left(\alpha\left\|\hat{P}\hat{Q}\right\|_F + \beta\left\|\hat{Q}\right\|_F + \delta\right)\eta_{Q_2}(\hat{P},\hat{Q})$$

So the relative residual $RR_{Q_2}(\hat{P},\hat{Q}) = \frac{\|R\|_F}{\alpha\|\hat{P}\hat{Q}\|_F + \beta\|\hat{Q}\|_F + \delta}$ is bounded below by the backward error

$$RR_{Q_2}(\hat{P},\hat{Q}) \le \eta_{Q_2}(\hat{Q})$$

Define $z = \begin{bmatrix} \alpha^{-1}\operatorname{vec}(\Delta A) \\ \beta^{-1}\operatorname{vec}(\Delta B) \\ \delta^{-1}\operatorname{vec}(\Delta D) \end{bmatrix}$ and as $\|z\|_2^2 \ge \eta_{Q_2}(\hat{P},\hat{Q})^2$ and $\|z\|_2^2 \le 3\eta_{Q_2}(\hat{P},\hat{Q})^2$, $\frac{1}{\sqrt{3}}\|z\|_2 \le \eta_{Q_2}(\hat{P},\hat{Q}) \le \|z\|_2$ using the Kronecker / vectorized representation

$$\underbrace{\begin{bmatrix}\alpha(\hat{P}\hat{Q})' \otimes I_{n_y} & \beta\hat{Q}' \otimes I_{n_y} & \delta I_{n_y\cdot n_e}\end{bmatrix}}_{\equiv H}\cdot z = \underbrace{-\operatorname{vec}(R)}_{\equiv r}$$

where H has dimensions $n_y n_e \times 3n_y n_e$ and $H\cdot z = r$ is an underdetermined system in z with the minimum 2-norm solution

$$z = H^+ r$$

So $\|z\|_2 = \left\|H^+\cdot r\right\|_2$ and $\eta_{Q_2}(\hat{P},\hat{Q}) \le \left\|H^+\cdot r\right\|_2 \le \left\|H^+\right\|_2 \cdot \|R\|_F$
as $\left\|X^+\right\|_2 = \sigma_{min}(X)^{-1}$ and $\sigma_{min}(X)^2 = \lambda_{min}(XX^*)$

$$\left\|H^+\right\|_2^{-2} = \sigma_{min}(H)^2 = \lambda_{min}(HH^*) = \lambda_{min}\left(\alpha^2(\hat{P}\hat{Q})'(\overline{\hat{P}\hat{Q}}) \otimes I_{n_y} + \beta^2\hat{Q}'\overline{\hat{Q}} \otimes I_{n_e} + \delta^2 I_{n_y\cdot n_e}\right) \\ \ge \alpha^2\sigma_{min}^2(\hat{P}\hat{Q}) + \beta\sigma_{min}^2(\hat{Q}) + \delta^2$$

So

$$RR_{Q_2}(\hat{P},\hat{Q}) \le \eta_{Q_2}(\hat{P},\hat{Q}) \le \mu_{Q_2}(\hat{P},\hat{Q})RR_{Q_2}(\hat{P},\hat{Q})$$

where $\mu_{Q_2}(\hat{P},\hat{Q}) = \frac{\alpha\|\hat{P}\hat{Q}\|_F + \beta\|\hat{Q}\|_F + \delta}{\left(\alpha^2\sigma_{min}^2(\hat{P}\hat{Q}) + \beta^2\sigma_{min}^2(\hat{Q}) + \delta^2\right)^{1/2}}$

5.6. **Proof of Theorem 4 - Joint backward error joint of P and Q.** For approximate solutions to $\hat{P}$ and $\hat{Q}$

$$\eta_{PQ}(\hat{P},\hat{Q}) = \min\Big\{\epsilon : (A+\Delta A)\hat{P}\begin{bmatrix}\hat{P} & \hat{Q}\end{bmatrix} + (B+\Delta B)\begin{bmatrix}\hat{P} & \hat{Q}\end{bmatrix} + \begin{bmatrix}C+\Delta C & D+\Delta D\end{bmatrix} = 0,$$
$$\|\Delta A\|_F \le \epsilon\alpha, \quad \|\Delta B\|_F \le \epsilon\beta, \quad \|\Delta C\|_F \le \epsilon\gamma, \quad \|\Delta D\|_F \le \epsilon\delta\Big\}$$

$$\Delta A\hat{P}\begin{bmatrix}\hat{P} & \hat{Q}\end{bmatrix} + \Delta B\begin{bmatrix}\hat{P} & \hat{Q}\end{bmatrix} + \begin{bmatrix}\Delta C & \Delta D\end{bmatrix} = -\begin{bmatrix}R_1 & R_2\end{bmatrix}$$
$$R_1 = A\hat{P}^2 + B\hat{P} + C \qquad R_2 = A\hat{P}\hat{Q} + \hat{B}Q + D$$

$$\left\|\begin{bmatrix}X & Y\end{bmatrix}\right\|_F = \left(\|X\|_F^2 + \|Y\|_F^2\right)^{1/2}, \quad \text{if } \|X\|_F \le l\epsilon \text{ and } \|X\|_F \le m\epsilon: \quad \left\|\begin{bmatrix}X & Y\end{bmatrix}\right\|_F \le (l^2+m^2)^{1/2}\epsilon$$

$$\begin{aligned}\left\|\begin{bmatrix}R_P & R_Q\end{bmatrix}\right\|_F &\le \|\Delta A\|_F\left\|\hat{P}\begin{bmatrix}\hat{P} & \hat{Q}\end{bmatrix}\right\|_F + \|\Delta B\|_F\left\|\begin{bmatrix}\hat{P} & \hat{Q}\end{bmatrix}\right\|_F + \left\|\begin{bmatrix}\Delta C & \Delta D\end{bmatrix}\right\|_F \\ &\le \left(\alpha\left\|\hat{P}\begin{bmatrix}\hat{P} & \hat{Q}\end{bmatrix}\right\|_F + \beta\left\|\begin{bmatrix}\hat{P} & \hat{Q}\end{bmatrix}\right\|_F + (\gamma^2+\delta^2)^{1/2}\right)\eta_{PQ}(\hat{P},\hat{Q}) \\ &= \left(\alpha\left\|\hat{P}\begin{bmatrix}\hat{P} & \hat{Q}\end{bmatrix}\right\|_F + \beta\left\|\begin{bmatrix}\hat{P} & \hat{Q}\end{bmatrix}\right\|_F + \left\|\begin{bmatrix}\gamma & \delta\end{bmatrix}\right\|_F\right)\eta_{PQ}(\hat{P},\hat{Q})\end{aligned}$$

