FUNDAÇÃO GETULIO VARGAS
ESCOLA de PÓS-GRADUAÇÃO em ECONOMIA

Guilherme Sohnlein Exel

Global Optimization of the CUE Objective Function by Eigenvalue Methods

Rio de Janeiro
2024

Guilherme Sohnlein Exel

Global Optimization of the CUE Objective Function by Eigenvalue Methods

Dissertação para obtenção do grau de mestre apresentada à Escola de Pós-Graduação em Economia

Área de concentração: Econometria

Orientador: Marcelo J. Moreira

Rio de Janeiro
2024

Dados Internacionais de Catalogação na Publicação (CIP)
Ficha catalográfica elaborada pelo Sistema de Bibliotecas/FGV

Exel, Guilherme Sohnlein
Global optimization of the CUE objective function by eigenvalue methods / Guilherme Sohnlein Exel. – 2024.

55 f.

Dissertação (mestrado) – Escola Brasileira de Economia e Finanças.
Orientador: Marcelo J. Moreira.
Inclui bibliografia.

1. Econometria – Processamento de dados. 2. Autovalores. 3. Metodo dos momentos (Estatistica). 4. Otimização matemática. I. Moreira, Marcelo J. II. Fundação Getulio Vargas. Escola Brasileira de Economia e Finanças. III. Título.

CDD – 330.015195

Elaborada por Marcelle Costal de Castro dos Santos – CRB-7-RJ-007517/O

Guilherme Sohnlein Exel

Global Optimization of the CUE Objective Function by Eigenvalue Methods

Dissertação para obtenção do grau de mestre apresentada à Escola de Pós-Grauação em Economia. Área de concentração: Econometria

E aprovado em 26/03/2024
Pela comissão organizadora

Prof. Dr.
Marcelo Jovita Moreira

Prof. Dr.
Marcelo Castello Branco Sant'Anna

Prof. Dr.
Whitney Kent Newey

Prof. Dr.
Mahrad Sharifvaghefi

Acknowledgments

Lengthy discussions with my advisor, Marcelo J. Moreira, and Luan Borelli were extremely important for this work. This thesis is part of a collaborative work with them, Mahrad Sharifvaghefi, and Whitney Newey. This study was financed in part by the Getulio Vargas Foundation (FGV), the University of Pittsburgh, and the Coordenação de Aperfeiçoamento de Pessoal de Nível Superior, Brazil (CAPES) — Finance Code 001.

Resumo

Desde sua concepção em Hansen et al. (1996), o Estimador de GMM de Atualização Contínua (CUE) tem representado um desafio para os métodos de otimização numérica devido à estrutura de sua função objetivo. Mesmo em modelos lineares simples, ela pode apresentar múltiplos mínimos locais e seções planas que frequentemente derrotam tanto os métodos de descida de gradiente quanto os de busca em grade. Esse problema de longa data na literatura foi recentemente resolvido por Moreira et al. (2023). Construindo sobre este resultado, propomos um método alternativo que é simples de implementar e mantém garantias semelhantes de produzir resultados globalmente ótimos. Mostramos como as condições de primeira ordem da função objetivo do CUE podem ser reformuladas na forma de um problema de autovalores e especializadas para casos em que algoritmos eficientes estão disponíveis.

Palavras-chave: Estimador de atualização contínua, Método generalizado dos momentos, Problemas de autovalores não lineares, Problemas de autovalores multiparamétricos.

Abstract

Since its inception in Hansen et al. (1996), the Continuously Updating GMM Estimator (CUE) has posed a challenge for numerical optimization methods due to the structure of its criterion function. Even in simple linear models, it can present multiple local minima and flat sections which often defeat both gradient descent and grid search methods. This longstanding problem in the literature has recently been solved by Moreira et al. (2023). We build upon their insight and propose an alternative method which is simple to implement and retains similar guarantees of producing globally optimal results. We show how the first order conditions of the CUE criterion function can be restated in the form of an eigenvalue problem and specialize to cases for which efficient algorithms are available.

Keywords: Continuously updating estimator, Generalized method of moments, Nonlinear eigenvalue problems, Multiparameter eigenvalue problems.

List of Figures

Contents

1 Introduction

Since its introduction in Hansen (1982b), the Generalized Method of Moments (GMM) has become a cornerstone in econometrics. It provides both versatility in accommodating a wide range of models and robustness by requiring only mild assumptions on the underlying data process. In its agnosticism towards the probability distribution of the data, the GMM offers the further advantage of alleviating the computational burden of estimation. While Maximum Likelihood Estimation may require cumbersome numerical integration, the GMM objective function is a quadratic form over vector-valued functions that are often easy to compute and present nice mathematical properties.

The GMM works by considering equations that are predicted by theory to hold in expectation and selecting parameters that match the equations in the sample to their population counterparts. Whenever there are more equations than parameters, not all equations can be satisfied at once. It is then necessary to consider a measure of how closely the sample equations are satisfied at each parameter value. Hansen (1982b) find that the appropriate measure is itself dependent on the parameters being estimated, which creates a kind of circularity. The recommended procedure it to obtain a preliminary estimate of the parameter which is then used as an input for a second step of estimation. While consistency of the first step estimator is enough to achieve asymptotic efficiency in the second step, concerns about the finite sample performance of a poor preliminary estimator motivate refinements of this procedure. Hansen et al. (1996) consider a simultaneous estimation procedure which eschews a preliminary estimate and allows both the sample equations and the measure of how closely they are satisfied to vary with the parameter, which requires only a single step of estimation. This is the Continuously Updating Estimator (CUE), which is the object of our discussion.

The CUE is of particular interest in the theory of weak identification. When the population moment equations are nearly satisfied in a neighborhood of the true parameter, the parameter is said to be weakly identified. While this does not affect consistency of estimators, it invalidates inferences based on usual asymptotic approximations. A vast and growing literature on weak identification finds that the CUE plays a crucial role in the development of correctly sized tests and estimation of confidence intervals when instruments are weak.

In a linear instrumental variables setting under the assumption of conditional heteroskedasticity, the CUE reduces to the Limited Information Maximum Likelihood (LIML) estimator of Anderson and Rubin (1949). The objective function of the LIML evaluated at the true parameter can be shown to be asymptotically chi-square distribution regardless of the strength of the instruments, which is the basis for the fully robust AR test. In the single-instrument case, Moreira (2009a) shows that the AR test is uniformly most powerful unbiased, but its power decays quickly as the number of instruments increases. The Conditional Likelihood Ratio test of Moreira (2003) builds upon the AR and has been shown to have very good power properties in the over-identified case. The CUE is plays a crucial role in subsequent work that consider extensions to either the nonlinear or heteroskedastic cases. A non-comprehensive list of such works include Stock and Wright (2000), Kleibergen (2005), Andrews and Mikusheva (2016) and Moreira and Moreira (2019).

Newey and Smith (2004) find that the CUE is part of the family of Generalized Empirical

Likelihood (GEL) estimators, which offer an alternative method of estimation for models defined by moment conditions. While the GMM searches for parameters that approximately satisfy the moment conditions, GEL estimators reweigh the data and ensure that all moment conditions are satisfied exactly. It subsumes a number of other estimators as special cases, such as the Empirical Likelihood estimator of Owen (1988); Qin and Lawless (1994); Imbens (1993) and the Exponential Tilting estimator of Kitamura and Stutzer (1997); Imbens et al. (1998a). GEL estimators minimize a measure of the discrepancy between the empirical distribution of the data and the distribution implied by the weights subject to the moment conditions being satisfied. The simultaneous estimation of weights and parameters gives rise to a high dimensional constrained minimization problem where the dimension exceeds the sample size, which can be intractable in large samples. Smith (1997) shows that GEL estimators can be equivalently obtained from an unconstrained saddle point problem where the number of auxiliary parameters is equal to the number of moment conditions, which substantially simplifies the computation of estimates. In the particular case of the CUE, these auxiliary parameters can be concentrated out and the saddle-point problem simplifies to the problem of minimizing the CUE objective function introduced in Hansen et al. (1996).

GEL estimators provide a welcome alternative to the standard GMM for multiple reasons. For one, they all share the same invariance property of not having to rely on a choice of weight matrix to be used in a first step, which affects the numerical value of the final estimates. While this choice has no bearing on the asymptotic distribution, it introduces an ambiguity in the estimation procedure. Furthermore, while the two-step estimator has been shown by Chamberlain (1987) to achieve the semiparameteric efficiency bound, a number of studies suggest it may present severe bias in small samples and produce confidence intervals with poor convergence rates (Burnside and Eichenbaum, 1996; Hall and Horowitz, 1996). In particular, Altonji and Segal (1996) suggest that even an inefficient first step estimator using the identity matrix may perform better in small samples than the two-step estimator. The poor performance of two-step procedures in finite samples is found to be related to degree of overidentification, which has motivated an extensive literature that considers alternative asymptotic approximations in which the the number of moment conditions is allowed to increase with the sample size (Han and Phillips, 2006; Chao and Swanson, 2005; Newey and Windmeijer, 2009). Newey and Smith (2004) show that the asymptotic bias of the Empirical Likelihood estimator does not grow with the number of moment conditions, and find that the CUE as well as other GEL estimators share this property under a symmetry condition. Imbens et al. (2010) also find the CUE to have lower higher-order variance than the bias-corrected GMM estimator. Standard errors for the CUE that account for many moments are provided in Newey (2004).

In spite of its attractive properties, the CUE has seen little use in practice. The reason is twofold: firstly, simulation evidence in Hansen et al. (1996) and subsequent studies (Hausman et al., 2011; Mittelhammer et al., 2005; Hausman et al., 2012) show the CUE tends to have higher dispersion compared to alternative estimators, which discourages its use in applications. Secondly, the CUE objective function is much less well behaved than that of estimators that employ fixed weight matrices, as the quadratic structure of the function is lost when the weight matrices also depend on the parameter. This complicates numerical optimization, since local search

algorithms are sensitive initial solutions and susceptible to getting stuck on local minima, as well as failing to converge over flat sections of the function. Hansen et al. (1996) report evaluating the objective function over a grid to check whether unusually large parameter estimates were caused by the local search algorithm spuriously searching towards an asymptote. Imbens et al. (1998b) also report problems in getting the CUE to converge in simulations, and caution for the presence of multiple local minima in the objective function. The recent textbook of Hansen (2022) emphasizes the computational challenge of minimizing the CUE objective function and observes that it is not commonly employed in applications.

Recently, Moreira et al. (2023) have proposed a novel method for global optimization of the CUE objective function. Their method leverages the particular structure of the function to obtain all solutions to the first order conditions and guarantee that a globally optimal parameter estimate is found. To our knowledge, that is the first time that a method with such guarantees has been employed in this literature, and defeats the longstanding problem of correct computation of the CUE. In this work, we leverage their insight and propose an alternative method that circumvents computationally intensive steps such as symbolic matrix inversion. We show how the first order conditions of the CUE objective function can be restated as a nonlinear eigenvalue problem and efficiently solved via well established numerical algorithms. To achieve this, we employ the same auxiliary parameters that are present in the GEL characterization of the CUE. We show how the auxiliary parameters can be interpreted as eigenvectors, while the parameters of interest are interpreted as eigenvalues.

The standard eigenvalue problem is ubiquitous in the whole of Economics, and the generalized eigenvalue problem is familiar to econometricians from its role in the computation of the LIML. However, further generalizations of eigenvalue problems are virtually absent in the econometrics literature. Problems where eigenvalues appear nonlinearly have been garnering increasing attention in physics and engineering, where they arise in areas such as fluid mechanics (Voss, 2003), acoustics (Mehrmann and Schröder, 2011) and materials science (Singh and Ram, 2002). Problems with multiple eigenvalues appearing simultaneously in the same equations are also known to arise in the study of boundary value problems for partial differential equations (Källström and Sleeman, 1976). This dissertation adds the CUE to the growing list of applications of such problems.

Our method closely resembles the approach of Vermeersch and Moor (2019), that show how the Least Squares estimation of ARMA models can be stated as a nonlinear multiparameter eigenvalue problem. While novel and theoretically appealing, their approach presents two disadvantages in the context of ARMA models which are not present in our application to the CUE. Firstly, they only consider models of order ARMA(1,1) or higher as their approach is of little interest in single parameter case. This is due to how the AR and the MA parameters appear in different parts of the Least-Squares objective function. For the CUE, we find that our method is especially suited for the single-parameter linear case, which is of great practical interest as it finds an application in linear instrumental variables models with a single endogenous regressor. Secondly, the dimensions of the matrices in the ARMA model grow with the sample size, which when combined with the burden of estimating multiple parameters only allow for estimation over very small datasets. This prompts them to not recommend their method for most applications.

In contrast, matrix dimensions in the CUE do not depend on sample size, as they only grow with the number of moment conditions. In the single-parameter linear case, we are able to compute precise estimates for samples of any size and multiple hundreds of moment conditions in seconds. We can therefore confidently recommend our method for the single-parameter case. While still resource-intensive for estimating multiple parameters, it proves to be be viable for models of moderate size.

The structure of this work is as follows: Section 2 motivates the CUE and presents our main result of restating the first order conditions of the CUE objective function as an eigenvalue problem. Section 3 specializes our method to the single-parameter case and discusses the computation of confidence sets based on the CUE objective. Section 3.4 extends the result of Section 2 to the problem of computing criterion-based confidence intervals. Section 4 presents our method for the multiparameter case. Section 5 concludes and presents directions for future work on the general nonlinear case.

2 The Continuously Updating Estimator

Consider a sample z_i, $i = 1, \ldots, N$ of strictly stationary and ergodic observations on a data vector z. Define the parameter $\theta_0 \in \mathbb{R}^p$ as the unique solution to the $m > p$ moment conditions

$$\mathbb{E}[g(z_i, \theta_0)] = 0, \tag{2.1}$$

where $g(z, \theta)$ is a vector of known moment functions. For notational convenience, define $g_i(\theta) = g(z_i, \theta)$. Denote the sample analog of the left hand side of (2.1) by $\hat{g}(\theta) = N^{-1} \sum_i^N g_i(\theta)$, where the dependence on the sample size N is omitted.

A GMM estimator of θ_0 minimizes

$$Q(\theta) = \hat{g}(\theta)' \hat{W} \hat{g}(\theta) \tag{2.2}$$

for some positive semidefinite weight matrix $\hat{W}$ such that $\hat{W} \xrightarrow{p} W$ positive definite. While consistency of such estimators is achieved for arbitrary weight matrices under mild conditions, asymptotic efficiency requires $W = \Omega(\theta_0)^{-1}$, where

$$\Omega(\theta_0) = \operatorname{avar}(N^{\frac{1}{2}} \hat{g}(\theta_0)) = \lim_{N \to \infty} \operatorname{var} \left(N^{-\frac{1}{2}} \sum_i^N g_i(\theta_0) \right) = \lim_{N \to \infty} N^{-1} \sum_i^N \sum_j^N \mathbb{E} \left[g_i(\theta_0) g_j(\theta_0)' \right] \tag{2.3}$$

is the asymptotic variance of the moment conditions. A heteroskedasticity and autocorrelation consistent (HAC) estimator of matrix $\Omega(\theta_0)$ is given by

$$\hat{\Omega}^{\text{HAC}}(\theta) = \hat{\Gamma}_0 + \sum_i^{N-1} \omega_i (\hat{\Gamma}_i + \hat{\Gamma}_i'), \qquad \text{where} \quad \hat{\Gamma}_i = N^{-1} \sum_{j=i+1}^N g_i(\theta) g_{i-j}(\theta)' \tag{2.4}$$

and ω_i are kernel weights such as the Bartlett weights in Newey and West (1987). Although the methods developed in this work can be easily applied to (2.4), we assume i.i.d. data to simplify exposition. With $\mathbb{E}[g_i(\theta_0) g_i(\theta_0)'] = \mathbb{E}[g_j(\theta_0) g_j(\theta_0)']$ and $\mathbb{E}[g_i(\theta_0) g_j(\theta_0)'] = 0$ for $i \neq j$, (2.3) simplifies to $\Omega(\theta_0) = \mathbb{E}\left[g_i(\theta_0) g_i(\theta_0)' \right]$, which can be estimated by

$$\hat{\Omega}(\theta) = N^{-1} \sum_i^N g_i(\theta) g_i(\theta)'. \tag{2.5}$$

This introduces a complication due to $\hat{\Omega}(\theta)$ being a function of the unknown parameter we wish to estimate. One alternative is to perform a two-step estimation procedure: start with a preliminary consistent estimator $\hat{\theta}^{(1)}$ obtained by setting $\hat{W}$ to some fixed positive definite matrix (commonly, the identity matrix) and proceed by estimating the parameter a second time, where $\hat{\theta}^{(1)}$ is used for a plug-in estimator of the asymptotic variance:

$$\hat{\theta}^{(2)} = \arg\min_\theta \hat{g}(\theta)' \hat{\Omega} \left(\hat{\theta}^{(1)} \right)^{-1} \hat{g}(\theta). \tag{2.6}$$

This is known as the two-step GMM estimator. While the first step estimator is likely inefficient, it is only required to be consistent in order to obtain $\hat{\Omega}(\hat{\theta}^{(1)}) \xrightarrow{p} \Omega(\theta_0)$ under regularity conditions. As a consequence, this simple two-step procedure is enough to achieve asymptotic

efficiency (Chamberlain, 1987). Since an inefficient estimator is used in the first step simply because an efficient one is not available, it is natural to consider performing a third step by using $\hat{W} = \hat{\Omega}(\hat{\theta}^{(1)})^{-1}$. The probability limit of $\hat{\Omega}(\theta)$ evaluated at any consistent estimator of θ_0 will be the same, so a third step will not change the asymptotic distribution of the estimator. However, we may expect that it will have a better performance in finite samples. This procedure could be iterated any number of times, each step using the estimator from the previous step:

$$\hat{\theta}^{(k+1)} = \arg\min_{\theta} \hat{g}(\theta)' \hat{\Omega} \left(\hat{\theta}^{(k)}\right)^{-1} \hat{g}(\theta). \tag{2.7}$$

When either a set number of iterations is reached or some measure of convergence is achieved, the iteration ends and the last estimator in the sequence defines the Iterated GMM estimator. Hansen et al. (1996) carry this idea to its logical conclusion with the Continuously Updating Estimator (CUE), in which the estimator of the asymptotic variance depends directly on the argument of the objective function. They define

$$\hat{\theta}_{\text{CUE}} = \arg\min_{\theta} \ \ Q(\theta), \tag{2.8}$$

where

$$Q(\theta) = \hat{g}(\theta)' \hat{\Omega}(\theta)^{-1} \hat{g}(\theta). \tag{2.9}$$

Despite the apparent similarity of (2.9) with the objective functions in (2.6) and (2.7), it can have a very different shape and be more challenging for numerical optimization. The two-step objective function is a quadratic form over the sample moments, which makes it usually unbounded and often convex. Meanwhile, the CUE objective function is always bounded above by 1. To see this, let $G(\theta)$ be an $N \times m$ matrix that stacks $g_i(\theta)'$ over i. We have $\hat{g}(\theta) = N^{-1} G(\theta)' 1_N$ where 1_N denotes the $N \times 1$ vector of ones and $\hat{\Omega}(\theta) = N^{-1} G(\theta)' G(\theta)$. Then,

$$Q(\theta) = N^{-1} 1_N' G(\theta) (G(\theta)' G(\theta))^{-1} G(\theta)' 1_N = N^{-1} 1_N' P_{G(\theta)} 1_N \leq N^{-1} 1_N' 1_N = 1.$$

Where $P_{G(\theta)}$ denotes the projection matrix of $G(\theta)$ and the inequality uses the fact that its largest eigenvalue is equal to 1. This shows that a (nonconstant) CUE objective function can never be convex, since a function being bounded and convex in $\mathbb{R}^p$ implies that it is constant. While it can still be quasi-convex –which is sufficient for Newton's method to converge to the global minimum– that is often not the case, as illustrated in the following example.

Example 1 (Instrumental variables estimation)**.** Consider the linear instrumental variables model:

$$\begin{aligned} y_{1i} &= y_{2i}' \beta_0 + u_i \\ y_{2i} &= \Pi_0' z_i + v_i, \qquad i = 1, \ldots, N, \end{aligned} \tag{2.10}$$

where y_{1i} is a scalar endogenous variable, y_{2i} is a $p \times 1$ vector of endogenous regressors and z_i is an $m \times 1$ vector of instruments. The exogeneity condition $\mathbb{E}[u_i z_i] = \mathbb{E}[(y_{1i} - y_{2i}' \beta_0) z_i] = 0$ suggests the estimation of β_0 by GMM with the moment function $g_i(\beta) = (y_{1i} - y_{2i}' \beta) z_i$, which is linear in the parameters.

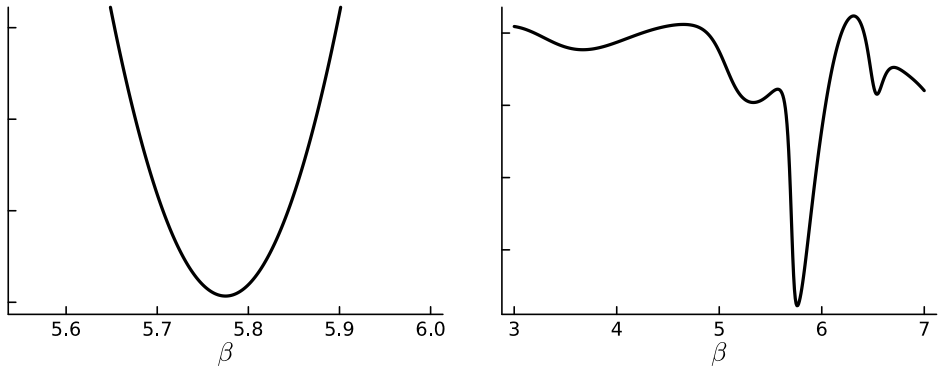
Denote $y_i = (y_{1i}, y_{2i}')'$ and let $b = (1, -\beta')'$. We can write $g_i(\beta) = z_i\, y_i' b$ and

$$\hat{\Omega}(\beta) = N^{-1} \sum_i (y_i' b)^2 z_i z_i' = (b' \otimes I_m)\Sigma(b \otimes I_m), \qquad \text{where} \quad \Sigma \coloneqq N^{-1} \left(\sum_i y_i y_i' \otimes z_i z_i' \right). \tag{2.11}$$

Let Z and Y be matrices that vertically stack observations z_i' and y_i' over i, respectively. We can write the sample moments as $\hat{g}(\theta) = N^{-1} Z'Yb$ and the CUE objective function in (2.9) simplifies to

$$Q(\beta) = b'Y'Z \left((b' \otimes I_m)\Sigma(b \otimes I_m) \right)^{-1} Z'Yb. \tag{2.12}$$

A two-step estimator minimizes $b'Y'Z\hat{W}Z'Yb$ for constant $\hat{W}$, which is a simple convex quadratic form in b. Since $(b \otimes I_m)\Sigma(b \otimes I_m)$ is a positive semidefinite quadratic matrix function of β, we may expect (2.12) to somewhat resemble a quadratic function. However, that is often not the case. Figure 1 depicts objective functions of the two-step and the CUE for scalar β. Both are obtained from a typical outcome of the simulation design we later describe in Section 3.2.



(a) Two-step (b) CUE

Figure 1: Objective functions for different GMM estimators.

While the minimizer of the function in Figure 1a can be found analytically as the unique solution to its first order condition, the first order condition of the function in Figure 1b has no analytical solution and the multiplicity of local minima suggests local search algorithms may be inadequate.

Example 2 (Dynamic panel data)**.** Consider a panel data model with individual fixed effects given by the equation

$$y_{i,t} = \rho_0 y_{i,t-1} + \eta_{0,i} + u_{i,t}, \qquad \text{for } i = 1, \ldots, N \quad \text{and } t = 1, \ldots, T, \tag{2.13}$$

where $y_{i,t}$ are observable variables and $u_{i,t}$ are unobservable errors with mean zero given $y_{i,t-1}, \ldots, y_{i,1}$. We are interested in estimating the parameter ρ_0, while the fixed effects $\eta_{0,i}$ are incidental parameters. Following Arellano and Bond (1991), we can obtain moment conditions that do not depend on $\eta_{0,i}$ by considering the first difference of (2.13) in t:

$$y_{i,t} - y_{i,t-1} = \rho_0 \left(y_{i,t-1} - y_{i,t-2} \right) + u_{i,t} - u_{i,t-1}. \tag{2.14}$$

Parameter ρ_0 can be estimated from (2.14) by using lags as instruments, which yields moment functions

$$g_t(\rho) = \begin{pmatrix} y_{i,t-2} \\ \vdots \\ y_{i,1} \end{pmatrix} \left(y_{i,t} - y_{i,t-1} - \rho \left(y_{i,t-1} - y_{i,t-2} \right) \right). \tag{2.15}$$

Note that the size of vector $g_t(\rho)$ depends on t. Since there are $t-2$ lags available for each $t = 3, \ldots, T$, we obtain $(T-1)(T-2)/2$ moment conditions in total, which is potentially very large. As previously discussed, a large degree of overidentification may make standard GMM procedures unsuitable for the estimation of ρ_0.

Parameter ρ_0 can also be estimated using moment conditions based on the second moments of $y_{i,t}$ (Chamberlain, 1984). Suppose we have the initial conditions $y_{i,0} = 0$ and assume $u_{i,t}$ is i.i.d. with variance σ_0^2. We can write (2.14) in vector form by stacking observations over t:

$$y_i = \rho_0 J y_i + \eta_{0,i} 1_T + u_i, \qquad \text{where} \quad J = \begin{bmatrix} 0 & 0 \\ I_{T-1} & 0 \end{bmatrix}.$$

Denote the sample variance matrix of y_i by

$$\hat{S}_N = (N-1)^{-1} \sum_i^N (y_i - \overline{y})(y_i - \overline{y})', \qquad \text{where } \overline{y} = N^{-1} \sum_i^N y_i.$$

From $y_i = B_0(\eta_{0,i} 1_T + u_i)$ with $B_0 = (I_T - \rho_0 J)^{-1}$, we obtain $\mathbb{E}[\hat{S}_N] = B_0(\lambda_{N,0} 1_T 1_T' + I_T \sigma_0^2) B_0'$, where $\lambda_{N,0} = (N-1)^{-1} \sum_i^N (\eta_{0,i} - \overline{\eta}_0)^2 / \sigma_0^2$ and $\overline{\eta}_0 = N^{-1} \sum_i^N \eta_{0,i}$. We can therefore consider estimation of the parameter vector $\theta_0 = (\rho_0, \lambda_{N,0}, \sigma_0^2)'$ from

$$\hat{g}(\theta) = \operatorname{vech}(\hat{S}_N) - \sigma^2 \operatorname{vech}(B(\lambda_N 1_T 1_T' + I_T) B'), \tag{2.16}$$

where $\operatorname{vech}(\cdot)$ denotes the vectorization of the lower triangular part of a matrix. While this approach does not completely eliminate incidental parameters, it reduces the number of incidental parameters from a vector of size N to a single scalar. Using an argument of invariance under orthogonal transformations, Moreira (2009b) shows that this approach preserves information in the data that is potentially wasted by only considering first differences. Note that a GMM estimator applied to (2.16) also fits the class of minimum distance estimators.

2.1 CUE first order conditions as an eigenvalue problem

In this section, we show how the first order conditions of the CUE objective function can be restated in the form of an eigenvalue problem. Our approach involves the introduction of auxiliary variables and is motivated by the characterization of the CUE as a special case of the GEL estimator of Smith (1997). Newey and Smith (2004) obtain this equivalence by concentrating out the auxiliary variables and observing that the concentrated GEL objective yields the objective function in (2.9). We leverage linearities that are present in this formulation and show how

the auxiliary variables can be interpreted as an eigenvector and $\hat{\theta}_{\text{CUE}}$ can be interpreted as an eigenvalue.

To motivate or method, we briefly present the GEL and discuss how the CUE can be obtained as a special case. The GEL solves the saddle point problem

$$\hat{\theta}_{\text{GEL}} \coloneqq \arg\min_{\theta} \sup_{x \in X(\theta)} N^{-1} \sum_{i}^{N} \rho(x' g_i(\theta)), \tag{2.17}$$

where the set $X(\theta)$ constrains x to values for which $x' g_i(\theta)$ is in the domain of ρ. Different variations of the GEL employ different specifications for ρ. The Empirical Likelihood estimator (Qin and Lawless, 1994; Owen, 1988) is obtained for $\rho(v) = \ln(1 - v)$, and the Exponential Tilting estimator (Imbens et al., 1998a; Kitamura and Stutzer, 1997) is obtained for $\rho(v) = -\exp(v)$. Newey and Smith (2004) show that the CUE is also a special case of the GEL for $\rho(v) = -(1 + v)^2/2$.

Solving (2.17) numerically may be challenging since it is a saddle point problem with a parameter space of dimension $p + m$, which is potentially very large. Both of these complications are eliminated if x can be concentrated out, which is possible when ρ is quadratic. We can write the GEL objective as

$$N^{-1} \sum_{i}^{N} -(1 + x' g_i(\theta))^2 = -1 + 2x' \hat{g}(\theta) - x' \hat{\Omega}(\theta) x.$$

It is easy to see that the first order conditions with respect to x yield $x = -\hat{\Omega}(\theta)^{-1} \hat{g}(\theta)$ and the concentrated GEL objective simplifies to $Q(\theta) - 1$. To restate the first order conditions $\nabla Q(\theta) = 0$ as an eigenvalue problem, we similarly define[1]

$$x(\theta) \coloneqq \hat{\Omega}(\theta)^{-1} \hat{g}(\theta) \tag{2.18}$$

and write $Q(\theta) = \hat{g}(\theta)' x(\theta)$. Then,

$$\begin{aligned} \nabla Q(\theta) &= (\mathsf{D}\, \hat{g}(\theta))'\, x(\theta) + (\mathsf{D}\, x(\theta))'\, \hat{g}(\theta) \\ &= (\mathsf{D}\, \hat{g}(\theta))'\, x(\theta) + (I_p \otimes \hat{g}(\theta)')\, \operatorname{vec}(\mathsf{D}\, x(\theta)), \end{aligned} \tag{2.19}$$

where the D operator denotes the Jacobian. From (2.18),

$$\mathsf{D}\, x(\theta) = \hat{\Omega}(\theta)^{-1} (\mathsf{D}\, \hat{g}(\theta)) - \left(\hat{\Omega}(\theta)^{-1} \otimes \hat{g}(\theta)' \hat{\Omega}(\theta)^{-1}\right) \left(\mathsf{D} \operatorname{vec}(\hat{\Omega}(\theta))\right)$$

and we can write

$$\operatorname{vec}(\mathsf{D}\, x(\theta)) = (I_p \otimes \hat{\Omega}(\theta)^{-1}) \left(\operatorname{vec}(\mathsf{D}\, \hat{g}(\theta)) - \left(\mathsf{D}\, \hat{\Omega}(\theta)\right) x(\theta)\right), \tag{2.20}$$

where, with some abuse of notation, we define the $mp \times m$ matrix $\mathsf{D}\, \hat{\Omega}(\theta)$ as a reshaping of the $m^2 \times p$ Jacobian matrix $\mathsf{D} \operatorname{vec}(\hat{\Omega}(\theta))$ that vertically stacks $\frac{\mathrm{d}}{\mathrm{d}\theta_j} \hat{\Omega}(\theta)$ over $j = 1, \ldots, m$.

[1] $x(\theta)$ is defined with a positive sign for convenience.

The definition of $x(\theta)$ in (2.18) together with (2.19) and (2.20) yield the following system of $p + m + pm$ equations:

$$\left(\mathsf{D}\,\hat{\Omega}(\theta)\right) x(\theta) + (I_p \otimes \hat{\Omega}(\theta))\,\mathrm{vec}\,(\mathsf{D}\,x(\theta)) - \mathrm{vec}\,(\mathsf{D}\,\hat{g}(\theta)) = 0, \tag{2.21}$$

$$\hat{\Omega}(\theta)x(\theta) - \hat{g}(\theta) = 0, \tag{2.22}$$

$$(\mathsf{D}\,\hat{g}(\theta))'\,x(\theta) + (I_p \otimes \hat{g}(\theta)')\,\mathrm{vec}\,(\mathsf{D}\,x(\theta)) = 0. \tag{2.23}$$

Since this system is linear in $x(\theta)$ and $\mathrm{vec}(\mathsf{D}\,x(\theta)$, we write it in matrix form as

$$\begin{bmatrix} \mathsf{D}\,\Omega(\theta) & I_p \otimes \hat{\Omega}(\theta) & \mathrm{vec}(\mathsf{D}\,\hat{g}(\theta)) \\ \hat{\Omega}(\theta) & 0 & \hat{g}(\theta) \\ \mathsf{D}\,\hat{g}(\theta)' & I_p \otimes \hat{g}(\theta)' & 0 \end{bmatrix} \begin{pmatrix} x(\theta) \\ \mathrm{vec}\,(\mathsf{D}\,x(\theta)) \\ -1 \end{pmatrix} = 0. \tag{2.24}$$

Denote the $(m(p+1)+p) \times (m(p+1)+1)$ matrix on the left hand side of (2.24) by $L(\theta)$. We obtain that any solution in θ to (2.24) will also be a solution to

$$L(\theta)v = 0, \qquad v \neq 0. \tag{2.25}$$

for some $v \in \mathbb{R}^{m(p+1)+1}$. The only possible solutions to (2.25) that are not also solutions to (2.24) must be such that the last component of v equals zero. Examination of equations (2.22) and (2.21) reveal that this can only happen when $\hat{\Omega}(\theta)$ is singular, which we assume throughout to not be the case.

Equation (2.25) resembles an eigenvalue problem, where one searches for a nonzero vector v and a scalar λ such that $(A - \lambda I)v = 0$ for some square matrix A. However, it differs form a standard eigenvalue problem in two important ways. Firstly, $L(\theta)$ is a nonlinear matrix function of θ. Even when moment functions are linear, $\hat{\Omega}(\theta)$ will be quadratic in θ, and therefore so will $L(\theta)$. Secondly, θ will often be a vector of multiple parameters, while the standard definition of an eigenvalue requires it to be a scalar. Furthermore, $L(\theta)$ will be a rectangular matrix with more rows than columns when there is more than one parameter, while the standard eigenvalue problem is only defined for square matrices.

These complications give rise to generalizations of the standard eigenvalue problem that we discuss in the following sections. We focus on cases where $g_i(\theta)$ is a polynomial function of θ, which allows for the use of efficient and well established numerical methods to find all solutions of (2.25) at once. The next section specializes to the the single-parameter case.

3 The single-parameter case

When $p = 1$ (θ is a scalar), equation (2.24) simplifies to

$$\begin{bmatrix} \frac{d}{d\theta}\hat{\Omega}(\theta) & \hat{\Omega}(\theta) & \frac{d}{d\theta}\hat{g}(\theta) \\ \hat{\Omega}(\theta) & 0 & \hat{g}(\theta) \\ \frac{d}{d\theta}\hat{g}(\theta)' & \hat{g}(\theta)' & 0 \end{bmatrix} \begin{pmatrix} x(\theta) \\ \frac{d}{d\theta}x(\theta) \\ 1 \end{pmatrix} = 0, \tag{3.1}$$

which is a (single-parameter) nonlinear eigenvalue problem. The particular form of this nonlinearity is determined by the form of the moment functions in θ. In the polynomial case (including the linear case of degree 1 polynomials), equation (3.1) can be transformed into a GEP and solved with the QR algorithm. Section 3.2 specializes to the case of linear moment functions, and section 3.3 discusses the more general polynomial case.

3.1 *Mathematical background I*: Nonlinear eigenvalue problems

The methods developed throughout this work concern the transformation of systems of nonlinear equations into (generalizations of) eigenvalue problems. It is therefore important to have a baseline understanding of both the theory and the methods involved in these problems. Eigenvalues play a crucial role in linear algebra, and find wide applications in many areas of mathematics. This has motivated the development of efficient numerical methods for the computation of eigenvalues of very large matrices. Most of the problems we discuss can be solved numerically in the standard software library LAPACK (Anderson et al., 1999), which is available on a wide range of programming languages. In this section, we recall the standard eigenvalue problem and proceed by discussing generalizations of increasing complexity.

Definition 3.1 (standard eigenvalue problem)**.** An eigenvalue-eigenvector pair $(\lambda, v) \in \mathbb{C} \times \mathbb{C}^n$ of a square matrix $A \in \mathbb{C}^{n\times n}$ is a solution to

$$Av = \lambda v, \qquad v \neq 0. \tag{3.2}$$

From $Av = \lambda v \iff (A - \lambda I_n)v = 0$, we have that λ is an eigenvalue of A if and only if the matrix $A - \lambda I_n$ is singular. Therefore, the eigenvalues of A are the solutions to $\det(A - \lambda I_n) = 0$. The left hand side of this equation is a degree n polynomial in λ known as the characteristic polynomial of matrix A. It follows from the Fundamental Theorem of Algebra that A has at least one eigenvalue, and if $A \in \mathbb{R}^{n\times n}$ with n odd, at least one eigenvalue is real.

Up to $n = 4$, eigenvalues can be obtained analytically as the roots of the characteristic polynomial, while larger eigenvalue problems require iteration. The standard direct[2] method for solving the eigenvalue problem is the QR algorithm of Francis (1961). Since its introduction in 1961, a number of improvements have been made to increase the efficiency of the QR algorithm. Modern implementations usually converges after one or two iterations and require $\mathcal{O}(n^2)$ arithmetic operations. For symmetric matrices, the number of operations is considerably

[2]By direct, it is meant that under very general circumstances the method converges after a fixed number of iterations.

lower on average. For a rigorous exposition of the QR algorithm we refer to Golub and Van Loan (2013).

Note that for any square matrix A, the leading term of $\det(A - \lambda I_n)$ is $(-1)^n\lambda^n \not\equiv 0$. As a consequence, the degree of the characteristic polynomial of A is always precisely equal to n. This observation may seem trivial at first, but a similar result does not hold for the generalized eigenvalue problem (GEP):

Definition 3.2 (Generalized eigenvalue problem)**.** A generalized eigenvalue-eigenvector pair $(\lambda, v) \in \mathbb{C} \times \mathbb{C}^n$ of an ordered pair of square matrices $A, B \in \mathbb{C}^{n\times n}$ is a solution to

$$Av = \lambda Bv \qquad v \neq 0. \tag{3.3}$$

The generalized eigenvalue problem over matrices A and B is denoted by (A, B). The matrix function of λ given by $A - \lambda B$ is known as a matrix pencil. For simplicity, I omit the word "generalized" when referring to generalized eigenvalues and eigenvectors of (A, B), as it is implied by context. It should be noted that when B is invertible, we have $Av = \lambda Bv \iff B^{-1}Av = \lambda v$ and (3.3) reduces to a standard eigenvalue problem over matrix $B^{-1}A$. Dividing (3.3) on both sides by λ yields that if λ is an eigenvalue of (A, B), then $\frac{1}{\lambda}$ is an eigenvalue of (B, A), with the same associated eigenvector. Division by λ is of course not possible when $\lambda = 0$. When (B, A) has a zero eigenvalue, it is said that (A, B) has a missing eigenvalue or an eigenvalue at infinity, which is the case whenever B is singular.