As $\left\|\begin{bmatrix}R_P & R_Q\end{bmatrix}\right\|_F = \left(\|R_P\|_F^2 + \|R_Q\|_F^2\right)^{1/2}$

$$\|R_i\|_F = \left(\left\|\begin{bmatrix}R_P & R_Q\end{bmatrix}\right\|_F^2 - \|R_{-i}\|_F^2\right)^{1/2} \le \left\|\begin{bmatrix}R_1 & R_2\end{bmatrix}\right\|_F, \; i \in \{P,Q\}$$
$$\|R_i\|_F \le \left(\alpha\left\|\hat{P}\begin{bmatrix}\hat{P} & \hat{Q}\end{bmatrix}\right\|_F + \beta\left\|\begin{bmatrix}\hat{P} & \hat{Q}\end{bmatrix}\right\|_F + \left\|\begin{bmatrix}\gamma & \delta\end{bmatrix}\right\|_F\right)\eta_{PQ}(\hat{P},\hat{Q}), \; i \in \{P,Q\}$$
$$RR_{PQ}(\hat{P}) \le \frac{\alpha\left(\|\hat{P}^2\|_F^2 + \|\hat{P}\hat{Q}\|_F^2\right)^{1/2} + \beta\left(\|\hat{P}\|_F^2 + \|\hat{Q}\|_F^2\right)^{1/2} + (\gamma^2+\delta^2)^{1/2}}{\alpha\|\hat{P}^2\|_F + \beta\|\hat{P}\|_F + \gamma}\eta_{PQ}(\hat{P},\hat{Q})$$

as $(a^2+b^2)^{1/2} \le \sqrt{2}\max\{a,b\}$

$$\begin{aligned}&\alpha\left(\|\hat{P}^2\|_F^2 + \|\hat{P}\hat{Q}\|_F^2\right)^{1/2} + \beta\left(\|\hat{P}\|_F^2 + \|\hat{Q}\|_F^2\right)^{1/2} + (\gamma^2+\delta^2)^{1/2} \\ &\le \sqrt{2}\left(\alpha\max\left\{\|\hat{P}^2\|_F, \|\hat{P}\hat{Q}\|_F\right\} + \beta\max\left\{\|\hat{P}\|_F, \|\hat{Q}\|_F\right\} + \max\{\gamma,\delta\}\right)\end{aligned}$$

$$RR_{PQ}(\hat{P}) \le \sqrt{2}\eta_{PQ}(\hat{P},\hat{Q})$$
$$RR_{PQ}(\hat{Q}) \le \sqrt{2}\eta_{PQ}(\hat{P},\hat{Q})$$

Define $z = \begin{bmatrix}\alpha^{-1}\operatorname{vec}(\Delta A)\\ \beta^{-1}\operatorname{vec}(\Delta B)\\ \gamma^{-1}\operatorname{vec}(\Delta C)\\ \delta^{-1}\operatorname{vec}(\Delta D)\end{bmatrix}$ which has dimensions $3n_{y^2} + n_y n_e \times 1$ using the Kronecker / vectorized representation

$$\underbrace{\begin{bmatrix}\alpha\begin{bmatrix}(\hat{P}^2)'\\ (\hat{P}\hat{Q})'\end{bmatrix}\otimes I_n & \beta\begin{bmatrix}\hat{P}'\\ \hat{Q}'\end{bmatrix}\otimes I_n & \gamma\begin{bmatrix}I_{n_y}\\ 0_{n_e\times n_y}\end{bmatrix}\otimes I_{n_y} & \delta\begin{bmatrix}0_{n_y\times n_e}\\ I_{n_e}\end{bmatrix}\otimes I_{n_y}\end{bmatrix}}_{\equiv H}\cdot z = -\underbrace{\begin{bmatrix}\operatorname{vec}(R_P)\\ \operatorname{vec}(R_Q)\end{bmatrix}}_{\equiv r}$$

where each of the 4 blocks of H has dimensions $(ny_+ n_e)n_y \times n_{y^2}$, r has dimensions $n_{y^2} + n_y n_e \times 1$ and $H \cdot z = r$ is an underdetermined system in z with the minimum 2-norm solution

$$z = H^+ r$$

So $\|z\|_2 = \|H^+ \cdot r\|_2$ and $\eta_{PQ}(\hat{P},\hat{Q}) \le \|H^+ \cdot r\|_2 \le \|H^+\|_2 \cdot \|r\|_2 = \|H^+\|_2 \cdot \left\| \begin{bmatrix} R_P & R_Q \end{bmatrix} \right\|_F$

$$\begin{aligned}
\|H^+\|_2^{-2} = \sigma_{min}^2(H) = \lambda_{min}(HH^*) = \lambda_{min}\Big(& \alpha^2(\hat{P}^{2\prime}\overline{\hat{P}^2}) \otimes I_n + \alpha^2\left((\hat{P}\hat{Q})'(\overline{\hat{P}\hat{Q}})\right) \otimes I_n \\
& + \beta^2(\hat{P}'\overline{\hat{P}}) \otimes I_n + \beta^2\left((\hat{Q}'\overline{\hat{Q}})\right) \otimes I_n \\
& + \gamma^2 \begin{bmatrix} I_{n_y} \\ 0_{n_e \times n_y} \end{bmatrix} \begin{bmatrix} I_{n_y} & 0_{n_y \times n_e} \end{bmatrix} \otimes I_{n_y} \\
& + \delta^2 \begin{bmatrix} 0_{n_y \times n_e} \\ I_{n_e} \end{bmatrix} \begin{bmatrix} 0_{n_e \times n_y} & I_{n_e} \end{bmatrix} \otimes I_{n_y} \Big) \\
\ge \alpha^2\left(\sigma_{min}^2(\hat{P}) + \sigma_{min}^2(\hat{P}\hat{Q})\right) + \beta^2\left(\sigma_{min}^2(\hat{P}) + \sigma_{min}^2(\hat{Q})\right) + \gamma^2 + \delta^2 \\
\ge \max\Big\{ & \alpha^2\sigma_{min}^2(\hat{P}^2) + \beta^2\sigma_{min}^2(\hat{P}) + \gamma, \\
& \alpha^2\sigma_{min}^2(\hat{P}\hat{Q}) + \beta^2\sigma_{min}^2(\hat{Q}) + \delta^2 \Big\} \\
\ge \min\Big\{ & \alpha^2\sigma_{min}^2(\hat{P}^2) + \beta^2\sigma_{min}^2(\hat{P}) + \gamma, \\
& \alpha^2\sigma_{min}^2(\hat{P}\hat{Q}) + \beta^2\sigma_{min}^2(\hat{Q}) + \delta^2 \Big\}
\end{aligned}$$