From $Av - \lambda Bv \iff (A - \lambda B)v = 0$, we again have that the eigenvalues of (A, B) are the roots of the polynomial $\det(A - \lambda B)$, which generalizes the characteristic polynomial of a matrix to the pair (A, B). Its leading term is equal to $\det(B)\lambda^n$, which sheds light on the phenomenon of eigenvalues at infinity: when B is singular, the degree of $\det(A - \lambda B)$ is strictly lower than n, and therefore it has less than n complex solutions.

Note that when B is singular, the GEP can't be solved by applying the QR algorithm to $B^{-1}A$. We may hope to apply the QR algorithm to $A^{-1}B$ and obtain the reciprocal eigenvalues, but it may also be the case that A is singular. Fortunately, the QZ algorithm of Moler and Stewart (1973) for solving the GEP does not require inversion of either A or B. Even when B is invertible, the QZ algorithm is more efficient than first computing the inverse of B and then applying the QR to $B^{-1}A$.

A well known application of the GEP in econometrics is the computation of the Limited Information Maximum Likelihood (LIML) estimator of Anderson and Rubin (1949). We discuss the LIML and its relation to the CUE in the following continuation of Example 1.

Example 1 (Continued)**.** Recall the linear instrumental variables model in (2.10), which can be written in vector form as

$$\begin{aligned} y_1 &= Y_2\beta_0 + U, \\ Y_2 &= Z\Pi_0 + V. \end{aligned} \tag{3.4}$$

It is well known that the Two-Stage Least Squares (2SLS) estimator is equivalent to the GMM with the weight matrix $\hat{W} = N^{-1}Z'Z$, which is a scalar multiple of the optimal weight

matrix under conditional homoskedasticity. Indeed, if we impose $\mathbb{E}[u_i^2|z_i] = \sigma_u^2$ constant, then $\mathbb{E}\left[(y_{1i} - y_{2i}'\beta_0)^2 z_i z_i'\right] = \sigma_u^2 \mathbb{E}[z_i z_i']$ and we can estimate $\Omega(\theta)$ by

$$\widetilde{\Omega}(\beta) = N^{-1}(b'Y'Yb)Z'Z.$$

Unlike $\hat{\Omega}(\beta)$ in (2.11), this estimator depends on β only through the scalar factor $b'Y'Yb$. In two-step estimation, this factor enters the objective function as a constant and does not affect the minimizer. Similarly, the CUE also reduces to a particular estimator when conditional homoskedasticity is imposed. Using $\widetilde{\Omega}(\beta)$ instead of $\hat{\Omega}(\beta)$ in (2.9), we obtain

$$Q_{\text{LIML}}(\beta) = \frac{b'Y'Z(Z'Z)^{-1}Z'Yb}{b'Y'Yb} = \frac{b'Y'P_ZYb}{b'Y'Yb}, \tag{3.5}$$

where P_Z denotes the projection matrix of Z. The minimizer of (3.5) is the LIML estimator of Anderson and Rubin (1949). The LIML is reported to have e better small sample properties than 2SLS in the presence of weak instruments and is consistent under many instruments asymptotics (Bekker, 1994). While (3.5) can also present multiple local minima, the particular structure of its objective function allows for computing the minimizer by solving a generalized eigenvalue problem. To see this, note that (3.5) is homogeneous in b, meaning that any scalar multiple of b yields the same objective function value. We can therefore exchange the normalization coming from $b = (1, -\beta')'$ —which effectively fixes the first entry of b equal to one— by any alternative normalization and optimize with respect to all entries of $b = (b_1, \ldots, b_{p+1})'$. The value of β implied by b can be recovered from $\beta = -(b_2, \ldots, b_{p+1})/b_1$ whenever $b_1 \neq 0$. A particularly useful normalization is obtained by setting the denominator of (3.5) equal to one, which yields the problem

$$\min_b \; b'Y'P_ZYb \quad \text{subject to } \; b'Y'Yb = 1.$$

From the first order conditions, we obtain

$$Y'P_ZYb = \lambda Y'Yb, \tag{3.6}$$

where λ is the lagrange multiplier associated with the normalization constraint. Equation (3.6) defines a GEP over the pair of matrices $(Y'P_ZY,\, Y'Y)$, where b is the eigenvector and λ as the eigenvalue. We can therefore compute all critical points of Q_{LIML} as solutions to a generalized eigenvalue problem. Multiplying (3.6) on the left by b', we obtain $\lambda = b'Y'P_ZYb$, which is simply the value of the objective. Therefore, the minimizer is given by the eigenvector associated with the smallest eigenvalue of $(Y'P_ZY,\, Y'Y)$. Ratios of quadratic forms such as (3.5) are known as generalized Rayleigh quotients, and their relationship to GEPs is an important topic in linear algebra (Magnus and Neudecker, 2019, Section 11.5).

This characterization of the LIML as a generalized eigenvector greatly simplifies its computation in practice, as one can employ the QZ algorithm instead of relying on numerical optimization methods for minimizing $Q_{\text{LIML}}(\beta)$.

While our method for the CUE understands the parameter as an eigenvalue instead of an

eigenvector, we also find a possible interpretation of the CUE objective function as a sort of Rayleigh quotient. Since this interpretation does not lend itself to practical methods for obtaining estimates, we leave this discussion to Appendix A.

The equation for the GEP (A, B) can be written as $P(\lambda)v = 0$, where $P(\lambda) = A - \lambda B$ is a linear matrix function of λ. A straightforward generalization of the GEP can be obtained by replacing $P(\lambda)$ by a polynomial matrix function, which is known as the polynomial eigenvalue problem (PEP):

Definition 3.3 (Polynomial eigenvalue problem)**.** An eigenvalue-eigenvector pair $(\lambda, v) \in \mathbb{C} \times \mathbb{C}^n$ of the polynomial eigenvalue problem over d ordered square matrices $A_0, A_1, \ldots, A_d \in \mathbb{C}^{n\times n}$ is a solution to the equation

$$\sum_{i=0}^{d} \lambda^i A_i v = 0, \qquad v \neq 0. \tag{3.7}$$

The matrix function of λ given by $P(\lambda) = \sum_i^d \lambda^i A_i$ is known as a polynomial matrix pencil of degree d. Evidently, the GEP is a PEP of degree 1. The eigenvalues of the PEP are the roots of $\det(P(\lambda))$, which is a polynomial of degree nd in λ, from which we learn that the PEP will have nd eigenvalues in general. It is subject to the same concerns as the GEP regarding infinite eigenvalues and regularity. $P(\lambda)$ has an infinite eigenvalue if $\tilde{P}(\lambda) = \lambda^d P(\frac{1}{\lambda})$ has a zero eigenvalue, which is the case whenever A_d is singular. $\tilde{P}(\lambda)$ is called the *reversal* of $P(\lambda)$, and its nonzero finite eigenvalues are the reciprocals of the eigenvalues of $P(\lambda)$.

While it may seem like the general PEP is much more difficult to solve than the GEP, that is fortunately not the case. A degree d PEP over $n \times n$ matrices can be transformed into a simple GEP over a pair of $nd \times nd$ matrices. That is achieved by the procedure of linearization. Because it is a case of particular interest, we first present a linearization for the PEP with $d = 2$, also known as the quadratic eigenvalue problem.

Proposition 3.1 (Linearization of quadratic matrix polynomials)**.** *(i) If (λ, v) is an eigenvalue-eigenvector pair of the PEP*

$$(A_0 + A_1\lambda + A_2\lambda^2)v = 0, \tag{3.8}$$

then (λ, u) with $u = (v', \lambda v')'$ is an eigenvalue-eigenvector pair of the GEP

$$\left(\begin{bmatrix} 0 & K \\ A_0 & A_1 \end{bmatrix} - \begin{bmatrix} K & 0 \\ 0 & -A_2 \end{bmatrix} \lambda \right) u = 0 \tag{3.9}$$

for any full rank matrix K of same dimensions as A_i.

(ii) If (λ, u) is an eigenvalue-eigenvector pair of (3.9), then $u = \begin{pmatrix} v \\ \lambda v \end{pmatrix}$ for v such that (λ, v) is an eigenvalue-eigenvector pair of (3.8).

Proof. For part (ii), let λ and $u = (w', z')'$ satisfy (3.9). Then,

$$-\lambda K w + K z = 0,$$
$$A_0 w + A_1 z + \lambda A_2 z = 0.$$

Since K is full rank, the first equation implies $z = \lambda w$. Substitute in the second equation to get $A_0 w + \lambda A_1 w + \lambda^2 A_2 w = 0$, from which we have that λ and $v = w$ satisfies (3.8) and $u = (v', \lambda v')$. For part (i), take $w = v$, $z = \lambda v$ and verify that both equations are satisfied. □

Proposition (3.1) is adapted from Gohberg et al. (1982). Note that if the A_i are all symmetric and A_0 is full rank, we can choose $K = A_0$ and (3.9) becomes a symmetric matrix pencil, which saves arithmetic operations in the QZ algorithm.

Obtaining symmetry preserving linearizations of higher degree matrix polynomials is more involved, although possible (Fassbender et al., 2017). We do not explore such lineralizations both because of their complexity and because practical experience shows that the gains in efficiency are marginal for medium sized problems.

A simple linearization that works for matrix polynomials of any degree is the Frobenius companion form:

Proposition 3.2 (Linearization of matrix polynomials)**.** *(i) If (λ, v) is an eigenvalue-eigenvector pair of the PEP*

$$\sum_i^d \lambda^i A_i v = 0, \tag{3.10}$$

then (λ, u) with $u = (1, \lambda, \ldots, \lambda^{d-1})' \otimes v$ is an eigenvalue-eigenvector pair of the GEP

$$\left(\begin{bmatrix} A_0 & A_1 & A_2 & \cdots & A_{d-1} \\ 0 & I_n & 0 & \cdots & 0 \\ 0 & 0 & I_n & \cdots & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & 0 & \cdots & I_n \end{bmatrix} - \lambda \begin{bmatrix} 0 & 0 & \cdots & 0 & -A_d \\ I_n & 0 & \cdots & 0 & 0 \\ 0 & I_n & \cdots & 0 & 0 \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ 0 & 0 & \cdots & I_n & 0 \end{bmatrix} \right) u = 0. \tag{3.11}$$

(ii) If (λ, u) is an eigenvalue-eigenvector pair of (3.11), then $u = (1, \lambda, \ldots, \lambda^{d-1})' \otimes v$ for v such that (λ, v) is an eigenvalue-eigenvector pair of (3.10).

Proof. For part (ii), let λ and $u = (w_0', \ldots, w_{d-1}')'$ satisfy (3.11). From the second to the last block rows of the linear pencil in (3.11), we obtain equations $w_{i+1} = \lambda w_i$ for $i = 0, \ldots, d-2$, which implies $w_i = \lambda^i w_0$ for $i = 0, \ldots, d-1$. Substitute in the equation obtained from the first block row get $A_0 w_0 + \cdots + \lambda^d A_d w_0 = 0$, from which we have that λ and $v = w_0$ satisfy (3.10) and $u = (1, \lambda, \ldots, \lambda^{d-1})' \otimes v$. For part (i), take $w_i = \lambda^i v$ for $i = 0, \ldots, d-1$ and verify that all equations are satisfied. □

Proposition (3.2) is also adapted from Gohberg et al. (1982). The GEP in (3.11) can again be solved by the QZ algorithm. Before proceeding to the next section, we briefly mention what would be the next logical step in our sequence of generalizations: the rational eigenvalue problem (REP):

Definition 3.4 (Rational eigenvalue problem)**.** An eigenvalue-eigenvector pair $(\lambda, v) \in \mathbb{C} \times \mathbb{C}^n$ of the rational eigenvalue problem over d square matrices $A_0, A_1, \ldots, A_d \in \mathbb{C}^{n\times n}$ and d pairs of scalar polynomials $p_i(\lambda)$, $q_i(\lambda)$ with coefficients in $\mathbb{C}$ is a solution to the equation

$$\sum_i^d A_i \frac{p_i(\lambda)}{q_i(\lambda)} v = 0, \qquad v \neq 0. \tag{3.12}$$

Note that scalar coefficients were not present in the definition of the PEP because they are assumed to be incorporated into the coefficient matrices. One brute-force approach to solve the REP is to multiply (3.12) by the product of the denominator terms:

$$\sum_i^d A_i \left(\prod_{j\neq i}^d q_i(\lambda) \right) p_i(\lambda) v = 0, \qquad v \neq 0. \tag{3.13}$$

which yields a PEP that can be solved by Proposition (3.2). This comes at the cost of introducing extraneous solutions that are roots of the denominator terms, but this problem can be worked around by checking solution by solution. A more serious problem is that the degree of (3.13) may be too high for a linearization to be computationally feasible. A better approach was introduced in Su and Bai (2011), which shows that direct linearizations of (3.12) are possible and result in considerably smaller GEPs. We do not present such linearizations here as they require a more involved exposition.

Equipped with the results of this Section, we are ready to tackle the problem of solving for the first order conditions of the single-parameter CUE objective function in Section 3.2.

3.2 Linear moment conditions

When the moment vector is linear in the parameter, we can write $g(z, \theta) = a_0(z) + a_1(z)\theta$. Since we only impose linearity with respect to θ, the coefficient vectors a_0 and a_1 are allowed to depend nonlinearly on the data. From $\mathbb{E}[g(z_i, \theta)] = \mathbb{E}[a_0(z)] + \mathbb{E}[a_1(z)]\theta$, we can write the sample analog of the moment vector in terms of the sample analogs of the coefficients. Define

$$\hat{a}_k = N^{-1} \sum_i^N a_k(z_i) \qquad \text{for } k = 0, 1.$$

which allows us to write $\hat{g}(\theta) = \hat{a}_0 + \hat{a}_1\theta$. Similarly for the asymptotic variance, note that

$$\Omega(\theta) = \mathbb{E}\left[a_0(z_i)a_0(z_i)'\right] + \mathbb{E}\left[a_0(z_i)a_1(z_i)' + a_1(z_i)a_0(z_i)'\right]\theta + \mathbb{E}\left[a_1(z_i)a_1(z_i)'\right]\theta^2,$$

which is a quadratic matrix function of θ. Define

$$\begin{gathered} \hat{C}_0 = N^{-1} \sum_i^N a_0(z_i)a_0(z_i)'; \qquad \hat{C}_1 = N^{-1} \sum_i^N \left(a_0(z_i)a_1(z_i)' + a_1(z_i)a_0(z_i)'\right); \\ \hat{C}_2 = N^{-1} \sum_i^N a_1(z_i)a_1(z_i)'. \end{gathered} \tag{3.14}$$

and we can then write the estimator of the asymptotic variance in (2.5) as $\hat{\Omega}(\theta) = \hat{C}_0 + \hat{C}_1\theta + \hat{C}_2\theta^2$. Differentiating with respect to θ, we obtain

$$\frac{d}{d\theta}\hat{g}(\theta) = \hat{a}_1 \quad \text{and} \quad \frac{d}{d\theta}\hat{\Omega}(\theta) = \hat{C}_1 + 2\hat{C}_2\theta.$$

Substitute in (3.1) and rearrange to obtain

$$\begin{aligned} L(\theta) &= \begin{bmatrix} \hat{C}_1 + 2\hat{C}_2\theta & \hat{C}_0 + \hat{C}_1\theta + \hat{C}_2\theta^2 & \hat{a}_1 \\ \hat{C}_0 + \hat{C}_1\theta + \hat{C}_2\theta^2 & 0 & \hat{a}_0 + \hat{a}_1\theta \\ \hat{a}_1' & \hat{a}_0' + \hat{a}_1'\theta & 0 \end{bmatrix} \\ &= \begin{bmatrix} \hat{C}_1 & \hat{C}_0 & \hat{a}_1 \\ \hat{C}_0 & 0 & \hat{a}_0 \\ \hat{a}_1' & \hat{a}_0' & 0 \end{bmatrix} + \begin{bmatrix} 2\hat{C}_2 & \hat{C}_1 & 0 \\ \hat{C}_1 & 0 & \hat{a}_1 \\ 0 & \hat{a}_1' & 0 \end{bmatrix}\theta + \begin{bmatrix} 0 & \hat{C}_2 & 0 \\ \hat{C}_2 & 0 & 0 \\ 0 & 0 & 0 \end{bmatrix}\theta^2 \\ &= A_0 + A_1\theta + A_2\theta^2, \end{aligned} \tag{3.15}$$

where all three $(2m+1) \times (2m+1)$ coefficient matrices A_i are symmetric. Equation (3.15) shows that when the moment functions are linear, $L(\theta)$ is a quadratic matrix function of θ and (3.1) reduces to a quadratic eigenvalue problem. We can use Proposition 3.1 to equivalently state this problem as regular GEP of dimensions $2(2m+1) \times 2(2m+1)$, which can be solved numerically with the QZ algorithm. Since the matrices in (3.15) are symmetric, a symmetry preserving linearization can be employed. Inspection of the right-most matrix in (3.15) shows that it has a column of zeros, so there is at least one eigenvalue at infinity. The real eigenvalues are precisely the critical points of the CUE objective function. The minimizer $\hat{\theta}_{\text{CUE}}$ can then be selected among the solutions as the one that achieves the smallest function value. In Example 1, we test the performance of this method in Monte Carlo experiments.

Example 1 (Continued)**.** Recall the linear instrumental variables model of Example 1, which we now specialize to the case of a single endogenous regressor:

$$\begin{aligned} y_{1i} &= y_{2i}\beta_0 + u_i, \\ y_{2i} &= z_i'\pi_0 + v_i. \end{aligned} \tag{3.16}$$

where β_0, y_{2i} and v_i are now scalars. To implement the method presented in this section, define:

$$\hat{a}_0 = N^{-1}\sum_i^N y_{1i}\, z_i\,; \qquad \hat{C}_0 = N^{-1}\sum_i^N y_{1i}^2 z_i z_i'\,;$$

$$\hat{a}_1 = -N^{-1}\sum_i^N y_{2i}\, z_i\,; \qquad \hat{C}_1 = -2N^{-1}\sum_i^N y_{1i}y_{2i}\, z_i z_i'\,;$$

$$\hat{C}_2 = N^{-1}\sum_i^N y_{2i}^2\, z_i z_i'\,.$$

Comparing with (2.11) and (2.12), we have

$$Z'Y = \begin{bmatrix} \hat{a}_0 & -\hat{a}_1 \end{bmatrix} \quad \text{and} \quad \Sigma = \begin{bmatrix} \hat{C}_2 & -\hat{C}_1/2 \\ -\hat{C}_1/2 & \hat{C}_2 \end{bmatrix}.$$

For Monte Carlo simulations, we let the structural form errors be distributed according to $v_i \sim \mathcal{N}(0,1)$ and $u_i = (v_i + \eta_i)/\sqrt{2}$, where $\eta_i \sim \mathcal{N}(0, 3z_{1i}^2)$ and $z_{1i} \sim \mathcal{U}(-1,1)$ is the first entry in the $m \times 1$ vector of instruments z_i. This generates heteroskedasticity and makes it so that the unconditional variance of u_i is equal to one. The full vector of instruments is given by $z_i = (z_{1i}, p_2(z_{1i}), \ldots, p_m(z_{1i}))'$ where $p_i(z)$ denotes the i-th normalized Legendre polynomial. The Legendre polynomials are defined recursively by $\widetilde{p}_0(z) = 1$, $\widetilde{p}_1(z) = z$ and

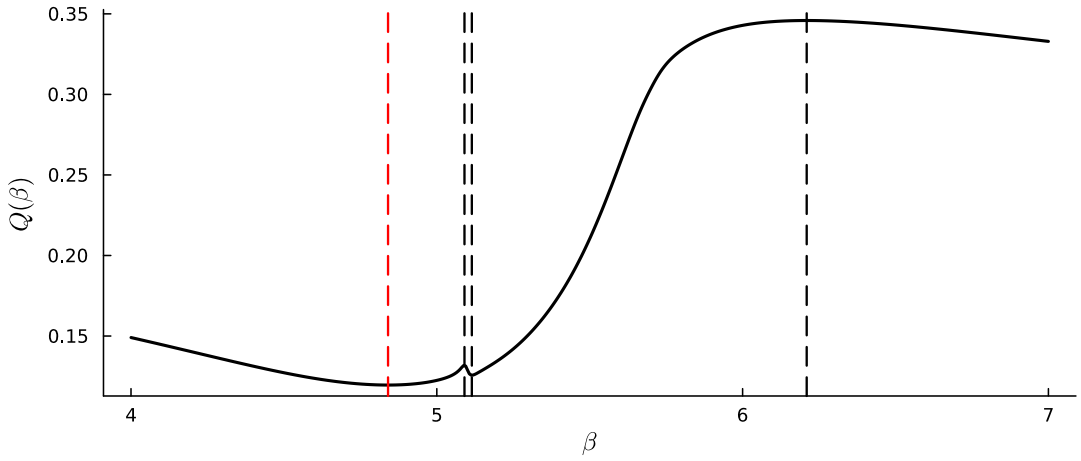
$$(k+1)\widetilde{p}_{k+1}(z) = (2k+1)\, z\, \widetilde{p}_k(z) - k\, \widetilde{p}_{k-1}(z) \;\text{ for }\; k \geq 2.$$

We then normalize to $p_k(z) \coloneqq \frac{2k+1}{2}\widetilde{p}_k(z)$. While this setup may seem convoluted at first, it is able to generate a sequence of any number of instruments that are functions of the same random variable z_{1i}, but all have zero mean, unit variance and, most importantly, are uncorrelated. This yields $\mathbb{E}[z_i z_i'] = I_m$. The absence of correlation comes from the fact that the Legendre polynomials are a family of orthogonal functions with respect to the L^2 norm on the interval $(-1,1)$. We set the sample size to $N = 500$ and the true parameter to $\theta_0 = 5$. The first stage coefficient is set to $\pi_0 = (\widetilde{\pi}, 0, \ldots, 0)'$ so that the true model only depends on z_{1i}.

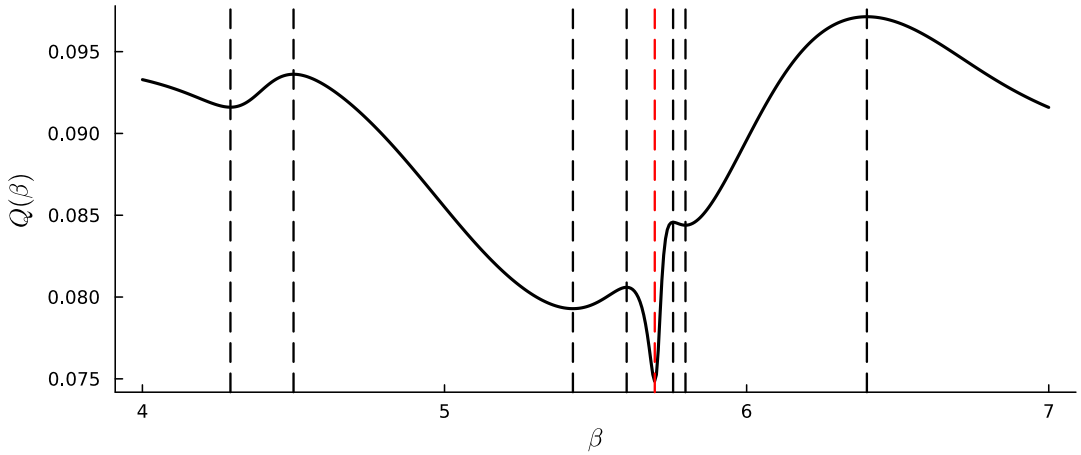
Figure 2 depicts typical objective functions and their critical points for different values of $\widetilde{\pi}$ and the number of instruments m. We can see that either increasing the number of instruments or decreasing the first stage coefficient makes the objective function present a more erratic behavior. The multiplicity of local minima indicate that local minimization routines are prone to failure. The objective function in (2b) shows that grid search methods can also perform poorly, as the true minimum is located between two local maxima that are close together. The objective function in (2c) is the most erratic, presenting multiple small bumps along its graph. Figure 3 depicts some of its the most ill-behaved sections up close. It shows that the eigenvalue approach is able to correctly identify all critical points, even the ones that are hard to spot by careful visual inspection of the graph.

To explore the performance of the CUE compared to alternative estimators, we set $\widetilde{\pi} = 1$ and vary the number of instruments from 1 to 159 at increments of 2. We perform 10000 simulations per increment. Figure 4 depicts the median bias for the 2SLS, two-step GMM, LIML and CUE estimators. The shaded bands range from the 5% to the 95% percentiles for each estimator. While the CUE presents higher dispersion than any of the other estimators, its median bias is virtually zero for all numbers of moments considered. Other estimators present clear biases that grow with the number of moments.

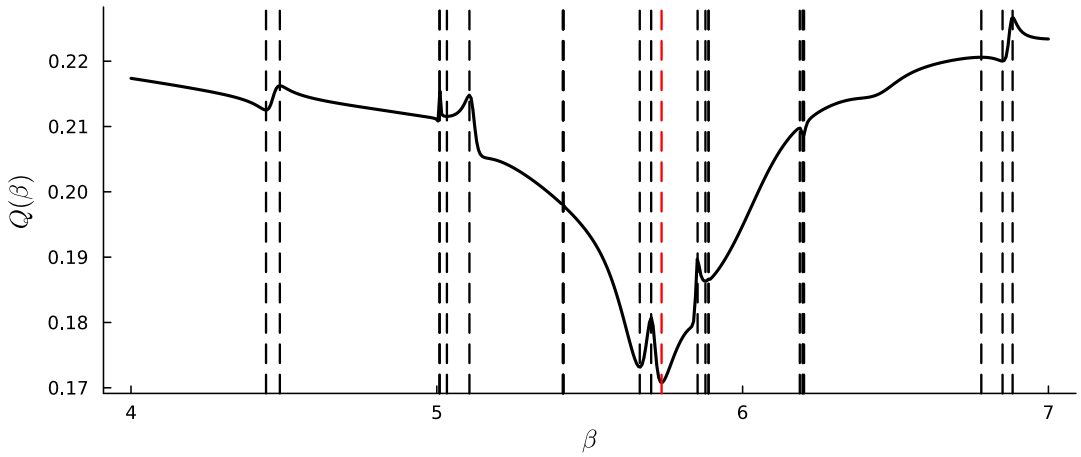
Table 1 presents benchmarking results for different numbers of moment conditions. We find that runtimes are very low even for large m, which shows that the method is viable for virtually all practical applications. The method is implemented in the *Julia* programming language (Bezanson et al., 2017) and generalized eigenvalues are computed using the `eigvals` function of the *LinearAlgebra.jl* base package. We run the algorithm 1000 times for each row in a 3GHz Intel i7-9700 CPU.



(a) $m = 50$, $\widetilde{\pi} = 1$



(b) $m = 50$, $\widetilde{\pi} = 1/10$



(c) $m = 100$, $\widetilde{\pi} = 1/10$

Figure 2: CUE objective functions and critical points for Example 1.

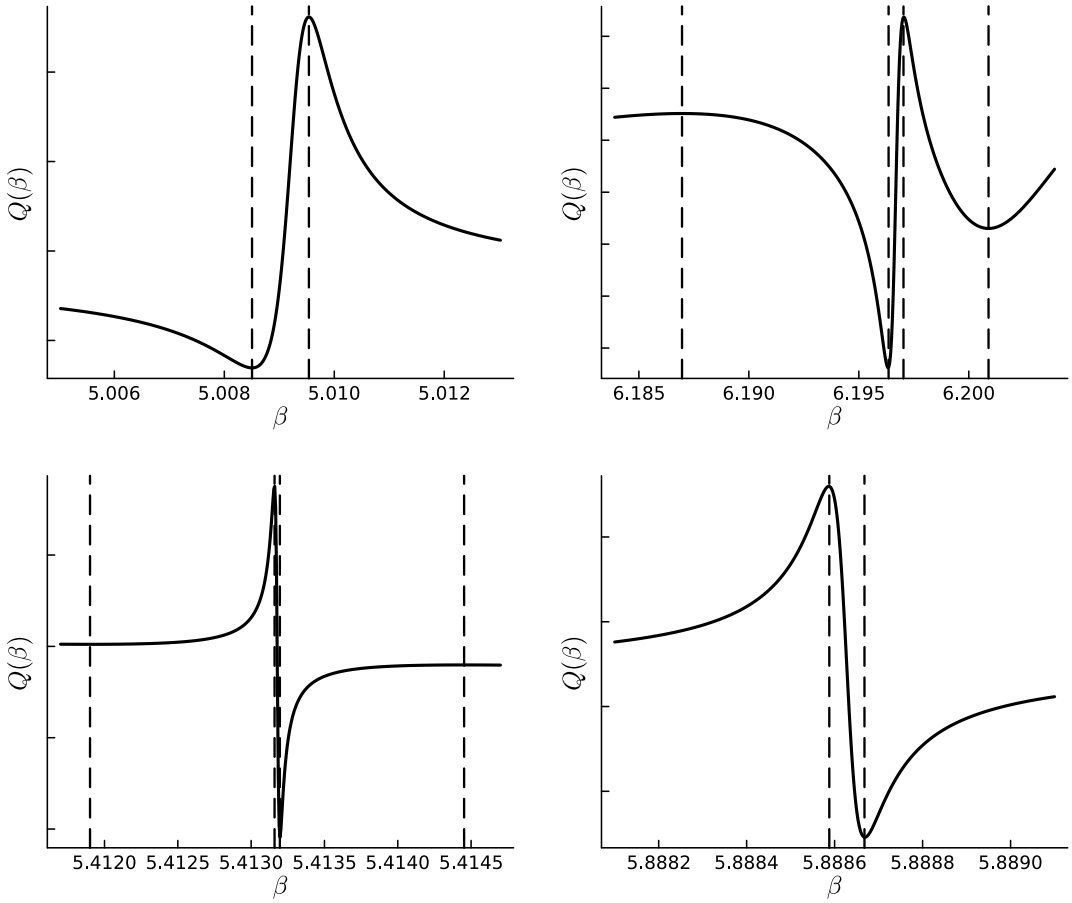


Figure 3: Close-up plots of the objective function in Figure 2c.

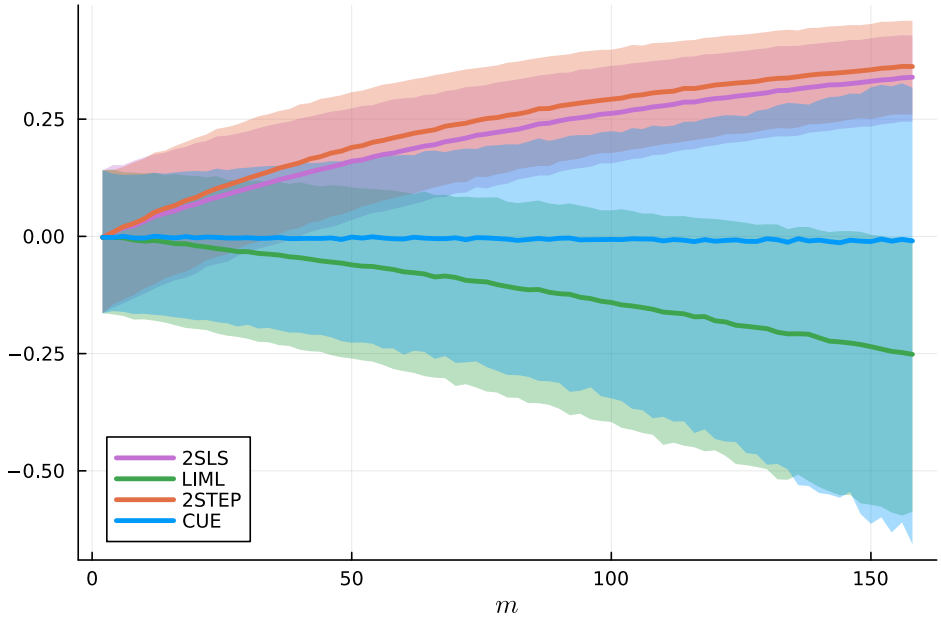


Figure 4: Median bias and nine decile ranges for Example 1.

m	mean time (s)	median time (s)
5	0.0006	0.0001
10	0.0010	0.0009
50	0.0482	0.0480
100	0.2543	0.2539
300	3.1109	3.1057
500	11.5571	11.5423
700	27.8682	27.8537

Table 1: Mean and median runtimes in seconds.

In the linear IV model, we have learned that when $\Sigma = Y'Y \otimes Z'Z$, the CUE and the LIML are identical. When that is the case, each eigenvalue of $L(\beta)$ will correspond to an eigenvector of $(Y'P_ZY, Y'Y)$, which is a GEP over $(p+1) \times (p+1)$ matrices and will therefore have at most $p+1$ solutions.

To see how matrix $L(\beta)$ simplifies when $\Sigma = Y'Y \otimes ZZ'$, consider the case of $p = 1$. With $\hat{\Omega}(\beta) = N^{-1}(b' \otimes I_k)(Y'Y \otimes ZZ')(b \otimes I_k) = N^{-1}(b'Y'Yb)Z'Z$, we have

$$L(\beta) = N^{-1} \begin{bmatrix} -2(y_2'Yb)Z'Z & (b'Y'Yb)Z'Z & -Z'y_2 \\ (b'Y'Yb)Z'Z & 0 & Z'Yb \\ -(Z'y_2)' & (Z'Yb)' & 0 \end{bmatrix}.$$

Note that the first $2k \times 2k$ block of $L(\beta)$ can be written as $N^{-1}M \otimes Z'Z$, where

$$M = \begin{bmatrix} -2(y_2'Yb) & (b'Y'Yb) \\ (b'Y'Yb) & 0 \end{bmatrix}.$$

Using Schur's determinant formula (Magnus and Neudecker, 2019), we have

$$\det(L(\beta)) = -\det(M \otimes Z'Z) \left(N^{-1} \begin{pmatrix} -Z'y_2 \\ Z'Yb \end{pmatrix}' (M \otimes Z'Z)^{-1} \begin{pmatrix} -Z'y_2 \\ Z'Yb \end{pmatrix} \right) \tag{3.17}$$

Since $\det(M \otimes Z'Z) = \det(M) \cdot \det(Z'Z)$ and $\det(M) = -(b'Y'Yb)^2 < 0$, this determinant can be disregarded when considering the solutions to $L(\beta) = 0$. It is easy to compute the 2×2 inverse matrix $M^{-1} = \frac{1}{b'Y'Yb} \begin{bmatrix} 0 & 1 \\ 1 & 2\frac{y_2'Yb}{b'Y'Yb} \end{bmatrix}$, from which we obtain

$$\begin{aligned} N^{-1} \begin{pmatrix} -Z'y_2 \\ Z'Yb \end{pmatrix}' (M \otimes Z'Z)^{-1} \begin{pmatrix} -Z'y_2 \\ Z'Yb \end{pmatrix} &= 2N^{-1} \left(\frac{(y_2'Yb)(b'Y'P_ZYb)}{(b'Y'Yb)^2} - \frac{y_2'P_ZYb}{b'Y'Yb} \right) \\ &= \frac{\partial}{\partial \beta} \left(N^{-1} \frac{b'Y'P_ZYb}{b'Y'Yb} \right) \end{aligned}$$

which is precisely the derivative of the LIML objective function as defined in (3.5).

3.3 Polynomial moment conditions

It is straightforward to generalize the procedure of Section 3.2 to the polynomial case. Suppose $g(z,\theta)$ a degree d vector of polynomials in θ:

$$g(z,\theta) = \sum_{k=0}^{d} a_k(z)\theta^k.$$

Define

$$\hat{a}_k = N^{-1}\sum_{i}^{N} a_k(z_i) \quad \text{for } k = 0,\ldots,d$$

and

$$\hat{C}_k = N^{-1}\sum_{\substack{j,\,\ell\le d\\ j+\ell=k}}\sum_{i}^{N} a_j(z_i)a_\ell(z_i)' \qquad \text{for } k = 0,\ldots,2d.$$

Note that matrices $\hat{C}_k$ particularize to the matrices in (3.14) when $d = 1$.

We can then write the sample moments as

$$\hat{g}(\theta) = \sum_{k=0}^{d} \hat{a}_k\theta^k$$

and the asymptotic variance estimator as

$$\hat{\Omega}(\theta) = \sum_{k=0}^{2d} \hat{C}_k\theta^k.$$

Differentiating with respect to θ, we obtain

$$\frac{d}{d\theta}\hat{g}(\theta) = \sum_{k=0}^{d-1}(k+1)\,\hat{a}_{k+1}\theta^k \qquad \text{and} \qquad \frac{d}{d\theta}\hat{\Omega}(\theta) = \sum_{k=0}^{2d-1}(k+1)\,\hat{C}_{k+1}\theta^k.$$

Substitute in (3.1) and rearrange to obtain

$$L(\theta) = \sum_{k=0}^{d-1}\left(\begin{bmatrix}(k+1)\hat{C}_{k+1} & \hat{C}_k & (k+1)\hat{a}'_{k+1}\\ \hat{C}_k & 0 & \hat{a}_k\\ (k+1)\hat{a}'_{k+1} & \hat{a}_k & 0\end{bmatrix}\theta^k\right) + \begin{bmatrix}(d+1)\hat{C}_{d+1} & \hat{C}_d & 0\\ \hat{C}_d & 0 & \hat{a}_d\\ 0 & \hat{a}_d & 0\end{bmatrix}\theta^d$$
$$+ \sum_{k=d+1}^{2d-1}\left(\begin{bmatrix}(k+1)\hat{C}_{k+1} & \hat{C}_k & 0\\ \hat{C}_k & 0 & 0\\ 0 & 0 & 0\end{bmatrix}\theta^k\right) + \begin{bmatrix}0 & \hat{C}_{2d} & 0\\ \hat{C}_{2d} & 0 & 0\\ 0 & 0 & 0\end{bmatrix}\theta^{2d}. \quad (3.18)$$

We can then write

$$L(\theta) = \sum_{k=0}^{2d} A_k\theta^k, \quad (3.19)$$

where A_k are $(2m+1) \times (2m+1)$ matrices defined accordingly. We obtain that if moment conditions are of degree d, the first order condition of the CUE objective function is equivalent to a PEP of degree $2d$. We can then apply Proposition 3.2 to transform this problem into a regular GEP that can be solved by the QZ algorithm. The resulting GEP has dimensions $d(2m+1) \times d(2m+1)$. Inspection of the right-most matrix in (3.18) shows that it has a column of zeros, so $\det(L(\theta))$ has degree at most $d(2m+1) - 1$.

Example 3 (Example 3 in Imbens and Spady (2002))**.** Let $z_i \overset{iid}{\sim} \mathcal{N}(\theta_0, 1)$. For GMM estimation of θ_0, consider the moment function based on the first five cumulants:

$$g(z, \theta) = \begin{pmatrix} z - \theta \\ z^2 - \theta^2 - 1 \\ z^3 - \theta^3 - 3\theta \\ z^4 - \theta^4 - 6\theta^2 - 3 \\ z^5 - \theta^5 - 10\theta^3 - 15\theta \end{pmatrix}.$$

We have $m = d = 5$. $L(\theta)$ is a degree 10 matrix polynomial over 11×11 matrices. Using Proposition 3.2, we solve for the first order conditions by computing generalized eigenvalues of 110×110 matrices.