Defining $RR_{PQ}(\hat{P},\hat{Q}) = \dfrac{\left\| \begin{bmatrix} R_P & R_Q \end{bmatrix} \right\|_F}{\alpha(\|\hat{P}^2\|_F + \|\hat{P}\hat{Q}\|_F) + \beta(\|\hat{P}\|_F + \|\hat{Q}\|_F) + (\gamma^2+\delta^2)^{1/2}}$

$$RR_{PQ}(\hat{P},\hat{Q}) \le \eta_{PQ}(\hat{P},\hat{Q}) \le \mu_{PQ}(\hat{P},\hat{Q}) RR_{PQ}(\hat{P},\hat{Q})$$

where $\mu_{PQ}(\hat{P},\hat{Q}) = \dfrac{\alpha(\|\hat{P}^2\|_F + \|\hat{P}\hat{Q}\|_F) + \beta(\|\hat{P}\|_F + \|\hat{Q}\|_F) + (\gamma^2+\delta^2)^{1/2}}{\left(\alpha^2\left[\sigma_{min}^2(\hat{P}^2) + \sigma_{min}^2(\hat{P}\hat{Q})\right] + \beta^2\left[\sigma_{min}^2(\hat{P}) + \sigma_{min}^2(\hat{Q})\right] + \gamma^2 + \delta^2\right)^{1/2}}$

using $c = (a^2+b^2)^{1/2} \quad \rightarrow \quad a = (c^2-b^2)^{1/2} \le c \Rightarrow a^2 = c^2 - b^2 \le c^2$

$$\|R_P\|_F \le \left\| \begin{bmatrix} R_P & R_Q \end{bmatrix} \right\|_F \le \left(\alpha\|\hat{P}^2\|_F + \beta\|\hat{P}\|_F + \gamma\right)\eta_{PQ}(\hat{P},\hat{Q})$$

So $RR_{PQ}(\hat{P}) \le \eta_{PQ}(\hat{P},\hat{Q})$

$$\|R_Q\|_F \le \left\| \begin{bmatrix} R_P & R_Q \end{bmatrix} \right\|_F \le \left(\alpha\|\hat{P}\hat{Q}\|_F + \beta\|\hat{Q}\|_F + \delta\right)\eta_{PQ}(\hat{P},\hat{Q})$$

So $RR_{PQ}(\hat{Q}) \le \eta_{PQ}(\hat{P},\hat{Q})$

$$\begin{aligned}
\eta_{PQ}(\hat{P},\hat{Q}) &\le \|H^+\|_2 \left\| \begin{bmatrix} R_P & R_Q \end{bmatrix} \right\|_F \\
&\le \|H^+\|_2 \sqrt{2}\max\{\|R_P\|_F, \|R_Q\|_F\} \\
&\le \|H^+\|_2 \sqrt{2}\min\{\alpha\|\hat{P}^2\|_F + \beta\|\hat{P}\|_F + \gamma, \alpha\|\hat{P}\hat{Q}\|_F + \beta\|\hat{Q}\|_F + \delta\} \\
&\quad \cdot \frac{\max\{\|R_P\|_F, \|R_Q\|_F\}}{\min\{\alpha\|\hat{P}^2\|_F + \beta\|\hat{P}\|_F + \gamma, \alpha\|\hat{P}\hat{Q}\|_F + \beta\|\hat{Q}\|_F + \delta\}} \\
&\le \|H^+\|_2 \sqrt{2}\min\{\alpha\|\hat{P}^2\|_F + \beta\|\hat{P}\|_F + \gamma, \alpha\|\hat{P}\hat{Q}\|_F + \beta\|\hat{Q}\|_F + \delta\}
\end{aligned}$$

$$\cdot \frac{\max\{RR_{PQ}(\hat{P}), RR_{PQ}(\hat{Q})\}}{\min\{\alpha\|\hat{P}^2\|_F + \beta\|\hat{P}\|_F + \gamma, \alpha\|\hat{P}\hat{Q}\|_F + \beta\|\hat{Q}\|_F + \delta\}}$$

$$\leq \sqrt{2}\frac{\min\{\alpha\|\hat{P}^2\|_F + \beta\|\hat{P}\|_F + \gamma, \alpha\|\hat{P}\hat{Q}\|_F + \beta\|\hat{Q}\|_F + \delta\}}{\max\left\{\left(\alpha^2\sigma_{min}^2(\hat{P}^2) + \beta^2\sigma_{min}^2(\hat{P}) + \gamma^2\right)^{1/2}\left(\alpha^2\sigma_{min}^2(\hat{P}\hat{Q}) + \beta^2\sigma_{min}^2(\hat{Q}) + \delta^2\right)^{1/2}\right\}}$$

$$\cdot \max\{RR_{PQ}(\hat{P}), RR_{PQ}(\hat{Q})\}$$

$$\leq \sqrt{2}\max\{\mu_P(\hat{P})RR_P(\hat{P}), \mu_{Q_2}(\hat{P},\hat{Q})RR_{Q_2}(\hat{Q})\}$$

5.7. **Proof of Theorem 5 - Condition number: Matrix Quadratic.** Consider the perturbed equation where $AP^2 + BP + C = 0$

$$(A+\Delta A)(P+\Delta P)^2 + (B+\Delta B)(P+\Delta P) + C + \Delta C = 0$$

measuring perturbations normwise as

$$\epsilon_P = \max\left\{\frac{\|\Delta A\|_F}{\alpha}, \frac{\|\Delta B\|_F}{\beta}, \frac{\|\Delta C\|_F}{\gamma}\right\}$$

can be written to first order as

$$AP\Delta P + A\Delta PP + \Delta AP^2 + B\Delta P + \Delta BP + \Delta C + \mathcal{O}(\epsilon^2) = 0$$

$$(AP+B)\Delta P + A\Delta PP = -\Delta AP^2 - \Delta BP - \Delta C + \mathcal{O}(\epsilon^2)$$

a generalized Sylvester equation in ΔP. Using the Kronecker / vectorized notation

$$\underbrace{\left(I_{n_y}\otimes(AP+B) + P'\otimes A\right)}_{V}\operatorname{vec}(\Delta P) = -\begin{bmatrix}\alpha\left(P^2\right)'\otimes I_{n_y} & \beta P'\otimes I_{n_y} & \gamma I_{n_{y^2}}\end{bmatrix}\begin{bmatrix}\alpha^{-1}\operatorname{vec}(\Delta A)\\ \beta^{-1}\operatorname{vec}(\Delta B)\\ \gamma^{-1}\operatorname{vec}(\Delta C)\end{bmatrix} + \mathcal{O}(\epsilon^2)$$

$$\frac{\|\Delta P\|_F}{\|P\|_F} = \left\|V^{-1}\begin{bmatrix}\alpha\left(P^2\right)'\otimes I_{n_y} & \beta P'\otimes I_{n_y} & \gamma I_{n_{y^2}}\end{bmatrix}\begin{bmatrix}\alpha^{-1}\operatorname{vec}(\Delta A)\\ \beta^{-1}\operatorname{vec}(\Delta B)\\ \gamma^{-1}\operatorname{vec}(\Delta C)\end{bmatrix}\right\|_2 / \|P\|_F + \mathcal{O}(\epsilon^2)$$

$$\leq \sqrt{3}\Psi(P)\epsilon \leq \sqrt{3}\Phi(P)\epsilon \leq \sqrt{3}\Theta(P)\epsilon$$

$$\Psi(P) = \left\|V^{-1}\begin{bmatrix}\alpha\left(P^2\right)'\otimes I_{n_y} & \beta P'\otimes I_{n_y} & \gamma I_{n_{y^2}}\end{bmatrix}\right\|_2 / \|P\|_F$$