In a numerical exercise, we set the true parameter to $\theta_0 = 5$ and the sample size to $N = 10$. Figure 5 depicts the CUE objective function for a particular realization of the data. The critical points found by computing generalized eigenvalues are indicated by dashed vertical lines. Computation of generalized eigenvalues is virtually instantaneous for a problem of this size, taking less than 5 milliseconds on average.

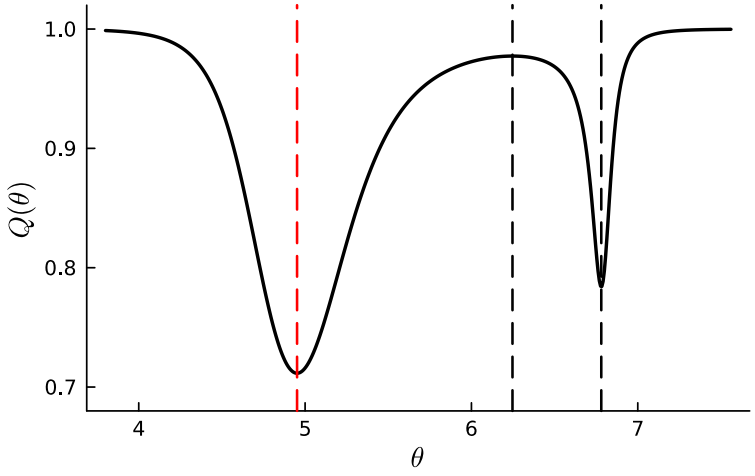


Figure 5: CUE objective function for Example 3.

3.4 Confidence sets based on the CUE objective function

Good econometric practice dictates that confidence sets should always be reported alongside point estimates. In order to refrain from specifying the probability distribution of the data, one usually appeals to asymptotic theory is order to construct confidence sets that are asymptotically

valid. Under appropriate assumptions regarding the moment conditions, the distribution of the data and the parameter space (Newey and McFadden, 1994, p. 2145), it can be established that any of the asymptotically equivalent GMM estimators discussed in Section 2 are asymptotically normal. That is, $\sqrt{N}(\hat{\theta}-\theta_0) \xrightarrow{d} N(0,V)$. This motivates the approximate $100(1-\alpha)\%$ confidence set for the i-th parameter estimate given by

$$\hat{\theta}_i \mp z_{\alpha/2}\sqrt{N^{-1}\hat{V}_{ii}},$$

where $z_{\alpha/2}$ is the $\alpha/2$ quantile of the standard normal distribution. Such confidence sets are appealing due to being symmetric intervals around the parameter estimate, as well as being very easy to compute. Indeed, one only needs an estimate for V, which comes free with the estimate of $\Omega(\theta_0)$ that is already required for the optimal weight matrix. It is not surprising that these are the most widely used confidence sets in applications. However, they are not without disadvantages. For one, the finite sample distribution of $\hat{\theta}$ is poorly approximated by a normal distribution when the model parameters are weakly identified by the moment conditions, as we have previously discussed. A more immediate motivation for considering such confidence sets comes from the invariance of the CUE objective function to transformations of the moment function of the form $h(z,\theta) = H(\theta)g(z,\theta)$, where $H(\theta)$ is a $m \times m$ nonsingular matrix function of the parameters. Hall (2005) describes h as a *curvature altering transformation* of $g(z,\theta)$. Unlike the CUE, the two-step and Iterated GMM estimators are not invariant to such transformations. This can be observed by inspecting their first order conditions. When the moment function $g(z,\theta)$ is used, the first order condition with respect to θ_i is

$$2\left(\frac{\partial}{\partial\theta_i}\hat{g}(\theta)\right)'\hat{W}\hat{g}(\theta) = 0.$$

If the moment function is replaced by $H(\theta)g(z,\theta)$, it becomes

$$2\left(\left(\frac{\partial}{\partial\theta_i}H(\theta)\right)\hat{g}(\theta) + H(\theta)\left(\frac{\partial}{\partial\theta_i}\hat{g}(\theta)\right)\right)'\hat{W}H(\theta)\hat{g}(\theta) = 0,$$

and there is no reason for the solutions to these equations to be the same in finite samples. Meanwhile, the invariance of the CUE can be easily verified by inspecting its objective function:

$$\begin{aligned} Q(\theta) &= \hat{g}(\theta)'H(\theta)'\left(N^{-1}\sum_i^N H(\theta)g_i(\theta)g_i(\theta)'H(\theta)'\right)^{-1}H(\theta)\hat{g}(\theta) \\ &= \hat{g}(\theta)'H(\theta)'\left(H(\theta)\hat{\Omega}(\theta)H(\theta)'\right)^{-1}H(\theta)\hat{g}(\theta) \\ &= \hat{g}(\theta)'\hat{\Omega}(\theta)^{-1}\hat{g}(\theta). \end{aligned}$$

In certain applications, one may argue that a particular choice of moment function is "natural" in the sense that it has a clear economic interpretation, and any transformation of the parameters

would obfuscate that interpretation. However, that is not always the case, as some specifications may require arbitrary choices in the parametrization of the moment conditions. This problem is always present in cases where some of the model parameters are identified only up to scale. Suppose, for example, that economic theory implies the moment condition $\mathbb{E}[\theta_1 - \theta_2 z_i] = 0$. identification of the ratio θ_1/θ_2 could be achieved by either setting one of the parameters to a constant or by considering a more complicated restriction such as $\theta_2 = \sqrt{1-\theta_1^2}$. While the CUE is invariant to any such transformation, the two-step and Iterated estimators are not. Whenever there is no reason to prefer one normalization over the other, invariance to normalization is clearly desirable.

To obtain asymptotically valid confidence intervals based on the CUE objective function, it is necessary to establish its asymptotic distribution when it is evaluated at the true parameter. Under standard assumptions, it can be established that $\hat{\Omega}(\theta_0) \xrightarrow{p} \Omega(\theta_0)$ and $N^{\frac{1}{2}}\hat{g}(\theta_0) \xrightarrow{d} \mathcal{N}(0, \Omega(\theta_0))$, which implies $NQ(\theta_0) \xrightarrow{d} \chi^2_m$. An asymptotically valid confidence set for $\hat{\theta}$ is then given by

$$C(\theta) = \{\theta \,|\, NQ(\theta) < q_{m,\alpha}\}, \tag{3.20}$$

where $q_{m,\alpha}$ is the $1-\alpha$ quantile of the χ^2_m distribution. It is important to note that confidence intervals of this form may be infinite or empty. Since the objective function of the CUE evaluated at the parameter estimate can also be a test statistic for over-identifying restrictions, an empty confidence set is evidence that the model is misspecified. A related problem is that a very small confidence set can either be an indication that the parameters are well identified by the moment conditions or that the model is misspecified, with no way to distinguish between the two just by inspecting the objective function. With these caveats, we can move on to the practical considerations on the computation of (3.20).

The method of Moreira et al. (2023) for computing CUE estimates generalizes easily to the problem of computing confidence sets. Our proposed method, however, requires a little more work. Instead of considering the first order conditions $\nabla Q(\theta) = 0$ as we did in Section 2.1, we now turn our attention to equation $Q(\theta) - q = 0$.

In the single-parameter case, continuity of Q implies that any confidence set can be written as a disjoint union of K open intervals:

$$C(\theta) = \{\theta \,|\, Q(\theta) < q\} = \bigcup_{k=1}^{K} \left(\underline{\theta_k}, \overline{\theta_k}\right), \tag{3.21}$$

where we allow for $\underline{\theta_1} = -\infty$ or $\overline{\theta_K} = +\infty$. We compute all boundary points of the intervals in (3.21) simultaneously by transforming the equation

$$\hat{g}(\theta)'\hat{\Omega}(\theta)^{-1}\hat{g}(\theta) = q \tag{3.22}$$

into a generalized eigenvalue problem. Using the same auxiliary vector $x(\theta) = \hat{\Omega}(\theta)^{-1}\hat{g}(\theta)$ defined in Section (2.1), we have obtain the system

$$\hat{g}(\theta)'x(\theta) - q = 0,$$
$$\hat{\Omega}(\theta)x(\theta) - \hat{g}(\theta) = 0.$$

Linearity in $x(\theta)$ allows us to write this system in matrix form:

$$\begin{bmatrix} \hat{\Omega}(\theta) & \hat{g}(\theta) \\ \hat{g}(\theta)' & q \end{bmatrix} \begin{pmatrix} x(\theta) \\ -1 \end{pmatrix} = 0. \tag{3.23}$$

Denote the matrix in the left hand side of (3.23) by $M(\theta; q)$. We obtain that the boundary points of confidence sets based on the inversion of the CUE objective function are the real eigenvalues of the nonlinear eigenvalue problem:

$$M(\theta; q)v = 0, \qquad v \neq 0.$$

When moment conditions are polynomial in θ we have that $M(\theta; q)$ is a matrix polynomial of degree at least two, for which the methods of Section (3.1) can be applied. For the sake of illustration, linear moment conditions yield

$$M(\theta; q) = \begin{bmatrix} \hat{C}_0 & \hat{a}_0 \\ \hat{a}_0' & q \end{bmatrix} + \begin{bmatrix} \hat{C}_1 & \hat{a}_1 \\ \hat{a}_1' & 0 \end{bmatrix} \theta + \begin{bmatrix} \hat{C}_2 & 0 \\ 0 & 0 \end{bmatrix} \theta^2,$$

where matrices $\hat{a}_i$ and $\hat{C}_i$ are defined as in Section 3.2. Figure 6 depicts 90% confidence intervals for the simulation setup in Example 1.

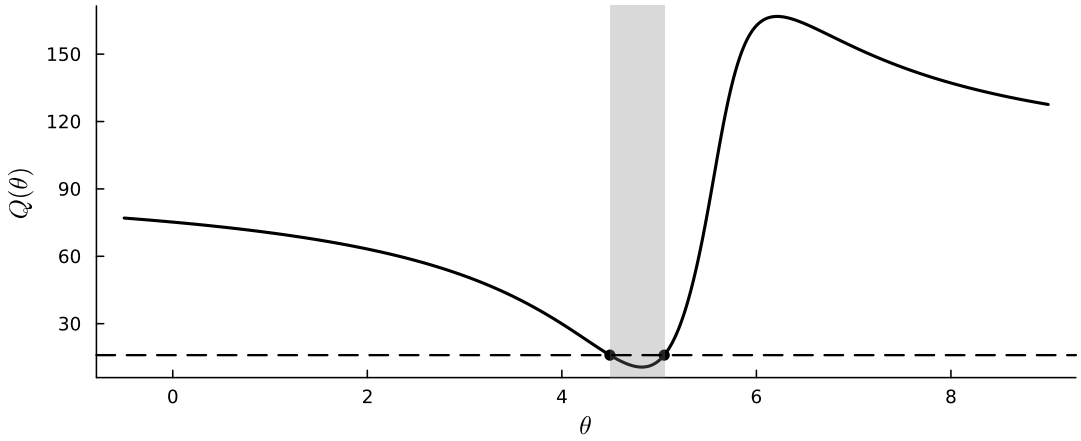
We end this section by observing that while $M(\theta)$ is $(m+1) \times (m+1)$, we can find an equivalent $m \times m$ matrix that has all the same eigenvalues. Another application of Schur's determinant formula yields

$$\det(M(\theta; q)) = \det(q) \det\Big(\hat{\Omega}(\theta) - \hat{g}(\theta)\hat{g}(\theta)'/q\Big) = \det\Big(q\,\hat{\Omega}(\theta) - \hat{g}(\theta)\hat{g}(\theta)'\Big).$$

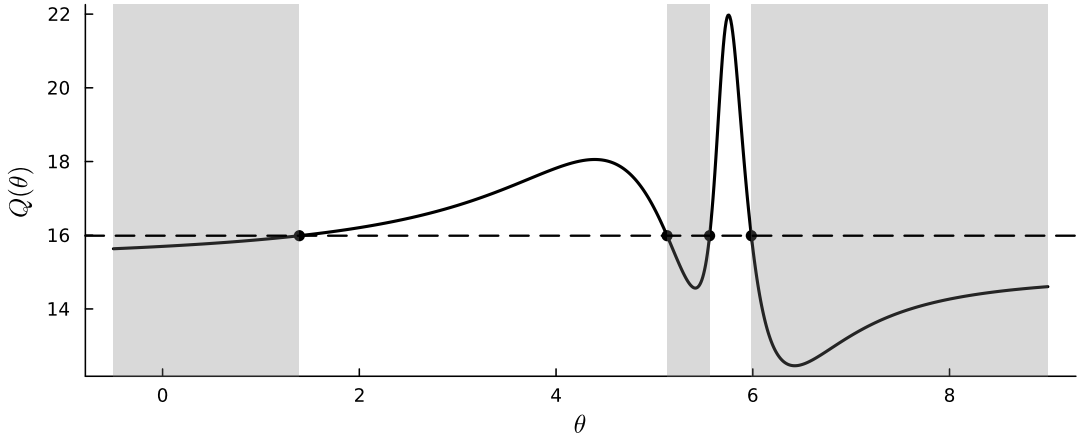
We can therefore apply the same procedure to $\overline{M}(\theta; q) = q\,\hat{\Omega}(\theta) - \hat{g}(\theta)\hat{g}(\theta)'$. In the case of linear moment conditions, for example, we obtain

$$\overline{M}(\theta; q) = q\,\hat{C}_0 - \hat{a}_0\hat{a}_0' + (q\,\hat{C}_1 + \hat{a}_1\hat{a}_1')\theta + q\,\hat{C}_2\theta^2.$$

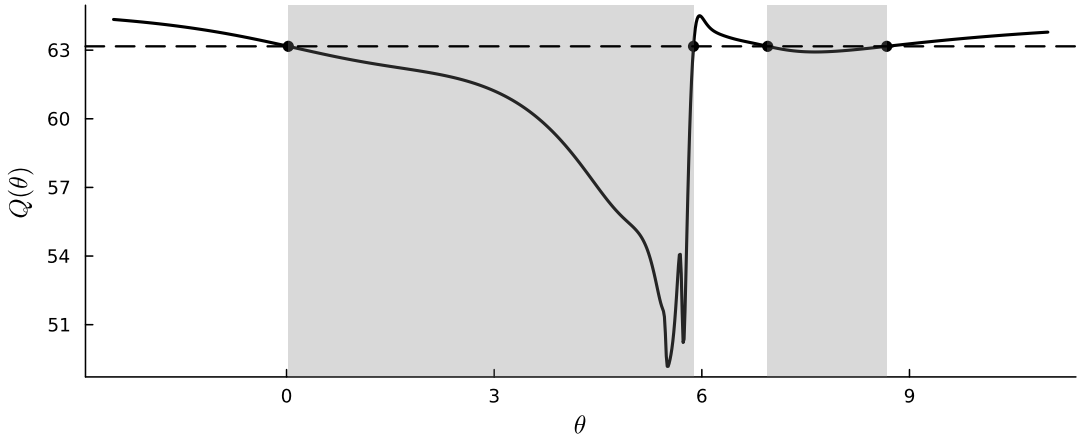
The relatively small dimensions of $\overline{M}(\theta; q)$ also motivate a different method for the minimization of $Q(\theta)$ from the one presented in Section (3). When m is either too large for a linearization of $L(\theta)$ to fit the memory, or if the resulting GEP is ill-conditioned, we can alternatively compute the CUE by an iterative algorithm that only requires solving for the eigenvalues of $\overline{M}(\theta; q)$. Note that if q is such that $Q(\hat{\theta}_{\text{CUE}}) > q$, equation (3.22) will have no solutions, and therefore $\overline{M}(\theta; q)$ will have no real eigenvalues. If, on the other hand, $Q(\hat{\theta}_{\text{CUE}}) < q$ and $\max_\theta Q(\theta) > q$, then $\overline{M}(\theta; q)$ will have at least one real eigenvalue. This result allows us to compute the CUE by performing a bissection algorithm in the image of $Q(\theta)$ that will eventually converge to $Q(\hat{\theta}_{\text{CUE}})$. A detailed description of this procedure is presented in the appendix under Algorithm 1.



(a) $m = 10$, $\widetilde{\pi} = 1$



(b) $m = 10$, $\widetilde{\pi} = 1/10$



(c) $m = 50$, $\widetilde{\pi} = 1/10$

Figure 6: 90% confidence intervals for Example 1.

4 The multiparameter case

As discussed in Section 2.1, our approach to formulating eigenvalue problems from first order conditions sees parameters as eigenvalues. It is then of no surprise that solving models with multiple parameters require solving eigenvalue problems with multiple eigenvalues.

We recall the main result of Section (2), which states that finding the critical points of the CUE objective function amounts to finding all pairs (θ, v) that solve

$$\underbrace{\begin{bmatrix} \mathsf{D}\,\hat{\Omega}(\theta) & I_p \otimes \hat{\Omega}(\theta) & \operatorname{vec}(\mathsf{D}\,\hat{g}(\theta)) \\ \hat{\Omega}(\theta) & 0 & \hat{g}(\theta) \\ \mathsf{D}\,\hat{g}(\theta)' & I_p \otimes \hat{g}(\theta)' & 0 \end{bmatrix}}_{=L(\theta)} v = 0, \qquad v \neq 0 \tag{4.1}$$

When $p > 1$, equation (4.1) differs from any eigenvalue problem studied in Section 3 in two ways: first, $\theta = (\theta_1, \ldots, \theta_p)'$ is now a vector, which means we have multiple eigenvalues. Second, $L(\theta)$ is an $(n + p - 1) \times n$ rectangular matrix, where $n = m(p + 1) + 1$.

Section 4.1 discusses the multiparameter eigenvalue problem, and Section 4.2 details our method for computing the CUE in the case of multiple parameters.

4.1 *Mathematical background II*: Multiparameter eigenvalue problems

In Section (3.1) we have studied two generalizations of the standard eigenvalue problem. Each is a proper generalization of the previous one in the sense that the standard eigenvalue problem is a GEP where one of the matrices in the pair is the identity matrix, and the GEP is a PEP of degree 1. This section explores a further generalization: the multiparameter eigenvalue problem (MEP). The MEP is also a proper generalization of the previous two, but its complexity warrants starting the exposition with a particular case: the two-parameter eigenvalue problem (2EP):

Definition 4.1 (Two parameter eigenvalue problem)**.** The vector $(\lambda, \mu)' \in \mathbb{C}^2$ is an eigenvalue and the Kronecker product $x \otimes y \in \mathbb{C}^n \otimes \mathbb{C}^n$ is an eigenvector of the two-parameter eigenvalue problem over matrices $A_i, B_i, C_i \in \mathbb{C}^{n \times n}$, $i = 1, 2$ if they jointly solve

$$C_1 x = \lambda A_1 x + \mu B_1 x \tag{4.2}$$

$$C_2 y = \lambda A_2 y + \mu B_2 y \qquad x, y \neq 0. \tag{4.3}$$

The 2EP is associated with two bivariate matrix pencils $P_i(\lambda, \mu) = A_i + \mu B_i - C_i$. Its eigenvalues are solutions to the system of two bivariate polynomial equations $\det(P_i(\lambda, \mu)) = 0$, each of degree n. Bézout's theorem (Cox et al., 2013) implies that the 2EP will have n^2 solutions in general.

Again, the 2EP (and the MEP in general) is subject to the same concerns about solutions at infinity and regularity as the GEP. A rigorous treatment of the spectral theory for MEPs can be found in the book by Atkinson (1972).

To establish a foundation for understanding the 2EP, it helps to draw an analogy to a different problem with which we are familiar: a linear system of two equations in two unknowns. Consider the problem of finding $\lambda, \mu \in \mathbb{R}$ that solve

$$\begin{aligned} c_1 &= a_1\lambda + b_1\mu, \\ c_2 &= a_2\lambda + b_2\mu. \end{aligned} \tag{4.4}$$

The standard way to proceed is to eliminate one of the variables (say, μ) from one of the equations (say, the first). We accomplish this by multiplying the first equation by b_1 and the second equation by b_2:

$$\begin{aligned} c_1b_2 &= a_1b_2\lambda + b_1b_2\mu, \\ b_1c_2 &= b_1a_2\lambda + b_1b_2\mu. \end{aligned}$$

The second equation yields $b_1c_2 - b_1a_2\lambda = b_1b_2\mu$, which we can substitute in the first to obtain λ as a solution to a single equation:

$$c_1b_2 = a_1b_2\lambda + b_1c_2 - b_1a_2\lambda \iff \lambda = \frac{c_1b_2 - b_1c_2}{a_1b_2 - b_1a_2}. \tag{4.5}$$

The solution to the 2EP follows a similar approach. The goal is to eliminate the term that contains μ from (4.2). Using the Kronecker product operator, we multiply (4.2) on the right by B_2y and (4.3) on the left by B_1x[3].

$$\begin{aligned} C_1x \otimes B_2y &= \lambda A_1x \otimes B_2y + \mu B_1x \otimes B_2y, \\ B_1x \otimes C_2y &= \lambda B_1x \otimes A_2y + \mu B_1x \otimes B_2y. \end{aligned}$$

The second equation yields $\mu B_1x \otimes B_2y = B_1x \otimes C_2y - \lambda B_1x \otimes A_2y$, which we can substitute in the first:

$$\begin{aligned} C_1x \otimes B_2y &= \lambda A_1x \otimes B_2y + B_1x \otimes C_2y - \lambda B_1x \otimes A_2y, \\ C_1x \otimes B_2y - B_1x \otimes C_2y &= \lambda \left(A_1x \otimes B_2y - B_1x \otimes A_2y\right). \end{aligned}$$

Using the mixed-product property of the Kronecker product[4], we obtain

$$(C_1 \otimes B_2 - B_1 \otimes C_2)(x \otimes y) = \lambda \left(A_1 \otimes B_2 - B_1 \otimes A_2\right)(x \otimes y).$$

Similarly, eliminating λ from equation (4.3) yields

[3] We adopt the convention that the Kronecker product takes precedence over matrix multiplication in the order of operations: $AB \otimes C = (AB) \otimes C$.

[4] $AB \otimes CD = (A \otimes C)(B \otimes D)$.

$$(A_1 \otimes C_2 - C_1 \otimes A_2)(x \otimes y) = \mu(A_1 \otimes B_2 - B_1 \otimes A_2)(x \otimes y).$$

Which are both GEPs over $n^2 \times n^2$ matrices. We have thus proved part (i) of Proposition 4.1, which is adapted from Atkinson (1972).

Proposition 4.1. *(i) If (λ, μ) is an eigenvalue and $x \otimes y$ is an eigenvector of the 2EP*

$$\begin{aligned} C_1 x &= \lambda A_1 x + \mu B_1 x, \\ C_2 y &= \lambda A_2 y + \mu B_2 y, \end{aligned} \tag{4.6}$$

then (λ, γ) with $\gamma = x \otimes y$ is an eigenvalue-eigenvector pair of the GEP

$$\Delta_1 \gamma = \lambda \Delta_0 \gamma, \tag{4.7}$$

and (μ, γ) is an eigenvalue-eigenvector pair of the GEP

$$\Delta_2 \gamma = \mu \Delta_0 \gamma, \tag{4.8}$$

where

$$\begin{aligned} \Delta_0 &= A_1 \otimes C_2 - C_1 \otimes A_2, \\ \Delta_1 &= C_1 \otimes B_2 - B_1 \otimes C_2, \\ \Delta_2 &= A_1 \otimes B_2 - B_1 \otimes A_2. \end{aligned} \tag{4.9}$$

(ii) If (λ, γ) is an eigenvalue-eigenvector pair of the GEP (4.7), there is μ such that (μ, γ) is an eigenvalue-eigenvector pair of the GEP (4.8) and $\gamma = (x \otimes y)$ such that (μ, λ) is an eigenvalue and $(x \otimes y)$ is an eigenvector of the MEP (4.6).

The proof of part (ii) is more involved, at it entails proving not only that the GEPs (4.7) and (4.8) share the same eigenvectors, but also that their eigenvectors have a Kronecker structure. Note that matrices Δ_i *do not* have a Kronecker structure, as the sum of Kronecker products is not itself a Kronecker product. We refer the reader to Atkinson (1972) for a complete proof.

The GEPs (4.7) and (4.8) once again can be numerically solved by the QZ algorithm. Note that it is not necessary to solve both GEPs in order to obtain the complete set of eigenvalues. Once we solve the first one for the $j = 1, \ldots, n^2$ eigenvalues λ_j and eigenvectors γ_j, the μ_j can be obtained from

$$\mu_j = \frac{\gamma_j^H \Delta_0 \gamma_j}{\gamma_j^H \Delta_2 \gamma_j}$$

and the x_j and y_j can be computed up to scale from the Kronecker product decomposition of γ_j (Van Loan and Pitsianis, 1993).

It is straightforward to generalize the 2EP to the general k-parameter linear MEP:

Definition 4.2 (Linear multiparameter eigenvalue problem)**.** The vector $(\lambda_1, \ldots, \lambda_k)' \in \mathbb{C}^k$ is an eigenvalue and the Kronecker product $x_1 \otimes \cdots \otimes x_k \in \mathbb{C}^n \otimes \cdots \otimes \mathbb{C}^n$ is an eigenvector of the multiparameter eigenvalue problem over matrices $A_{ij} \in \mathbb{C}^{n\times n}$, $i = 1, \ldots, k$, $j = 0, \ldots, k$ if they jointly solve

$$A_{i0}x_i = \sum_{j=1}^{k} \lambda_j A_{ij} x_i, \quad x_i \neq 0, \qquad i = 1, \ldots, k. \tag{4.10}$$

The standard procedure for solving the linear MEP is also a straightforward generalization of the procedure for the 2EP. We return to the analogy with the linear system in (4.4) for a moment. Note that the solution for λ we obtained in (4.5) can be expressed as

$$\lambda = \frac{\det(M_1)}{\det(M)}, \qquad \text{where} \quad M = \begin{bmatrix} a_1 & b_1 \\ a_2 & b_2 \end{bmatrix}, \quad M_1 = \begin{bmatrix} c_1 & b_1 \\ c_2 & b_2 \end{bmatrix}.$$

Which is the well known Cramer's rule for the solution of a system of linear equations. The procedure for obtaining matrices (4.9) generalizes to a matrix valued version of the determinant:

Definition 4.3 (Operator determinant)**.** The operator determinant of a $k \times k$ array of $n \times n$ matrices A_{ij}, i,$j = 1, \ldots, k$ is a $n^k \times n^k$ matrix Δ defined as

$$\Delta = \begin{vmatrix} A_{11} & \cdots & A_{1k} \\ \vdots & \ddots & \vdots \\ A_{k1} & \cdots & A_{kk} \end{vmatrix}_{\otimes} = \sum_{\sigma \in S_k} \operatorname{sgn}(\sigma) A_{\sigma(1)1} \otimes \cdots \otimes A_{\sigma(k)k}. \tag{4.11}$$

where S_k is the group of permutations of indices $i = 1, \ldots, k$ and $\operatorname{sgn}(\sigma) = 1$ ($\operatorname{sgn}(\sigma) = -1$) whenever the permutation $\sigma \in S_k$ is obtainable by an even (odd) number of element swaps.

Note that the formula in (4.11) is simply the well known Leibniz formula for determinants where scalars are replaced by matrices and scalar products are replaced by Kronecker products. Proposition 4.2 generalizes Proposition (4.1) much like Cramer's rule generalizes the result in (4.5).

Proposition 4.2. *(i) If $(\lambda_1, \ldots \lambda_k)$ is an eigenvalue and $x_1 \otimes \cdots \otimes x_k$ is an eigenvector of the MEP*

$$A_{i0}x_i = \sum_{j=1}^{k} \lambda_j A_{ij} x_i, \quad x_i \neq 0, \qquad i = 1, \ldots, k, \tag{4.12}$$

then (λ_j, γ) with $\gamma = x_1 \otimes \cdots \otimes x_k$ is an eigenvalue-eigenvector pair of the GEP

$$\Delta_j \gamma = \lambda_j \Delta_0 \gamma \tag{4.13}$$

for $j = 1, \ldots, k$, where

$$\Delta_0 = \begin{vmatrix} A_{11} & \cdots & A_{1k} \\ \vdots & \ddots & \vdots \\ A_{k1} & \cdots & A_{kk} \end{vmatrix}_{\otimes} \quad \textit{and}$$
$$\Delta_j = \begin{vmatrix} A_{11} & \cdots & A_{1(j-1)} & A_{10} & A_{1(j+1)} & \cdots & A_{1k} \\ \vdots & \ddots & \vdots & \vdots & \vdots & \ddots & \vdots \\ A_{k1} & \cdots & A_{k(j-1)} & A_{k0} & A_{k(j+1)} & \cdots & A_{kk} \end{vmatrix}_{\otimes}. \tag{4.14}$$

(ii) If (λ_j, γ) is an eigenvalue-eigenvector pair of the j-th GEP in (4.13), there are λ_i such that (λ_i, γ) is an eigenvalue-eigenvector pair for the remaining $i \neq j$ GEPs in (4.13) and $\gamma = x_1 \otimes \cdots \otimes x_k$ such that $(\lambda_1, \ldots \lambda_k)$ is an eigenvalue and $x_1 \otimes \cdots \otimes x_k$ is an eigenvector of the MEP (4.12).

To gain further insight on what Proposition 4.2 does, we again recall the linear system (4.4). In that problem, we are faced with a pair of equations that act simultaneously on two different spaces: the space where λ lives and the space where μ lives, which in that case are two identical copies of $\mathbb{R}$. Linear algebra tackles the problem by posing it as a single equation acting on a single space: the vector space $\mathbb{R}^2$ where $(\lambda, \mu)'$ lives. The theory of MEPs follows the same approach in taking the problem to a single space: the Kronecker product (or tensor product) space.

At the beginning of this section, we discussed how the MEP is a proper generalization of the PEP, so the next logical step is to discuss the polynomial MEP. However, we will not explore this problem in its full generality. The definition of the general polynomial MEP involves cumbersome notation, as the number of terms can be very large in general. However, it is often the case in applications that only a small number of terms is present. The strategy for solving the polynomial MEP is to find an appropriate linearization and then solve the resulting linear MEP using Proposition (4.2). Efficient linearizations should be particular to the terms that are actually present in the problem at hand, so a general linearization is hardly useful. We illustrate this point in Example 4. The last definition in this section will therefore concern only a particular case: the quadratic two-parameter eigenvalue problem (Q2EP).

Definition 4.4 (Quadratic two-parameter eigenvalue problem)**.** The vector $(\lambda, \mu)' \in \mathbb{C}^2$ is an eigenvalue and the Kronecker product $x_1 \otimes x_2 \in \mathbb{C}^n \otimes \mathbb{C}^n$ is an eigenvector of the quadratic two-parameter eigenvalue problem over matrices $A_i, B_i, C_i, D_i, E_i, F_i \in \mathbb{C}^{n \times n}$, $i = 1, 2$, if they jointly solve

$$A_i x_i + \lambda B_i x_i + \mu C_i x_i + \lambda^2 D_i x_i + \lambda\mu E_i x_i + \mu^2 F_i x_i = 0, \quad x_i \neq 0, \qquad i = 1, 2. \tag{4.15}$$

One possible linearization of the Q2EP is as follows:

$$\left(\begin{bmatrix} A_i & B_i & C_i \\ 0 & -I & 0 \\ 0 & 0 & -I \end{bmatrix} + \begin{bmatrix} 0 & D_i & E_i \\ I & 0 & 0 \\ 0 & 0 & 0 \end{bmatrix} \lambda + \begin{bmatrix} 0 & 0 & F_i \\ 0 & 0 & 0 \\ I & 0 & 0 \end{bmatrix} \mu \right) y_i = 0. \tag{4.16}$$

While the linearization (4.16) is easy to construct and follows a simple logic that is easy to extend to different cases, it is not as straightforward to solve as a generic linear MEP. Note that it is composed of $3n \times 3n$ matrices, and a generic MEP of this size has $6n^2$ solutions, more than the $4n^2$ solutions of a Q2EP. However, the linearized MEP is not generic, as all extra solutions are at infinity. This is caused by the block row of zeros in the second matrix and the block column of zeros in the last. In fact, on applying proposition (4.1) to (4.16), one obtains pencils (Δ_1, Δ_0) and (Δ_2, Δ_0) for which

$$\det(\Delta_1 - \lambda\Delta_0) \equiv \det(\Delta_2 - \lambda\Delta_0) \equiv 0.$$

These types of GEPs are known as *singular* GEPs, and can't be solved by the QZ algorithm. For the remainder of this section, we discuss the singular GEP and the alternative algorithms that they require.

Suppose that $A - \lambda B$ is singular for all $\lambda \in \mathbb{C}$, which is the case when A and B share a common null space. In this case, $\det(A - \lambda B)$ is identically zero and, by definition 3.2, every $\lambda \in \mathbb{C}$ is an eigenvalue of (A, B). Consider, for example,

$$\begin{bmatrix} 2 & 0 \\ 0 & 0 \end{bmatrix} v = \lambda \begin{bmatrix} 1 & 0 \\ 0 & 0 \end{bmatrix} v. \tag{4.17}$$

It is easy to see that $v = (0, 1)'$ is an eigenvector associated to any eigenvalue $\lambda \in \mathbb{C}$. However, the eigenvalue $\lambda = 2$ still holds particular significance as it is the only eigenvalue associated with the eigenvector $v = (1, 0)'$. While the rank of $\begin{bmatrix} 2-\lambda & 0 \\ 0 & 0 \end{bmatrix}$ equals one for most choices of λ, it equals zero when $\lambda = 2$. This motivates definition 4.5:

Definition 4.5 (Normal rank)**.** The normal rank of (A, B) is defined as

$$\operatorname{nrank}(A, B) = \max_{\lambda \in \mathbb{C}} \operatorname{rank}(A - \lambda B).$$

The generalized eigenvalue problem (A, B) is said to be *singular* when $\operatorname{nrank}(A, B) < n$ and it is said to be *regular* when $\operatorname{nrank}(A, B) = n$. With this, we can define a variation of an eigenvalue problem for the singular case which is not trivially solved by the whole of $\mathbb{C}$.

Definition 4.6 (Singular generalized eigenvalue problem)**.** A *regular* eigenvalue $\lambda \in \mathbb{C}$ of a singular generalized eigenvalue problem $(A, B) \in \mathbb{C}^{n\times n} \times \mathbb{C}^{n\times n}$ satisfies

$$\operatorname{rank}(A - \lambda B) < \operatorname{nrank}(A, B). \tag{4.18}$$

Its associated eigenvector $v \in \mathbb{C}^n$, $v \neq 0$ satisfies $Av = \lambda Bv$ and $Av \neq \tilde{\lambda} Bv$ for some $\tilde{\lambda} \neq \lambda$.

It is interesting to note that definition 4.6 can be easily extended to include GEPs over non-square matrices. By adding an appropriate number of columns or rows of zeroes to a pair of rectangular matrices, one can define a singular GEP over the resulting square matrices.

The singular GEP is ill-conditioned in the sense that small perturbations in its entries can radically change its eigenvalues. Consider the perturbation of (4.17) given by

$$\begin{bmatrix} 2 & \varepsilon_1 \\ \varepsilon_2 & 0 \end{bmatrix} v = \lambda \begin{bmatrix} 1 & \xi_1 \\ \xi_2 & 0 \end{bmatrix} v. \tag{4.19}$$

Problem (4.19) is no longer singular and has eigenvalues $\lambda_i = \varepsilon_i/\xi_i$, which can assume any value with ε_i and ξ_i arbitrarily small. Moreover, the eigenvalues no longer depend on the diagonal entries 2 and 1, which determine the eigenvalues of the singular problem (4.17).

Unfortunately, the ill-conditioning of the singular GEP defeats the QZ algorithm, which often returns completely spurious eigenvalues as it converges to some small regular perturbation of the original pencil. There are no routines in LAPACK suited for singular GEPs, so we must look elsewhere. The most well established algorithm for obtaining solutions for the singular GEP is the *staircase* algorithm of Demmel and Kågström (1993). The staircase algorithm reduces a singular pencil to its generalized Schur-staircase form, which is block upper triangular. If $\operatorname{nrank}(A, B) = m$, the center block of the staircase form is an $m \times m$ regular pencil that has the same regular eigenvalues as the original pencil. The QZ algorithm can then be used to compute the regular eigenvalues. A detailed exposition of this algorithm can be found in Bai et al.. It is implemented in the Fortran library GUPTRI (Demmel and Kågström, 1999). While elegant, the staircase algorithm is much more computationally demanding than the QZ algorithm.

Recently, a faster algorithm for solving the singular GEP was introduced in Hochstenbach et al. (2019), which is based on rank-completing perturbations. That is, perturbations such as (4.19) that complete the rank of the singular pencil while retaining its regular eigenvalues. Consider, for example, the perturbation of (4.17) given by

$$\begin{bmatrix} 2 & 0 \\ 0 & \varepsilon \end{bmatrix} v = \lambda \begin{bmatrix} 1 & 0 \\ 0 & \xi \end{bmatrix} v. \tag{4.20}$$

Again, (4.20) is a regular pencil. Its eigenvalues are ε/ξ and 2, and we see that this perturbation preserves the regular eigenvalue of the original pencil.

4.2 Models with multiple parameters

Equipped with the methods presented in Section (4.1), we are ready to discuss the case where $\theta = (\theta_1, \ldots, \theta_p)'$ is a vector of parameters. Recall the main result of Section (2), which states that finding the critical points of the CUE objective function amounts to solving the following nonlinear multiparameter eigenvalue problem:

$$\underbrace{\begin{bmatrix} \mathsf{D}\,\hat{\Omega}(\theta) & I_p \otimes \hat{\Omega}(\theta) & \operatorname{vec}(\mathsf{D}\,\hat{g}(\theta)) \\ \hat{\Omega}(\theta) & 0 & \hat{g}(\theta) \\ \mathsf{D}\,\hat{g}(\theta)' & I_p \otimes \hat{g}(\theta)' & 0 \end{bmatrix}}_{=L(\theta)} v = 0, \qquad v \neq 0. \tag{4.21}$$

Matrix $L(\theta)$ as an $(n + p - 1) \times n$ matrix where $n = m(p + 1) + 1$. In the single-parameter case ($p = 1$), we had that $L(\theta)$ was square and we could proceed as usual. In the multivariate case, every additional parameter adds to the difference between the number of rows and columns

of $L(\theta)$. The MEPs explored in Section 4.1 involve a number of square pencils that is equal to the number of parameters, so the case of a single rectangular pencil does not quite fit. We will follow the strategy of Plestenjak et al. (2023) and obtain p different square pencils by multiplying $L(\theta)$ by (possibly random) rectangular matrices. In the next subsection, we particularize to the two parameter linear case, but remark that the techniques we discuss do generalize to the polynomial case with any number of parameters. Similarly to the single-parameter case, we find that linearity of the moment condition implies that (4.21) is quadratic and, as Section (3.3) shows, the generalization from the quadratic to the polynomial case is no mystery. The difficult step is to find an appropriate linearization, which (as remarked in Section 4.1) crucially depends on the particular terms that are present in the problem.