5.8. **Proof of Corollary 2 - Condition number bound P.**

$$\Psi(P) = \left\|V^{-1}\begin{bmatrix}\alpha\left(P^2\right)'\otimes I_{n_y} & \beta P'\otimes I_{n_y} & \gamma I_{n_{y^2}}\end{bmatrix}\right\|_2 / \|P\|_F$$

$$\leq \|V^{-1}\|_2\left\|\begin{bmatrix}\alpha\left(P^2\right)'\otimes I_{n_y} & \beta P'\otimes I_{n_y} & \gamma\end{bmatrix}\right\|_2 / \|P\|_F$$

From Weyl-Courant-Fischer Theorem

$$\left\|\begin{bmatrix}X & Y\end{bmatrix}\right\|_2^2 = \lambda_{max}\left(XX^* + YY^*\right) \leq \lambda_{max}\left(XX^*\right) + \lambda_{max}\left(YY^*\right)$$

$$= \|X\|_2^2 + \|Y\|_2^2$$

And from the triangle inequality $\left\|\begin{bmatrix}X & Y\end{bmatrix}\right\|_2 \leq \|X\|_2 + \|Y\|_2$ so

$$\Psi(P) \leq \|V^{-1}\|_2\frac{\alpha\|P^2\|_F + \beta\|P\|_F + \gamma}{\|P\|_F} = \Phi(P) \qquad \text{cite Highan+Kim Davis}$$

Note that this is equivaalent to Higham and Kim (2001) and Davis (1981) up to the norm inequality: $\|X\|_2 \leq \|X\|_F$.

5.9. **Proof of Theorem 6 - Condition number: Linear equation 1.** Consider the perturbed equation where $FQ + D = 0$.

$$(F + \Delta F)(Q + \Delta Q) + D + \Delta D = 0$$

measuring perturbations normwise as

$$\epsilon = \max\left\{\frac{\|\Delta F\|_F}{\phi}, \frac{\|\Delta D\|_F}{\delta}\right\}$$

can be written to first order as

$$\begin{aligned}
F\Delta Q + \Delta F Q &= -\Delta G + \mathcal{O}(\epsilon^2) \\
F\Delta Q &= -\Delta F Q - \Delta G + \mathcal{O}(\epsilon^2) \\
&= -\begin{bmatrix}\Delta F & \Delta G\end{bmatrix}\begin{bmatrix} Q \\ I_{n_e}\end{bmatrix} + \mathcal{O}(\epsilon^2) \\
\Delta Q &= -F^{-1}\begin{bmatrix}\frac{\Delta F}{\phi} & \frac{\Delta G}{\delta}\end{bmatrix}\begin{bmatrix}\phi Q \\ \delta I_{n_e}\end{bmatrix} + \mathcal{O}(\epsilon^2) \\
\frac{\|\Delta Q\|_F}{\|Q\|_F} &\leq \left\|F^{-1}\right\|_F \frac{\phi\|Q\|_F + \delta}{\|Q\|_F}\sqrt{2}\epsilon + \mathcal{O}(\epsilon^2)
\end{aligned}$$

as $\|F\|_F \|Q\|_F \geq \|D\|_F$

$$\begin{aligned}
\frac{\|\Delta Q\|_F}{\|Q\|_F} &\leq \left\|F^{-1}\right\|_F \|F\|_F \left(\frac{\phi}{\|F\|_F} + \frac{\delta}{\|F\|_F\|Q\|_F}\right)\sqrt{2}\epsilon + \mathcal{O}(\epsilon^2) \\
&\leq \left\|F^{-1}\right\|_F \|F\|_F \left(\frac{\phi}{\|F\|_F} + \frac{\delta}{\|D\|_F}\right)\sqrt{2}\epsilon \\
\Theta(Q) &= \left\|F^{-1}\right\|_F \|F\|_F \left(\frac{\phi}{\|F\|_F} + \frac{\delta}{\|D\|_F}\right)\sqrt{2}\epsilon
\end{aligned}$$

Using the Kronecker / vectorized notation

$$\begin{aligned}
\left(I_{n_e} \otimes F\right)\operatorname{vec}(\Delta Q) &= -\begin{bmatrix}\phi Q' \otimes I_{n_y} & \delta I_{n_e n_y}\end{bmatrix}\begin{bmatrix}\phi^{-1}\operatorname{vec}(\Delta F) \\ \delta^{-1}\operatorname{vec}(\Delta D)\end{bmatrix} + O(\epsilon^2) \\
\operatorname{vec}(\Delta Q) &= -\begin{bmatrix}\phi Q' \otimes F^{-1} & \delta I_{n_e} \otimes F^{-1}\end{bmatrix}\begin{bmatrix}\phi^{-1}\operatorname{vec}(\Delta F) \\ \delta^{-1}\operatorname{vec}(\Delta D)\end{bmatrix} + O(\epsilon^2) \\
\frac{\|\Delta Q\|_F}{\|Q\|_F} &\leq \frac{\left\|\begin{bmatrix}\phi Q' \otimes F^{-1} & \delta I_{n_e} \otimes F^{-1}\end{bmatrix}\right\|_2}{\|Q\|_F}\sqrt{2}\epsilon + \mathcal{O}(\epsilon^2)
\end{aligned}$$

$$\begin{aligned}
\Psi(Q) &= \frac{\left\|\left(I_{n_e} \otimes F\right)^{-1}\begin{bmatrix}\phi Q' \otimes I_{n_y} & \delta I_{n_e n_y}\end{bmatrix}\right\|_2}{\|Q\|_F} \\
&\leq \frac{\left\|I_{n_e} \otimes F^{-1}\right\|_2 \left\|\begin{bmatrix}\phi Q' \otimes I_{n_y} & \delta I_{n_e n_y}\end{bmatrix}\right\|_2}{\|Q\|_F} \\
&= \frac{\left\|I_{n_e}\right\|_2 \left\|F^{-1}\right\|_2 \left\|\begin{bmatrix}\phi Q' \otimes I_{n_y} & \delta I_{n_e n_y}\end{bmatrix}\right\|_2}{\|Q\|_F}
\end{aligned}$$

5.10. **Proof of Corollary 3 - Condition number bound Linear equation 1.**

$$\begin{aligned}
\Psi(Q) &= \frac{\left\| \left(I_{n_e} \otimes F\right)^{-1} \left[\phi Q' \otimes I_{n_y} \quad \delta I_{n_e n_y}\right] \right\|_2}{\|Q\|_F} \\
&\leq \frac{\left\| I_{n_e} \otimes F^{-1} \right\|_2 \left\| \left[\phi Q' \otimes I_{n_y} \quad \delta I_{n_e n_y}\right] \right\|_2}{\|Q\|_F} \\
&= \frac{\left\| I_{n_e} \right\|_2 \left\| F^{-1} \right\|_2 \left\| \left[\phi Q' \otimes I_{n_y} \quad \delta I_{n_e n_y}\right] \right\|_2}{\|Q\|_F} \\
&\leq \left\| F^{-1} \right\|_2 \frac{\phi \|Q\|_F + \delta}{\|Q\|_F} = \Phi(Q)
\end{aligned}$$

5.11. **Proof of Theorem 7 - Condition number: Linear equation 2.** Consider the perturbed equation where $(AP+B)Q+D=0$

$$[(A+\Delta A)(P+\Delta P)+B+\Delta B](Q+\Delta Q)+D+\Delta D=0$$

Side calculations:

$$\begin{aligned}
\|AP+B\|_F \|Q\|_F &\geq \|D\|_F \\
\|A\|_F \|PQ\|_F + \|BQ\|_F &\geq \|D\|_F \\
APQ + BQ + D &= 0 \\
\|AP+B\|_F &\leq \|A\|_F \|P\|_F + \|B\|_F
\end{aligned}$$

measuring the perturbations normwise as

$$\epsilon = \max\left\{ \frac{\|\Delta P\|_F}{\xi}, \frac{\|\Delta A\|_F}{\alpha}, \frac{\|\Delta B\|_F}{\beta}, \frac{\|\Delta D\|_F}{\delta}, \right\}$$

can be written to first order as

$$\begin{aligned}
(AP+B)\Delta Q &= -A\Delta PQ - \Delta APQ - \Delta BQ - \Delta D + \mathcal{O}(\epsilon^2) \\
\Delta Q &= -(AP+B)^{-1}(A\Delta PQ + \Delta APQ + \Delta BQ + \Delta D) + \mathcal{O}(\epsilon^2) \\
\|\Delta Q\|_F &\leq \left\|(AP+B)^{-1}\right\|_F \Big(\|A\|_F \|Q\|_F \|\Delta P\|_F + \|PQ\|_F \|\Delta A\|_F \\
&\qquad + \|Q\|_F \|\Delta B\|_F + \|\Delta D\|_F \Big) + \mathcal{O}(\epsilon^2) \\
&\leq \left\|(AP+B)^{-1}\right\|_F \left(\xi \|A\|_F \|Q\|_F + \alpha \|PQ\|_F + \beta \|Q\|_F + \delta\right) \sqrt{4}\epsilon + \mathcal{O}(\epsilon^2) \\
\frac{\|\Delta Q\|_F}{\|Q\|_F} &\leq \left\|(AP+B)^{-1}\right\|_F \|AP+B\|_F \left(\frac{\xi \|A\|_F + \beta}{\|AP+B\|_F} + \frac{\alpha \|PQ\|_F}{\|Q\|_F \|AP+B\|_F} + \frac{\delta}{\|D\|_F} \right) \sqrt{4}\epsilon + \mathcal{O}(\epsilon^2) \\
&\leq \left\|(AP+B)^{-1}\right\|_F \|AP+B\|_F \left(\frac{\xi \|A\|_F + \alpha \|P\|_F + \beta}{\|AP+B\|_F} + \frac{\delta}{\|D\|_F} \right) \sqrt{4}\epsilon + \mathcal{O}(\epsilon^2)
\end{aligned}$$

This is wrong! From $\|AP+B\|_F \leq \|A\|_F \|P\|_F + \|B\|_F$ it follows

$$\begin{aligned}
\|B\|_F &\geq \|AP+B\|_F \\
\|A\|_F &\geq \frac{\|AP+B\|_F}{\|P\|_F} \\
\|P\|_F &\geq \frac{\|AP+B\|_F}{\|A\|_F}
\end{aligned}$$

$$\frac{\|\Delta Q\|_F}{\|Q\|_F} \leq \left\|(AP+B)^{-1}\right\|_F \|AP+B\|_F \left(\frac{\xi}{\|P\|_F}+\frac{\alpha}{\|A\|_F}+\frac{\beta}{\|B\|_F}+\frac{\delta}{\|D\|_F}\right)\cdot 2\cdot\epsilon+\mathcal{O}(\epsilon^2)$$

using the Kronecker / vectorized notation

$$\left(I_{n_e}\otimes(AP+B)\right)\mathrm{vec}(\Delta Q) = -\begin{bmatrix}\xi Q'\otimes A & \alpha(PQ)'\otimes I_{n_y} & \beta Q'\otimes I_{n_y} & \delta I_{n_e n_y}\end{bmatrix}\begin{bmatrix}\xi^{-1}\mathrm{vec}(\Delta P)\\ \alpha^{-1}\mathrm{vec}(\Delta A)\\ \beta^{-1}\mathrm{vec}(\Delta B)\\ \delta^{-1}\mathrm{vec}(\Delta D)\end{bmatrix}+\mathcal{O}(\epsilon^2)$$

$$\mathrm{vec}(\Delta Q) = \begin{bmatrix}\xi Q'\otimes(AP+B)^{-1}A & \alpha(PQ)'\otimes(AP+B)^{-1} & \beta Q'\otimes(AP+B)^{-1} & \delta I_{n_e}\otimes(AP+B)^{-1}\end{bmatrix}$$
$$\cdot\begin{bmatrix}\xi^{-1}\mathrm{vec}(\Delta P)\\ \alpha^{-1}\mathrm{vec}(\Delta A)\\ \beta^{-1}\mathrm{vec}(\Delta B)\\ \delta^{-1}\mathrm{vec}(\Delta D)\end{bmatrix}+\mathcal{O}(\epsilon^2)$$

$$\frac{\|\Delta Q\|_F}{\|Q\|_F} \leq \sqrt{4}\Psi(Q)\epsilon \leq \sqrt{4}\Phi(Q)\epsilon \leq \sqrt{4}\Theta(Q)\epsilon$$

$$\Psi(Q) = \frac{\left\|\left(I_{n_e}\otimes(AP+B)\right)^{-1}\begin{bmatrix}\xi Q'\otimes A & \alpha(PQ)'\otimes I_{n_y} & \beta Q'\otimes I_{n_y} & \delta I_{n_e n_y}\end{bmatrix}\right\|_2}{\|Q\|_F}$$

5.12. **Proof of Corollary 4 - Condition number bound Linear equation 2.**

$$\Psi(Q) = \frac{\left\|\left(I_{n_e}\otimes(AP+B)\right)^{-1}\begin{bmatrix}\xi Q'\otimes A & \alpha(PQ)'\otimes I_{n_y} & \beta Q'\otimes I_{n_y} & \delta I_{n_e n_y}\end{bmatrix}\right\|_2}{\|Q\|_F}$$
$$\leq \left\|\left(I_{n_e}\otimes(AP+B)\right)^{-1}\right\|_2 \frac{\xi\|Q\|_F\|A\|_F+\alpha\|PQ\|_F+\beta\|Q\|_F+\delta}{\|Q\|_F} = \Phi(Q)$$