Similar to what we did in the univariate case, we can write any linear bivariate moment condition as $g(z,\theta) = a_0(z) + a_1(z)\theta_1 + a_2(z)\theta_2$, where the coefficient vectors a_i do not depend on the parameters. The sample moments are $\hat{g}(\theta) = \hat{a}_0 + \hat{a}_1\theta_1 + \hat{a}_2\theta_2$ and we write the covariance matrix estimator as

$$\hat{\Omega}(\theta) = \hat{C}_{00} + \hat{C}_{10}\theta_1 + \hat{C}_{01}\theta_2 + \hat{C}_{20}\theta_1^2 + \hat{C}_{11}\theta_1\theta_2 + \hat{C}_{02}\theta_2^2,$$

where we define $\hat{C}_{ij}$ as the sum of every term in the expansion of $\sum_i^N g_i(\theta)g_i(\theta)'$ that multiplies the monomial $\theta_1^i\theta_2^j$. The use of multi-indices in this form is especially helpful when dealing with many parameters.

Substitute in (4.21) to obtain

$$\begin{aligned} L(\theta) &= \begin{bmatrix} \hat{C}_{10} & \hat{C}_{00} & 0 & \hat{a}_1 \\ \hat{C}_{01} & 0 & \hat{C}_{00} & \hat{a}_2 \\ \hat{C}_{00} & 0 & 0 & \hat{a}_0 \\ \hat{a}_1' & \hat{a}_0' & 0 & 0 \\ \hat{a}_2' & 0 & \hat{a}_0' & 0 \end{bmatrix} + \cdots + \begin{bmatrix} 0 & \hat{C}_{02} & 0 & 0 \\ 0 & 0 & \hat{C}_{02} & 0 \\ \hat{C}_{02} & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \end{bmatrix} \theta_2^2, \\ &= A_{00} + A_{10}\theta_1 + A_{01}\theta_2 + A_{20}\theta_1^2 + A_{11}\theta_1\theta_2 + A_{02}\theta_2^2, \end{aligned} \tag{4.22}$$

which is a rectangular quadratic two-parameter eigenvalue problem over $(n+1)\times n$ matrices. For the general case of p parameters and degree d, Plestenjak et al. (2023) show that a generic rectangular problem has $d^p\binom{n+p-1}{p}$ eigenvalues. This result tells us that (4.22) can have at most $2n(n+1)$ eigenvalues, but since the matrices associated with the higher order terms have a column of zeroes, we know that some will be at infinity. One way to proceed it to multiply $L(\theta)$ on the left by two different full rank matrices P_1 and P_2 of dimensions $n\times(n+1)$ and consider the Q2EP

$$P_iA_{00}x_i + \theta_1P_iA_{10}x_i + \theta_2P_iA_{01}x_i + \theta_1^2P_iA_{20}x_i + \theta_1\theta_2P_iA_{11}x_i + \theta_2^2P_iA_{02}x_i, \quad x_i \neq 0, \quad i = 1,2. \tag{4.23}$$

For generic P_i of full rank, Plestenjak et al. (2023) show that all solutions of the original

problem are preserved, and suggest using simple row selection matrices $P_1 = \begin{bmatrix} I_n & 0 \end{bmatrix}$ and $P_2 = \begin{bmatrix} 0 & I_n \end{bmatrix}$. However, we have seen in Section (3.1) that the Q2EP has $4n^2$ solutions in general, so this procedure introduces $2n(n-1)$ extraneous solutions. The extra solutions are caused by replacing the single $n \times 1$ vector of the original problem $L(\theta)\gamma = 0$ by two different vectors x_i of the same size, each associated to one of the equations $P_i L(\theta) x_i = 0$ which ignore the additional constraint $x_1 = x_2$. This is not a problem since the true solutions can be easily recovered by discarding any solution with $x_1 \neq x_2$ (or, depending on the normalization, $x_1 \neq \alpha x_2$ for some $\alpha \neq 0$). This does introduce inefficiency as some solutions are computed only to be later discarded.

Example 1 (Continued)**.** Consider a setting similar to Example 1, but with two endogenous variables. The parameter of interest is now a vector $\beta = (\beta_1, \beta_2)$. We do not elaborate on the specifics of the simulation design as it only involves minor changes to accommodate two parameters. We set $N = 50$, $K = 10$ and $\beta_0 = (5, 3)$. Figure 7 depicts typical objective functions and their critical points. We again see that the function presents multiple critical points, some of which are local minima.

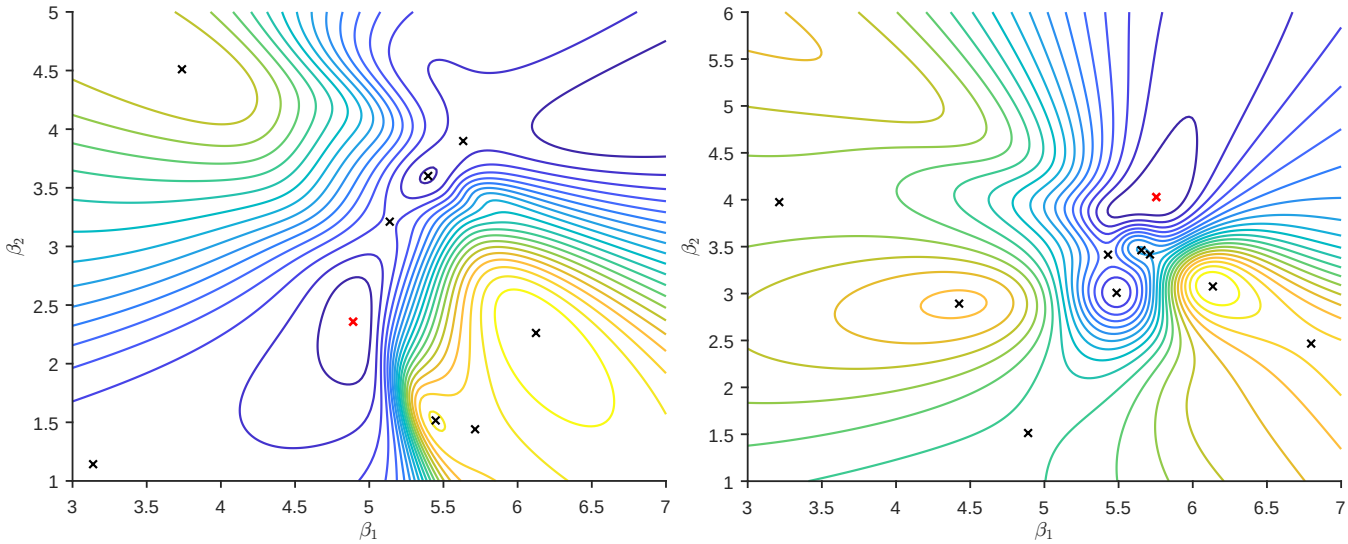


Figure 7: Contour plots for Example 1 with two parameters.

We proceed with an example of a three parameter rational model that allows us to illustrate how the multiparameter case can be generalized. The exposition of the model is based on Hall (2005).

Example 4 (Inventory Holdings (Eichenbaum, 1989))**.** Consider a representative firm that faces production costs given by

$$c_t^q = \varepsilon_t q_t + \frac{\alpha_0}{2} q_t^2,$$

where q_t denotes production at time t and v_t is a stochastic shock to the marginal cost of production. While the presence of shocks creates an incentive for the firm to hold inventory and smooth production costs, it also faces inventory holdings costs given by

$$c_t^y = \frac{\delta_0}{2}\left(y_t - \gamma_0 s_t\right)^2 + \frac{\eta_0}{2} y_t^2,$$

where y_t denotes inventory holdings and s_t denotes sales. The firm maximizes a discounted sum of expected profits:

$$\mathbb{E}_i\left[\sum_j^{\infty} \beta_0^j \left(p_{i+j} s_{i+j} - c_{i+j}^y - c_{i+j}^q\right)\right],$$

where p_t is the price of sale at time t. The expectation is conditional on the information set at time t. When ε_t is i.i.d., the Euler equation implies

$$\mathbb{E}_t\left[h_t(\theta_0)\right] = 0,$$

where

$$h_t(\theta_0) = y_{t+1} - \left(\lambda_0 + (\lambda_0\beta_0)^{-1}\right) y_i + \beta_0^{-1} y_{t-1} + s_{t+1} - \phi_0 \beta_0^{-1} s_t. \tag{4.24}$$

The parameters of interest are $\theta_0 = (\lambda_0, \phi_0, \beta_0)$. We define $\phi_0 = 1 - \delta_0\gamma_0/\alpha_0$ since the parameters δ_0, γ_0 and α_0 are not individually identified. by the law of iterated expectations, we can use any relevant variable in the information set at time t as an instrument to obtain moment conditions $\mathbb{E}[z_t h_t(\theta)] = 0$. Eichenbaum (1989) uses lagged values of inventory holding and sales.

From (4.24), we see that the moment conditions are rational functions of the parameters, which we have not encountered until now. As discussed in Section 3.1, we could work with the rational moment conditions and consider linearizations of rational pencils. However, as we have shown in Section 3.4, the CUE is invariant to multiplication of the moment conditions by functions of the parameters. To eliminate all denominators, we multiply (4.24) by $\lambda\,\beta$ to obtain [5]

$$\tilde{h}_t(\theta) = \lambda\beta y_{t+1} - \left(\lambda^2\beta + 1\right) y_i + \lambda y_{t-1} + \lambda\beta s_{t+1} - \lambda\phi s_t. \tag{4.25}$$

Which is a multivariate polynomial of degree 3. Since the variance estimator is always quadratic on the moment conditions, matrix $L(\theta)$ becomes a matrix polynomial of degree 6:

$$\begin{aligned} L(\theta) = {} & A_{000} + A_{100}\lambda + A_{110}\lambda\phi + A_{101}\lambda\beta + A_{200}\lambda^2 + A_{210}\lambda^2\phi + A_{201}\lambda^2\beta + A_{220}\lambda^2\phi^2 \\ & + A_{211}\lambda^2\beta\phi + A_{202}\lambda^2\beta^2 + A_{301}\lambda^3\beta + A_{311}\lambda^3\beta\phi + A_{302}\lambda^3\beta^2 + A_{402}\lambda^4\beta^2. \end{aligned} \tag{4.26}$$

Again, the subindices are chosen such that A_{ijk} is the coefficient matrix in the term $\lambda^i\phi^j\beta^k$.

While a dense degree 6 polynomial in 3 variables has $\binom{6+3}{3} = 84$ terms, we find that (4.26) only has 14 terms. We can leverage this sparsity to find an appropriate linearization which is not overly large. In this case, we can find:

[5] Note that we could also reparametrize $\gamma = \beta^{-1}$ and multiply the moment conditions by λ.

$$\left(\begin{bmatrix} A_{000} & A_{100} & A_{200} & A_{210} & A_{201} & A_{301} & A_{302} \\ \cdot & -I & \cdot & \cdot & \cdot & \cdot & \cdot \\ \cdot & \cdot & -I & \cdot & \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot & -I & \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot & \cdot & -I & \cdot & \cdot \\ \cdot & \cdot & \cdot & \cdot & \cdot & -I & \cdot \\ \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & -I \end{bmatrix} + \begin{bmatrix} \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & A_{402} \\ I & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot \\ \cdot & I & \cdot & \cdot & \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot & \cdot & I & \cdot & \cdot \\ \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot \end{bmatrix} \lambda + \right.$$
$$\left. + \begin{bmatrix} \cdot & A_{110} & \cdot & A_{220} & A_{211} & A_{311} & \cdot \\ \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot \\ \cdot & \cdot & I & \cdot & \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot \end{bmatrix} \phi + \begin{bmatrix} \cdot & A_{101} & \cdot & \cdot & \cdot & \cdot & A_{202} \\ \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot \\ \cdot & \cdot & I & \cdot & \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot & \cdot & \cdot & I & \cdot \end{bmatrix} \beta \right) \begin{pmatrix} x \\ \lambda x \\ \lambda^2 x \\ \lambda^2 \phi x \\ \lambda^2 \beta x \\ \lambda^3 \beta x \\ \lambda^3 \beta^2 x \end{pmatrix} = 0.$$

Where blocks of zeros are denoted by dots $(\cdot)$ for readability. In the the three parameter case, matrices A_{ijk} are of size $(4m+3) \times (4m+1)$. The linearized pencil is then of size $(7(4m+1)+2) \times 7(4m+1)$.

Example 2 (Continued)**.** Recall the sample moments we obtain from considering second moments in the dynamic panel data model of Example 2:

$$\hat{g}(\theta) = \operatorname{vech}(\hat{S}_N) - \sigma^2 \operatorname{vech}(B(\lambda_N 1_T 1_T' + I_T)B').$$

Note that $\hat{g}(\theta)$ is a size $T(T+1)/2$ vector of polynomial functions of the parameters. While B is the inverse of a linear matrix function of ρ, we have

$$B = \begin{bmatrix} 1 & 0 & \cdots & 0 \\ \rho & 1 & \cdots & 0 \\ \vdots & \ddots & \ddots & \vdots \\ \rho^{T-1} & \cdots & \rho & 1 \end{bmatrix},$$

which is a degree $T-1$ matrix polynomial. As a result, $\hat{g}(\theta)$ is of degree $(T-1)^2+1$, which is potentially very large.

Recall that we wish to estimate three parameters $\theta_0 = (\rho_0, \lambda_{N,0}, \sigma_0^2)$ using the sample moments

$$\hat{g}(\theta) = \operatorname{vech}(\hat{S}_N) - \operatorname{vech}(B(\lambda_N 1_T 1_T' + I_T \sigma^2)B'), \qquad B = (I - \rho J)^{-1}.$$

In this case, the estimator $\hat{\Omega}(\theta)$ as defined in (2.5) will depend on the full vector of incidental parameters $(\eta_i, \ldots, \eta_N)$, as can be observed from

$$g_i(\theta) = \operatorname{vech}((y_i - \overline{y})(y_i - \overline{y})') - \operatorname{vech}(B(\eta_i 1_T 1_T' + I_T \sigma^2)B').$$

One alternative is to consider the analytical expression for ($(2\sigma^2)^{-1}$ times) the asymptotic

variance matrix under the assumption that errors are normal:

$$\overline{\Omega}(\theta) = D_T^+(B \otimes B)\left(2\lambda_N(1_T 1_T' \otimes I_T) + \sigma^2 I_{T^2}\right)(B' \otimes B')D_T^{+\prime}, \tag{4.27}$$

where D_T denotes the $T^2 \times T(T+1)/2$ duplication matrix that satisfies $D_T \operatorname{vech}(A) = \operatorname{vec}(A)$ whenever A is symmetric and $D_T^+ = (D_T' D_T)^{-1} D_T'$ denotes its generalized inverse (Magnus and Neudecker, 2019). In fact, given i.i.d errors, the normality assumption can be weakened to $\mathbb{E}[u_{i,t}^4] = 3\sigma^4$ and (4.27) still holds.

A further complication that arises is the potentially high polynomial degree of both $\overline{\Omega}(\theta)$ and $\hat{g}(\theta)$ coming from B, which is a degree $T-1$ matrix polynomial in ρ. However, the can write $Q(\theta)$ only in terms of B^{-1} (which is linear in ρ) by taking $C \coloneqq D_T'(B^{-1} \otimes B^{-1})D_T$ and observing that

$$Q(\theta) = \hat{g}(\theta)'\overline{\Omega}(\theta)^{-1}\hat{g}(\theta) = \hat{g}(\theta)'C(C\overline{\Omega}(\theta)C)^{-1}C\hat{g}(\theta),$$

where

$$C\hat{g}(\theta) = C \operatorname{vech}(\hat{S}_N) - D_T' \operatorname{vec}(\lambda_N 1_T 1_T' + I_T\sigma^2)$$

and

$$C\overline{\Omega}(\theta)C = D_T'((2\lambda_N 1_T 1_T' + \sigma^2 I_T) \otimes I_T)D_T,$$

which yields a much more tractable problem, and produces a matrix pencil $L(\theta)$ of degree 5.

5 Final remarks and directions for future research

In this work, we have presented a novel method for optimization of the CUE objective function, which builds upon the recent method by Moreira et al. (2023). In the univariate case, our method is faster and requires less numerical precision. In the multivariate case, it proves to handle small to medium sized problems well, but going from one to two parameters greatly complicates the procedure. We do note that numerical methods currently available for multiparameter eigenvalue problems are still in their infancy. While the spectral theory of multiparameter eigenvalue problems date back to Atkinson (1968), they have only become of interest in applied research in the last few years. We may expect that increasingly efficient methods being developed today may soon make our approach feasible for larger models.

While our method works for linear, polynomial and rational moment conditions, it is often the case in applied models that moment conditions include other forms of non-linearity. Hansen (1982a), for example, considers moment conditions of the form

$$\mathbb{E}\left[z_t\left(\beta_0 x_{t+1}^{\alpha_0} w_{t+1} - 1\right)\right] = 0, \tag{5.1}$$

for which our method is not appropriate. However, one possible approach for the general nonlinear case is motivated by what we find in Example 1. While the objective functions depicted in Figure 2 are very ill-behaved, the entries of matrix $L(\theta)$ are all either quadratic, constant or linear in θ. If we can expect the general nonlinear case to present a similarly well behaved structure in $L(\theta)$, we can consider an approach by approximation: first approximate the entries of $L(\theta)$ by either polynomial or rational functions, and then compute eigenvalues by linearization. For the single-parameter case, $L(\theta)$ is of dimensions $(2m+1)\times(2m+1)$, but only has $2\left(\binom{m+1}{2}+m\right)$ unique nonzero entries which require approximating. In fact, there is a substantial literature in nonlinear eigenvalue problems that follows similar procedures. A particularly interesting approach is the *infinite Arnoldi algorithm* of Jarlebring et al. (2012), which performs a dynamic rational approximation that achieves increasingly good approximations at every iteration. While we do not explore this literature further, we end this section by presenting a toy implementation of an approximation-linearization procedure that achieves surprisingly good results given its simplicity.

Example 5 (An experiment for the general nonlinear case)**.** Let $x_{ki} \sim \chi^2_1$ for $k = 1, 2, 3$ and $i = 1, \ldots, N$. We define

$$y_{ki} = x_{ki}^{\theta_0} + \varepsilon_{ki}, \qquad \varepsilon_{ki} \sim \mathcal{N}(0, 1),$$

and consider the moment function $g_i(\theta) = y_i - x_i^{\theta}$. With 3 moments, $L(\theta)$ is 7×7 matrix. We need to approximate 18 of its 49 entries, as all other entries are either repeated or zero.

A canonical method in approximation theory is Chebyshev interpolation, which approximates a function as a linear combination of orthogonal polynomials $t_i(x)$ given by the recurrence relation

$$t_0(x) = 1; \quad t_1(x) = x; \quad t_{i+1}(x) = 2x\, t_i(x) - t_{i-1}(x). \tag{5.2}$$

This choice of polynomials is particularly useful in approximations since they are orthogonal

in the interval $[-1, 1]$ with respect to the weight function $(1 - x^2)^{\frac{1}{2}}$, which assigns infinite weight to values at the boundary of the approximation interval. This mitigates the problem of oscillations at the edges of the interval leading to large errors (Runge's phenomenon). We employ the *Julia* package *ApproxFun* (Olver and Townsend, 2014) and set the degree of the approximating polynomials to $d = 100$. We obtain an approximation of $L(\theta)$ in the form

$$\widetilde{L}(\theta) = \sum_{i}^{d} A_i \, t_i(\theta), \tag{5.3}$$

where each A_i is a linear combination of matrices of ones and zeros weighted by the coefficients of the i-th Chebyshev polynomial in each approximation.