5.13. **Proof of Theorem 8 - Condition number: Linear equation 3.** Consider the perturbed equation where $(AP+B)Q+D=0$

$$[(A+\Delta A)(P+\Delta P)+B+\Delta B](Q+\Delta Q)+D+\Delta D=0$$

and where $AP^2+BP+C=0$ and ΔP satisfies

$$(A+\Delta A)(P+\Delta P)^2+(B+\Delta B)(P+\Delta P)+C+\Delta C=0$$

measuring perturbations normwise as

$$\epsilon = \max\left\{\frac{\|\Delta A\|_F}{\alpha},\frac{\|\Delta B\|_F}{\beta},\frac{\|\Delta C\|_F}{\gamma},\frac{\|\Delta D\|_F}{\delta}\right\}$$

can be written to first order as

$$(AP+B)\Delta Q = -A\Delta PQ-\Delta APQ-\Delta BQ-\Delta D+\mathcal{O}(\epsilon^2)$$

and

$$(AP+B)\Delta P+A\Delta PP = -\Delta AP^2-\Delta BP-\Delta C+\mathcal{O}(\epsilon^2)$$

using the Kronecker / vectorized notation

$$\left(I_{n_e}\otimes(AP+B)\right)\mathrm{vec}(\Delta Q) = -\left(Q'\otimes A\right)\mathrm{vec}(\Delta P)$$

$$
\begin{aligned}
&- \left[\alpha(PQ)' \otimes I_{n_y} \quad \beta Q' \otimes I_{n_y} \quad 0_{n_{y^2}} \quad \delta I_{n_e n_y}\right] \\
&\cdot \begin{bmatrix} \alpha^{-1}\operatorname{vec}(\Delta A) \\ \beta^{-1}\operatorname{vec}(\Delta B) \\ \gamma^{-1}\operatorname{vec}(\Delta C) \\ \delta^{-1}\operatorname{vec}(\Delta D) \end{bmatrix} \\
&+ \mathcal{O}(\epsilon^2)
\end{aligned}
$$

$$
\begin{aligned}
\underbrace{\left(I_{n_e} \otimes (AP+B) + P' \otimes A\right)}_{V} \operatorname{vec}(\Delta P) = &- \left[\alpha(P^2)' \otimes I_{n_y} \quad \beta P' \otimes I_{n_y} \quad \gamma I_{n_{y^2}} \quad 0_{n_e n_y}\right] \\
&\cdot \begin{bmatrix} \alpha^{-1}\operatorname{vec}(\Delta A) \\ \beta^{-1}\operatorname{vec}(\Delta B) \\ \gamma^{-1}\operatorname{vec}(\Delta C) \\ \delta^{-1}\operatorname{vec}(\Delta D) \end{bmatrix} \\
&+ \mathcal{O}(\epsilon^2)
\end{aligned}
$$

$$
\begin{aligned}
\left(I_{n_e} \otimes (AP+B)\right)\operatorname{vec}(\Delta Q) = &- \Big[\alpha\left((PQ)' \otimes I_{n_y} - \left(Q' \otimes A\right)V^{-1}\left(P^2\right)' \otimes I_{n_y}\right) \quad \cdots \\
&\beta\left(Q' \otimes I_{n_y} - \left(Q' \otimes A\right)V^{-1}P' \otimes I_{n_y}\right) \quad \cdots \\
&-\gamma\left(Q' \otimes A\right)V^{-1} \quad \cdots \\
&\delta I_{n_e n_y}\Big] \qquad \begin{bmatrix} \alpha^{-1}\operatorname{vec}(\Delta A) \\ \beta^{-1}\operatorname{vec}(\Delta B) \\ \gamma^{-1}\operatorname{vec}(\Delta C) \\ \delta^{-1}\operatorname{vec}(\Delta D) \end{bmatrix} + \mathcal{O}(\epsilon^2)
\end{aligned}
$$

Define $T = I_{n_{y^2}} - \left(I_{n_y} \otimes A\right)V^{-1}\left(P' \otimes I_{n_y}\right)$

note $\left(I_{n_y} \otimes A \cdot (AP+B)^{-1}\right)V = I_{n_y} \otimes A + P' \otimes A(AP+B)^{-1}A = V\left(I_{n_y} \otimes (AP+B)^{-1}A\right)$

so $I_{n_y} \otimes (AP+B)^{-1}A = V^{-1}\left(I_{n_y} \otimes A \cdot (AP+B)^{-1}\right)V$

and $V = \left(I_{n_{y^2}} + P' \otimes \left[A(AP+B)^{-1}\right]\right)\left(I_{n_y} \otimes (AP+B)\right)$

so $I_{n_y} \otimes (AP+B)^{-1} = V^{-1} + V^{-1}\left(P' \otimes \left[A(AP+B)^{-1}\right]\right)$

Hence

$$
\begin{aligned}
\left(I_{n_y} \otimes (AP+B)^{-1}\right)T &= I_{n_y} \otimes (AP+B)^{-1} - \left(I_{n_y} \otimes (AP+B)^{-1}A\right)V^{-1}\left(P' \otimes I_{n_y}\right) \\
&= I_{n_y} \otimes (AP+B)^{-1} - V^{-1}\left(I_{n_y} \otimes A(AP+B)^{-1}\right)VV^{-1}\left(P' \otimes I_{n_y}\right) \\
&= V^{-1} + V^{-1}\left(P' \otimes A(AP+B)^{-1}\right) - V^{-1}\left(P' \otimes A(AP+B)^{-1}\right) = V^{-1}
\end{aligned}
$$

$$
\begin{aligned}
\operatorname{vec}(\Delta Q) = &- \Big[\alpha\left(Q' \otimes I_{n_y}\right)V^{-1}\left(P' \otimes I_{n_y}\right) \quad \cdots \\
&\beta\left(Q' \otimes I_{n_y}\right)V^{-1} \quad \cdots \\
&-\gamma\left(Q' \otimes (AP+B)^{-1}A\right)V^{-1} \quad \cdots
\end{aligned}
$$

$$\delta I_{n_e} \otimes (AP+B)^{-1}\Big] \begin{bmatrix} \alpha^{-1}\operatorname{vec}(\Delta A) \\ \beta^{-1}\operatorname{vec}(\Delta B) \\ \gamma^{-1}\operatorname{vec}(\Delta C) \\ \delta^{-1}\operatorname{vec}(\Delta D) \end{bmatrix} + \mathcal{O}(\epsilon^2)$$

$$\frac{\|\Delta Q\|_F}{\|Q\|_F} \leq \sqrt{4}\Psi(Q)\epsilon \leq \sqrt{4}\Phi(Q)\epsilon \leq \sqrt{(4)}\Theta(Q)\epsilon$$

$$\begin{aligned}\Psi(Q) = \Big\| \Big[& \alpha\left(Q' \otimes I_{n_y}\right) V^{-1}\left(P' \otimes I_{n_y}\right) \quad \cdots \\ & \beta\left(Q' \otimes I_{n_y}\right) V^{-1} \quad \cdots \\ & -\gamma\left(Q' \otimes (AP+B)^{-1}A\right) V^{-1} \quad \cdots \\ & \delta I_{n_e} \otimes (AP+B)^{-1}\Big] \Big\|_2 / \|Q\|_F\end{aligned}$$