In Proposition 3.2, we presented a linearization for polynomial eigenvalue problems that works for polynomial pencils in a *monomial basis*. That is, pencils that are written as products of constant matrices and powers of λ. While we could rewrite (5.3) in terms of a monomial basis, we can also use a linearization that exploits the recurrence relation of the Chebyshev polynomials in (5.2):

$$\left(\begin{bmatrix} 0 & I & 0 & \cdots & 0 \\ I & 0 & I & \cdots & 0 \\ \vdots & \ddots & \ddots & \ddots & \vdots \\ 0 & \cdots & I & 0 & I \\ A_0 & \cdots & A_{d-3} & A_{d-2} & A_{d-1} \end{bmatrix} - \theta \begin{bmatrix} I & 0 & 0 & \cdots & 0 \\ 0 & 2I & & \ddots & \vdots \\ 0 & & \ddots & & 0 \\ \vdots & \ddots & & 2I & 0 \\ 0 & \dots & 0 & 0 & 2A_d \end{bmatrix} \right) \begin{pmatrix} t_0(\theta)x \\ t_1(\theta)x \\ \vdots \\ t_{d-2}(\theta)x \\ t_{d-1}(\theta)x \end{pmatrix} = 0. \tag{5.4}$$

Which we can then solved by the QZ algorithm. For Monte Carlo simulations, we set $\theta_0 = 1$ and we consider an interval of approximation to $[-2, 2]$. Figure 8 depicts objective functions and approximated critical points for $N = 10$. The shaded area is outside the interval of approximation. In this particular case, the method seems to work remarkably well inside the interval of approximation.

While our approximation-linearization method for the general nonlinear case is still incipient, we hope that this toy implementation serves as a proof of concept to motivate future work.

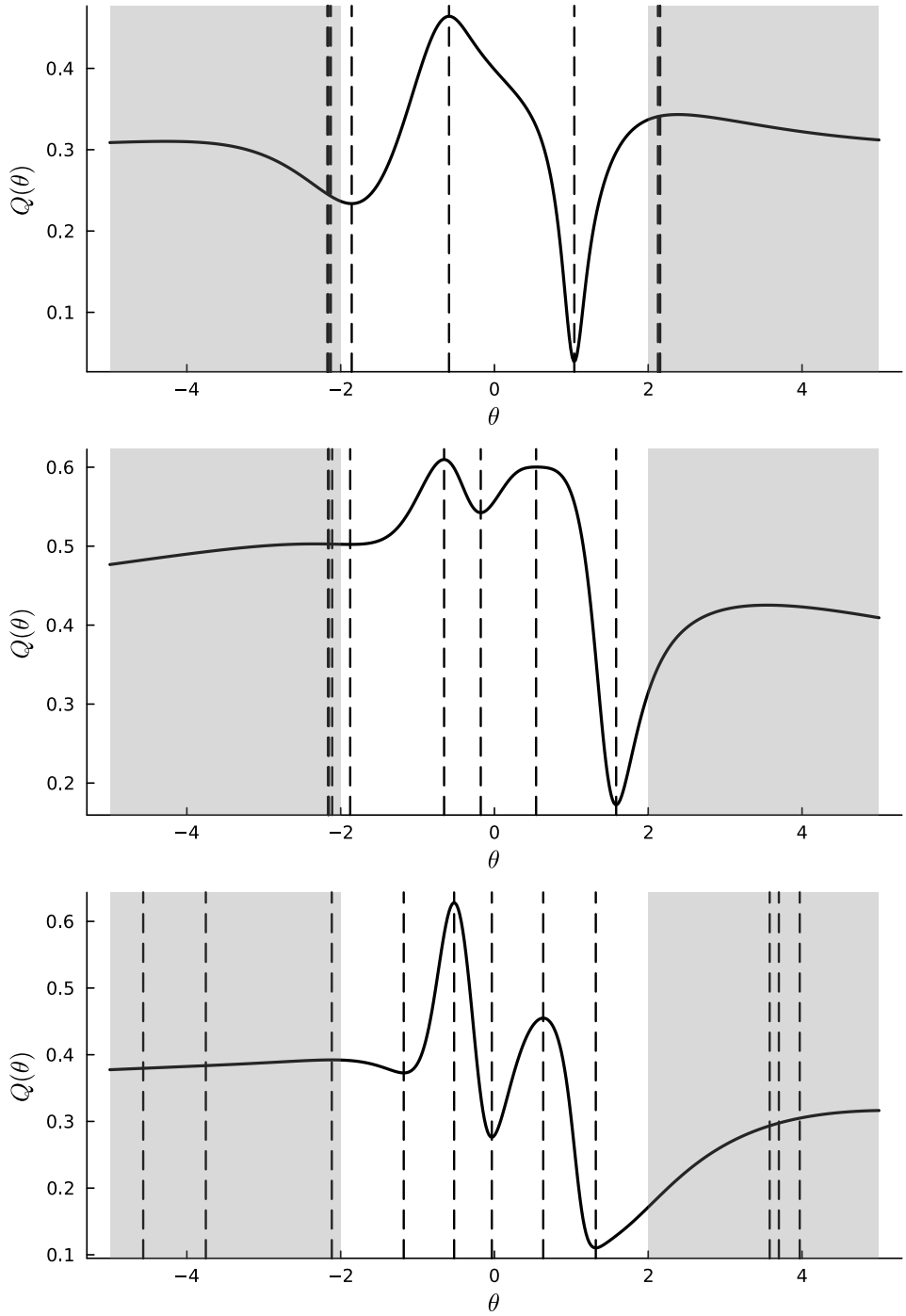


Figure 8: CUE objective functions and approximate critical points.

A Appendix

A bissection method for computing the CUE

Algorithm 1 describes in pseudocode the bissection procedure for computing CUE estimates discussed in Section (3.4).

Algorithm 1 Bissection method for computing the CUE

```
θ̄ ← {0}                                          ▷ initial guess
L ← 0
U ← Q(θ̄)                                         ▷ lower and upper limits of the bissection interval.
while not halting_conditions do
    θ̃ ← real eigenvalues of M̄ (θ; (L+U)/2)
    if θ̃ is empty then
        L ← (L+U)/2
    else
        U ← (L+U)/2
        θ̄ ← θ̃
    end if
end while
Return θ̂_CUE ← mean(θ̄)
```

A Rayleigh quotient interpretation of the CUE objective function

Recall the expression for the CUE objective function in Example (1):

$$Q(\beta) = b'Y'Z\left((b \otimes I_m)\Sigma(b \otimes I_m)\right)^{-1} Z'Yb. \tag{A.1}$$

Using the fact that $Z'Yb = (b' \otimes I_m)\,\mathrm{vec}(Z'Y)$, we write the problem of minimizing (A.1) as

$$\min_{\beta,B} \quad r'B\left(B'\Sigma B\right)^{-1} B'r \quad \text{subject to} \quad B = (b \otimes I_m), \tag{A.2}$$

where $r \coloneqq \mathrm{vec}(Z'Y)$. Consider a relaxation of problem (A.2) without the constraint $B = (b \otimes I_m)$. In the relaxed problem, we were free to choose all $pm \times m$ entries of B. Homogeneity in the columns of B allow for introducing the normalization $B'\Sigma B = I_m$. We then have $r'B(B'\Sigma B)^{-1}B'r = r'BB'r = \mathrm{tr}(B'rr'B)$ and (A.2) is equivalent to

$$\min_{B} \mathrm{tr}(B'rr'B) \quad \text{subject to} \quad B'\Sigma B = I_m. \tag{A.3}$$

The first order conditions of (A.3) are

$$rr'B = \Sigma B\Lambda,$$

where Λ is an $m \times m$ diagonal matrix of lagrange multipliers (Boyd and Vandenberghe, 2004). The first order conditions can are solved by taking the m diagonal elements of Λ to be any of the mp eigenvectors of the GEP (rr', Σ) and the columns of B to be their associated eigenvalues. Since (A.3) is a minimization problem and the first order conditions imply $\mathrm{tr}(B'rr'B) = \mathrm{tr}(B'\Sigma B\Lambda) = \mathrm{tr}(\Lambda)$, the minimum is equal to the sum of the m smallest eigenvalues. Unfortunately, since rr' is rank 1, the GEP (rr', Σ) has a single nonzero eigenvalue, which reveals that this relaxation does not provide much insight on the original problem. However, it is interesting to note that what keeps problem (A.2) from having a solution from a generalized eigenvalue problem is the simple linear constraint $B = (b \otimes I_m)$.

Bibliography

Joseph G. Altonji and Lewis M. Segal. Small-sample bias in gmm estimation of covariance structures. *Journal of Business & Economic Statistics*, 14(3):353–366, 1996. ISSN 07350015.

E. Anderson, Z. Bai, C. Bischof, S. Blackford, J. Demmel, J. Dongarra, J. Du Croz, A. Greenbaum, S. Hammarling, A. McKenney, and D. Sorensen. *LAPACK Users' Guide*. Society for Industrial and Applied Mathematics, Philadelphia, PA, third edition, 1999. ISBN 0-89871-447-8 (paperback).

T. W. Anderson and Herman Rubin. Estimation of the parameters of a single equation in a complete system of stochastic equations. *The Annals of Mathematical Statistics*, 20(1):46–63, 1949.

Isaiah Andrews and Anna Mikusheva. Conditional inference with a functional nuisance parameter. *Econometrica*, 84(4):1571–1612, 2016.

Manuel Arellano and Stephen Bond. Some tests of specification for panel data: Monte carlo evidence and an application to employment equations. *The Review of Economic Studies*, 58 (2):277–297, 1991.

F. V. Atkinson. Multiparameter spectral theory. *Bulletin of the American Mathematical Society*, 74(1):1 – 27, 1968.

F.V. Atkinson. *Multiparameter Eigenvalue Problems: Matrices and compact operators*. Mathematics in science and engineering. Academic Press, 1972. ISBN 9780120658015.

Z. Bai, G. Sleijpen, H. van der Vorst, R. Lippert, and A. Edelman. *9. Nonlinear Eigenvalue Problems*, pages 281–314.

Paul A. Bekker. Alternative approximations to the distributions of instrumental variable estimators. *Econometrica*, 62(3):657–681, 1994.

Jeff Bezanson, Alan Edelman, Stefan Karpinski, and Viral B Shah. Julia: A fresh approach to numerical computing. *SIAM review*, 59(1):65–98, 2017.

S.P. Boyd and L. Vandenberghe. *Convex Optimization*. Number pt. 1 in Berichte über verteilte messysteme. Cambridge University Press, 2004.

Craig Burnside and Martin Eichenbaum. Small-sample properties of gmm-based wald tests. *Journal of Business & Economic Statistics*, 14(3):294–308, 1996. ISSN 07350015.

Gary Chamberlain. Chapter 22 panel data. volume 2 of *Handbook of Econometrics*, pages 1247–1318. Elsevier, 1984.

Gary Chamberlain. Asymptotic efficiency in estimation with conditional moment restrictions. *Journal of Econometrics*, 34(3):305–334, 1987. ISSN 0304-4076.

John C. Chao and Norman R. Swanson. Consistent estimation with a large number of weak instruments. *Econometrica*, 73(5):1673–1692, 2005.

D. Cox, J. Little, and D. OSHEA. *Ideals, Varieties, and Algorithms: An Introduction to Computational Algebraic Geometry and Commutative Algebra*. Undergraduate Texts in Mathematics. Springer New York, 2013.

James Demmel and Bo Kågström. The generalized schur decomposition of an arbitrary pencil $a - \lambda b$ —robust software with error bounds and applications. part i: theory and algorithms. *ACM Trans. Math. Softw.*, 19(2):160–174, jun 1993.

James Demmel and Bo Kågström. GUPTRI, software for singular pencils. http://www8.cs.umu.se/research/nla/singular_pairs/guptri/, 1999. Accessed: February 13, 2024.

Martin Eichenbaum. Some empirical evidence on the production level and production cost smoothing models of inventory investment. *American Economic Review*, 79(4):853–64, 1989.

Heike Fassbender, Javier Pérez, and Nikta Shayanfar. Constructing symmetric structure-preserving strong linearizations. *ACM Commun. Comput. Algebra*, 50(4):167–169, feb 2017.

J. G. F. Francis. The QR Transformation A Unitary Analogue to the LR Transformation—Part 1. *The Computer Journal*, 4(3):265–271, 01 1961. ISSN 0010-4620.

I. Gohberg, P. Lancaster, and L. Rodman. *Matrix Polynomials*. Classics in Applied Mathematics. Society for Industrial and Applied Mathematics (SIAM, 3600 Market Street, Floor 6, Philadelphia, PA 19104), 1982.

G.H. Golub and C.F. Van Loan. *Matrix Computations*. Johns Hopkins Studies in the Mathematical Sciences. Johns Hopkins University Press, 2013.

A.R. Hall. *Generalized Method of Moments*. Advanced texts in econometrics. Oxford University Press, 2005.

Peter Hall and Joel L. Horowitz. Bootstrap critical values for tests based on generalized-method-of-moments estimators. *Econometrica*, 64(4):891–916, 1996. ISSN 00129682, 14680262.

Chirok Han and Peter C. B. Phillips. Gmm with many moment conditions. *Econometrica*, 74(1): 147–192, 2006.

B. Hansen. *Econometrics*. Princeton University Press, 2022.

Lars Peter Hansen. Large sample properties of generalized method of moments estimators. *Econometrica*, 50(4):1029–1054, 1982a.

Lars Peter Hansen. Large sample properties of generalized method of moments estimators. *Econometrica*, 50(4):1029–1054, 1982b.

Lars Peter Hansen, John Heaton, and Amir Yaron. Finite-sample properties of some alternative gmm estimators. *Journal of Business & Economic Statistics*, 14(3):262–280, 1996.

Jerry Hausman, Randall Lewis, Konrad Menzel, and Whitney Newey. Properties of the cue estimator and a modification with moments. *Journal of Econometrics*, 165(1):45–57, 2011.

Jerry A. Hausman, Whitney K. Newey, Tiemen Woutersen, John C. Chao, and Norman R. Swanson. Instrumental variable estimation with heteroskedasticity and many instruments. *Quantitative Economics*, 3(2):211–255, 2012.

Michiel E. Hochstenbach, Christian Mehl, and Bor Plestenjak. Solving singular generalized eigenvalue problems by a rank-completing perturbation. *SIAM Journal on Matrix Analysis and Applications*, 40(3):1022–1046, 2019.

G. Imbens. A New Approach to Generalized Method on Moments Estimation. Harvard Institute of Economic Research Working Papers 1633, Harvard - Institute of Economic Research, 1993.

Guido Imbens, Stephen Donald, and Whitney Newey. Choosing the number of moments in conditional moment restriction models. *Whitney Newey*, 04 2010.

Guido W Imbens and Richard Spady. Confidence intervals in generalized method of moments models. *Journal of econometrics*, 107(1-2):87–98, 2002.

Guido W. Imbens, Richard H. Spady, and Phillip Johnson. Information theoretic approaches to inference in moment condition models. *Econometrica*, 66(2):333–357, 1998a.

Guido W. Imbens, Richard H. Spady, and Phillip Johnson. Information theoretic approaches to inference in moment condition models. *Econometrica*, 66(2):333–357, 1998b.

Elias Jarlebring, Wim Michiels, and Karl Meerbergen. A linear eigenvalue algorithm for the nonlinear eigenvalue problem. *Numerische Mathematik*, 122, 09 2012.

A Källström and BD Sleeman. A left definite multiparameter eigenvalue problem in ordinary differential equations. *Proceedings of the Royal Society of Edinburgh Section A: Mathematics*, 74:145–155, 1976.

Yuichi Kitamura and Michael Stutzer. An information-theoretic alternative to generalized method of moments estimation. *Econometrica*, 65(4):861–874, 1997.

Frank Kleibergen. Testing parameters in gmm without assuming that they are identified. *Econometrica*, 73(4):1103–1123, 2005.

J.R. Magnus and H. Neudecker. *Matrix Differential Calculus with Applications in Statistics and Econometrics*. Wiley Series in Probability and Statistics. Wiley, 2019.

Volker Mehrmann and Christian Schröder. Nonlinear eigenvalue and frequency response problems in industrial practice. *Journal of Mathematics in Industry*, 1, 07 2011.

Ron C Mittelhammer, George G Judge, and Ron Schoenberg. Empirical evidence concerning the finite sample performance of el-type structural equation estimation and inference methods. *Chapter*, 12:282–305, 2005.

C. B. Moler and G. W. Stewart. An algorithm for generalized matrix eigenvalue problems. *SIAM Journal on Numerical Analysis*, 10(2):241–256, 1973.

Humberto Moreira and Marcelo J. Moreira. Optimal two-sided tests for instrumental variables regression with heteroskedastic and autocorrelated errors. *Journal of Econometrics*, 213(2): 398–433, 2019.

Marcelo Moreira. Tests with correct size when instruments can be arbitrarily weak. *Journal of Econometrics*, 152:131–140, 10 2009a.

Marcelo J. Moreira. A conditional likelihood ratio test for structural models. *Econometrica*, 71 (4):1027–1048, 2003.

Marcelo J. Moreira. A maximum likelihood method for the incidental parameter problem. *The Annals of Statistics*, 37(6A):3660 – 3696, 2009b.

Marcelo J. Moreira, Mahrad Sharifvaghefi, and Whitney K. Newey. Robust method for estimation and testing in linear gmm settings with weak instruments: bridging theory and practice. Unpublished Manuscript, FGV/EPGE, 2023.

Whitney K Newey. Many weak moment asymptotics for the continuously updated gmm estimator. *Department of Economics, MIT*, September 2004.

Whitney K. Newey and Daniel McFadden. Chapter 36 large sample estimation and hypothesis testing. volume 4 of *Handbook of Econometrics*, pages 2111–2245. Elsevier, 1994.

Whitney K. Newey and Richard J. Smith. Higher order properties of gmm and generalized empirical likelihood estimators. *Econometrica*, 72(1):219–255, 2004.

Whitney K. Newey and Kenneth D. West. A simple, positive semi-definite, heteroskedasticity and autocorrelation consistent covariance matrix. *Econometrica*, 55(3):703–708, 1987.

Whitney K. Newey and Frank Windmeijer. Generalized method of moments with many weak moment conditions. *Econometrica*, 77(3):687–719, 2009.

Sheehan Olver and Alex Townsend. A practical framework for infinite-dimensional linear algebra. In *Proceedings of the 1st Workshop for High Performance Technical Computing in Dynamic Languages – HPTCDL '14*. IEEE, 2014.

Art B. Owen. Empirical likelihood ratio confidence intervals for a single functional. *Biometrika*, 75(2):237–249, 1988.

Bor Plestenjak, Michiel Hochstenbach, and Tomaz Kosir. Numerical methods for rectangular multiparameter eigenvalue problems, with applications to finding optimal arma and lti models. *Numerical Linear Algebra with Applications*, 2023:e2540, 11 2023.

Jin Qin and Jerry Lawless. Empirical Likelihood and General Estimating Equations. *The Annals of Statistics*, 22(1):300 – 325, 1994.

Kumar Singh and Yitshak Ram. Transcendental eigenvalue problem and its applications. *AIAA journal*, 40:1402–1407, 07 2002.

Richard J. Smith. Alternative semi-parametric likelihood approaches to generalised method of moments estimation. *The Economic Journal*, 107(441):503–519, 1997.

James H. Stock and Jonathan H. Wright. Gmm with weak identification. *Econometrica*, 68(5): 1055–1096, 2000.

Yangfeng Su and Zhaojun Bai. Solving rational eigenvalue problems via linearization. *SIAM J. Matrix Analysis Applications*, 32:201–216, 05 2011.

C. F. Van Loan and N. Pitsianis. *Approximation with Kronecker Products*, pages 293–314. Springer Netherlands, Dordrecht, 1993.

Christof Vermeersch and Bart De Moor. Globally optimal least-squares arma model identification is an eigenvalue problem. *IEEE Control Systems Letters*, 3(4):1062–1067, 2019.

Heinrich Voss. A rational spectral problem in fluid-solid vibration. *Electronic Transactions on Numerical Analysis*, 16:94 –, 01 2003.