5.14. **Proof of Corollary 5 - Condition number bound Linear equation 3.**

$$\begin{aligned}\Psi(Q) = \Big\| \Big[& \alpha\left(Q' \otimes I_{n_y}\right) V^{-1}\left(P' \otimes I_{n_y}\right) \quad \cdots \\ & \beta\left(Q' \otimes I_{n_y}\right) V^{-1} \quad \cdots \\ & -\gamma\left(Q' \otimes (AP+B)^{-1}A\right) V^{-1} \quad \cdots \\ & \delta I_{n_e} \otimes (AP+B)^{-1}\Big] \Big\|_2 / \|Q\|_F \\ \leq \left\|V^{-1}\right\|_2 & \frac{\alpha\|Q\|_F\|P\|_F + \beta\|Q\|_F + \gamma\|Q\|_F\left\|(AP+B)^{-1}A\right\|_2 + \delta\left\|(AP+B)^{-1}V\right\|_2}{\|Q\|_F} \\ \leq \left\|V^{-1}\right\|_2 & \frac{\alpha\|Q\|_F\|P\|_F + \beta\|Q\|_F}{\|Q\|_F} + \left\|V^{-1}\right\|_2\left\|(AP+B)^{-1}\right\|_2\frac{\gamma\|Q\|_F\|A\|_F}{\|Q\|_F} \\ & + \left\|(AP+B)^{-1}\right\|_2\frac{\delta}{\|Q\|_F} \\ = \Phi(Q)\end{aligned}$$

5.15. **Proof of Theorem 9 - Condition number: Joint P and Q.** Consider the perturbed equation

$$\begin{aligned}(A+\Delta A)(P+\Delta P)\left[P+\Delta P \quad Q+\Delta Q\right] \\ +(B+\Delta B)\left[P+\Delta P \quad Q+\Delta Q\right] \\ +\left[C+\Delta C \quad D+\Delta D\right] = \left[0 \quad 0\right]\end{aligned}$$

where

$$AP\left[P \quad Q\right] + B\left[P \quad Q\right] + \left[C \quad D\right] = \left[0 \quad 0\right]$$

measuring perturbations normwise as

$$\epsilon = \max\left\{\frac{\|\Delta A\|_F}{\alpha}, \frac{\|\Delta B\|_F}{\beta}, \frac{\|\Delta C\|_F}{\gamma}, \frac{\|\Delta D\|_F}{\delta}\right\}$$

can be written to first order as

$$(AP+B)\begin{bmatrix}\Delta P & \Delta Q\end{bmatrix}+A\begin{bmatrix}\Delta P & \Delta Q\end{bmatrix}\begin{bmatrix}P & Q\\ \underset{n_e\times n_y}{0} & \underset{n_e\times n_e}{0}\end{bmatrix}+(\Delta AP+\Delta B)\begin{bmatrix}P & Q\end{bmatrix}+\begin{bmatrix}\Delta C & \Delta D\end{bmatrix}=\begin{bmatrix}0 & 0\end{bmatrix}$$

$$\begin{aligned}(AP+B)\begin{bmatrix}\Delta P & \Delta Q\end{bmatrix}I_{n_y+n_e}+A\begin{bmatrix}\Delta P & \Delta Q\end{bmatrix}&\begin{bmatrix}P & Q\\ \underset{n_e\times n_y}{0} & \underset{n_e\times n_e}{0}\end{bmatrix}\\ &+I_{n_y}\Delta A\begin{bmatrix}P^2 & PQ\end{bmatrix}+I_{n_y}\Delta B\begin{bmatrix}P & Q\end{bmatrix}\\ &+I_{n_y}\Delta C\begin{bmatrix}I_{n_y} & \underset{n_y\times n_e}{0}\end{bmatrix}+I_{n_y}\Delta D\begin{bmatrix}\underset{n_e\times n_y}{0} & I_{n_e}\end{bmatrix}=\begin{bmatrix}0 & 0\end{bmatrix}\end{aligned}$$

a generalized Sylvester equation in $\begin{bmatrix}\Delta P & \Delta Q\end{bmatrix}$.

Using the Kronecker / vectorized notation

$$\underbrace{\left(I_{n_y+n_e}\otimes(AP+B)+\begin{bmatrix}P' & \underset{n_y\times n_e}{0}\\ Q' & \underset{n_e\times n_e}{0}\end{bmatrix}\otimes A\right)}_{=W}\begin{bmatrix}\operatorname{vec}(\Delta P)\\ \operatorname{vec}(\Delta Q)\end{bmatrix}$$

$$=-\begin{bmatrix}\begin{bmatrix}P^{2\prime}\\ Q'P'\end{bmatrix}\otimes I_{n_y} & \begin{bmatrix}P'\\ Q'\end{bmatrix}\otimes I_{n_y} & \begin{bmatrix}I_{n_y}\\ \underset{n_e\times n_y}{0}\end{bmatrix}\otimes I_{n_y} & \begin{bmatrix}\underset{n_y\times n_e}{0}\\ I_{n_e}\end{bmatrix}\otimes I_{n_y}\end{bmatrix}\begin{bmatrix}\operatorname{vec}(\Delta A)\\ \operatorname{vec}(\Delta B)\\ \operatorname{vec}(\Delta C)\\ \operatorname{vec}(\Delta D)\end{bmatrix}$$

$$=\underbrace{\begin{bmatrix}\alpha\begin{bmatrix}P^{2\prime}\\ Q'P'\end{bmatrix}\otimes I_{n_y} & \beta\begin{bmatrix}P'\\ Q'\end{bmatrix}\otimes I_{n_y} & \gamma\begin{bmatrix}I_{n_y}\\ \underset{n_e\times n_y}{0}\end{bmatrix}\otimes I_{n_y} & \delta\begin{bmatrix}\underset{n_y\times n_e}{0}\\ I_{n_e}\end{bmatrix}\otimes I_{n_y}\end{bmatrix}}_{=X}\begin{bmatrix}\alpha^{-1}\operatorname{vec}(\Delta A)\\ \beta^{-1}\operatorname{vec}(\Delta B)\\ \gamma^{-1}\operatorname{vec}(\Delta C)\\ \delta^{-1}\operatorname{vec}(\Delta D)\end{bmatrix}$$

$$\frac{\left\|\begin{bmatrix}\Delta P & \Delta Q\end{bmatrix}\right\|_F}{\left\|\begin{bmatrix}P & Q\end{bmatrix}\right\|_F}\leq\sqrt{4}\Psi(P,Q)\epsilon\leq\sqrt{4}\Phi(P,Q)\epsilon\leq\sqrt{4}\Theta(P,Q)\epsilon$$

$$\Psi(P,Q)=\frac{\left\|W^{-1}X\right\|_2}{\left\|\begin{bmatrix}P & Q\end{bmatrix}\right\|_F}$$

5.16. **Proof of Corollary 7 - Condition number bound joint.**

$$\begin{aligned}\Psi(P,Q)&=\frac{\left\|W^{-1}X\right\|_2}{\left\|\begin{bmatrix}P & Q\end{bmatrix}\right\|_F}\\ &\leq\left\|W^{-1}\right\|_2\frac{\alpha\left(\left\|P^2\right\|_2+\|QP\|_2\right)+\beta\left(\left\|P^2\right\|_2+\|Q\|_2\right)+\gamma+\delta}{\left\|\begin{bmatrix}P & Q\end{bmatrix}\right\|_F}=\Phi(P,Q)\end{aligned}$$

5.17. **Proof of Corollary 12 - A posterior forward error bound linear equation 3.** Set $\Delta A=\Delta B=0$, $\Delta C=R_P=A\hat{P}^2+B\hat{P}+C$, $\Delta D=R_Q=\left(A\hat{P}+B\right)\hat{Q}+D$

$$V\operatorname{vec}(\Delta P)=-\operatorname{vec}(R_P)$$

$$\left(I_{n_e}\otimes\left(A\hat{P}+B\right)\right)\operatorname{vec}(\Delta Q)=\left(\hat{Q}'\otimes A\right)V^{-1}\operatorname{vec}(R_P)-\operatorname{vec}(R_Q)$$

$$
\begin{aligned}
\operatorname{vec}(\Delta Q) &= \left(\hat{Q}' \otimes \left[\left(A\hat{P}+B\right)^{-1}A\right]\right)V^{-1}\operatorname{vec}(R_P) \\
&\quad -\left(I_{n_e}\otimes\left(A\hat{P}+B\right)^{-1}\right)\operatorname{vec}(R_Q) \\
\frac{\|\Delta Q\|_F}{\|\hat{Q}\|_F} &\le \left\|\left(\hat{Q}' \otimes \left[\left(A\hat{P}+B\right)^{-1}A\right]\right)V^{-1}\operatorname{vec}(R_P)\right. \\
&\quad \left. -\left(I_{n_e}\otimes\left(A\hat{P}+B\right)^{-1}\right)\operatorname{vec}(R_Q)\right\|_2 / \|\hat{Q}\|_F \\
&\le \left\|\left(A\hat{P}+B\right)^{-1}\right\|_2 \frac{\left\|\left(\hat{Q}'\otimes A\right)V^{-1}\operatorname{vec}(R_P)\right\|_2 + \|R_Q\|_F}{\|\hat{Q}\|_F} \\
&\le \left\|\left(A\hat{P}+B\right)^{-1}\right\|_2 \frac{\|R_Q\|_F}{\|\hat{Q}\|_F} \\
&\quad + \frac{\left\|\left(\hat{Q}'\otimes A\right)V^{-1}\operatorname{vec}(R_P)\right\|_2}{\|\hat{Q}\|_F}\left\|\left(A\hat{P}+B\right)^{-1}\right\|_2 \\
&\le \left\|\left(A\hat{P}+B\right)^{-1}\right\|_2 \frac{\|R_Q\|_F}{\|\hat{Q}\|_F} \\
&\quad + \left\|\left(A\hat{P}+B\right)^{-1}\right\|_2 \left\|V^{-1}\right\|_2 \|A\|_F \|R_P\|_F \\
&\le \left\|\left(A\hat{P}+B\right)^{-1}\right\|_2 \left(\frac{\|R_Q\|_F}{\|\hat{Q}\|_F} + \left\|V^{-1}\right\|_2 \|A\|_2 \|R_P\|_F\right)
\end{aligned}
$$

IMFS WORKING PAPER SERIES

Recent Issues

192 / 2023	Otmar Issing	On the importance of Central Bank Watchers
191 / 2023	Anh H. Le	Climate Change and Carbon Policy: A Story of Optimal Green Macroprudential and Capital Flow Management
190 / 2023	Athanasios Orphanides	The Forward Guidance Trap
189 / 2023	Alexander Meyer-Gohde Mary Tzaawa-Krenzler	Sticky information and the Taylor principle
188 / 2023	Daniel Stempel Johannes Zahner	Whose Inflation Rates Matter Most? A DSGE Model and Machine Learning Approach to Monetary Policy in the Euro Area
187 / 2023	Alexander Dück Anh H. Le	Transition Risk Uncertainty and Robust Optimal Monetary Policy
186 / 2023	Gerhard Rösl Franz Seitz	Uncertainty, Politics, and Crises: The Case for Cash
185 / 2023	Andrea Gubitz Karl-Heinz Tödter Gerhard Ziebarth	Zum Problem inflationsbedingter Liquiditätsrestriktionen bei der Immobilienfinanzierung
184 / 2023	Moritz Grebe Sinem Kandemir Peter Tillmann	Uncertainty about the War in Ukraine: Measurement and Effects on the German Business Cycle
183 / 2023	Balint Tatar	Has the Reaction Function of the European Central Bank Changed Over Time?
182 / 2023	Alexander Meyer-Gohde	Solving Linear DSGE Models with Bernoulli Iterations
181 / 2023	Brian Fabo Martina Jančoková Elisabeth Kempf Luboš Pástor	Fifty Shades of QE: Robust Evidence
180 / 2023	Alexander Dück Fabio Verona	Monetary policy rules: model uncertainty meets design limits
179 / 2023	Josefine Quast Maik Wolters	The Federal Reserve's Output Gap: The Unreliability of Real-Time Reliability Tests

178 / 2023	David Finck Peter Tillmann	The Macroeconomic Effects of Global Supply Chain Disruptions
177 / 2022	Gregor Boehl	Ensemble MCMC Sampling for Robust Bayesian Inference
176 / 2022	Michael D. Bauer Carolin Pflueger Adi Sunderam	Perceptions about Monetary Policy
175 / 2022	Alexander Meyer-Gohde Ekaterina Shabalina	Estimation and Forecasting Using Mixed-Frequency DSGE Models
174 / 2022	Alexander Meyer-Gohde Johanna Saecker	Solving linear DSGE models with Newton methods
173 /2022	Helmut Siekmann	Zur Verfassungsmäßigkeit der Veranschlagung Globaler Minderausgaben
172 / 2022	Helmut Siekmann	Inflation, price stability, and monetary policy – on the legality of inflation targeting by the Eurosystem
171 / 2022	Veronika Grimm Lukas Nöh Volker Wieland	Government bond rates and interest expenditures of large euro area member states: A scenario analysis
170 / 2022	Jens Weidmann	A new age of uncertainty? Implications for monetary policy
169 / 2022	Moritz Grebe Peter Tillmann	Household Expectations and Dissent Among Policymakers
168 / 2022	Lena Dräger Michael J. Lamla Damjan Pfajfar	How to Limit the Spillover from an Inflation Surge to Inflation Expectations?
167 / 2022	Gerhard Rösl Franz Seitz	On the Stabilizing Role of Cash for Societies
166 / 2022	Eva Berger Sylwia Bialek Niklas Garnadt Veronika Grimm Lars Other Leonard Salzmann Monika Schnitzer Achim Truger Volker Wieland	A potential sudden stop of energy imports from Russia: Effects on energy security and economic output in Germany and the EU
165 / 2022	Michael D. Bauer Eric T. Swansson	A Reassessment of Monetary Policy Surprises and High-Frequency Identification
164 / 2021	Thomas Jost Karl-Heinz Tödter	Reducing sovereign debt levels in the post-Covid Eurozone with a simple deficit rule