ECONSTOR
Make Your Publications Visible.

A Service of

ZBW Leibniz-Informationszentrum Wirtschaft
Leibniz Information Centre for Economics

Tryphonides, Andreas

Working Paper

Conditional moment restrictions and the role of density information in estimated structural models

SFB 649 Discussion Paper, No. 2017-016

Provided in Cooperation with:
Collaborative Research Center 649: Economic Risk, Humboldt University Berlin

Suggested Citation: Tryphonides, Andreas (2017) : Conditional moment restrictions and the role of density information in estimated structural models, SFB 649 Discussion Paper, No. 2017-016, Humboldt University of Berlin, Collaborative Research Center 649 - Economic Risk, Berlin

This Version is available at:
https://hdl.handle.net/10419/169206

Standard-Nutzungsbedingungen:

Die Dokumente auf EconStor dürfen zu eigenen wissenschaftlichen Zwecken und zum Privatgebrauch gespeichert und kopiert werden.

Sie dürfen die Dokumente nicht für öffentliche oder kommerzielle Zwecke vervielfältigen, öffentlich ausstellen, öffentlich zugänglich machen, vertreiben oder anderweitig nutzen.

Sofern die Verfasser die Dokumente unter Open-Content-Lizenzen (insbesondere CC-Lizenzen) zur Verfügung gestellt haben sollten, gelten abweichend von diesen Nutzungsbedingungen die in der dort genannten Lizenz gewährten Nutzungsrechte.

Terms of use:

Documents in EconStor may be saved and copied for your personal and scholarly purposes.

You are not to copy documents for public or commercial purposes, to exhibit the documents publicly, to make them publicly available on the internet, or to distribute or otherwise use the documents in public.

If the documents have been made available under an Open Content Licence (especially Creative Commons Licences), you may exercise further usage rights as specified in the indicated licence.

WWW.ECONSTOR.EU

Mitglied der Leibniz-Gemeinschaft

SFB 649 Discussion Paper 2017-016

Conditional moment restrictions and the role of density information in estimated structural models

Andreas Tryphonides*

* Humboldt-Universität zu Berlin, Germany

This research was supported by the Deutsche Forschungsgemeinschaft through the SFB 649 "Economic Risk".

http://sfb649.wiwi.hu-berlin.de
ISSN 1860-5664

SFB 649, Humboldt-Universität zu Berlin
Spandauer Straße 1, D-10178 Berlin

SFB 649 ECONOMIC RISK BERLIN

CONDITIONAL MOMENT RESTRICTIONS AND THE ROLE OF DENSITY INFORMATION IN ESTIMATED STRUCTURAL MODELS

ANDREAS TRYPHONIDES

HUMBOLDT UNIVERSITY

Abstract

While incomplete models are desirable due to their robustness to misspecification, they cannot be used to conduct full information exercises i.e. counterfactual experiments and predictions. Moreover, the performance of the corresponding GMM estimators is fragile in small samples. To deal with both issues, we propose the use of an auxiliary conditional model for the observables $f(X|Z,\varphi)$, where the equilibrium conditions $\mathbb{E}(m(X,\vartheta)|Z)=0$ are imposed on $f(X|Z,\varphi)$ using information projections, and (ϑ,φ) are estimated jointly. We provide the asymptotic theory for parameter estimates for a general set of conditional projection densities, under correct and local misspecification of $f(X|Z,\varphi)$. In either cases, efficiency gains are significant. We provide simulation evidence for the Mean Squared Error (MSE) both under the case of local and fixed density misspecification and apply the method to the prototypical stochastic growth model. Moreover, we illustrate that given $(\hat{\vartheta},\hat{\varphi})$ it is now feasible to do counterfactual experiments without explicitly solving for the equilibrium law of motion.

JEL Classification: C13, C14 , E10

Keywords: Incomplete models, Information projections, Small Samples, Shrinkage

Spandauerstr 1, 10178,Berlin,Germany.

E-mail address: andreas.tryphonides@hu-berlin.de.

Date: 17/07/2017.

This paper is a substantially revised version of chapter 3 of my PhD thesis (European University Institute). I thank Fabio Canova for his advice and the rest of the thesis committee: Peter Reinhard Hansen, Giuseppe Ragusa and Frank Schorfheide for comments and suggestions. Earlier versions of this paper (circulated with different titles) greatly benefited from discussions with Jack Porter, Raffaella Giacomini, George Tauchen and the participants at the 23rd MEG (Bloomington), the 1st IAAE (London), the 68th ESEM (Toulouse), the 4th International Conference in memory of Carlo Giannini (Pavia), the Econometrics Study Group (Bristol), the EUI Econometrics Working Group and the University of Wisconsin Madison lunch seminar. Any errors are my own.

1. Introduction

The use of estimated structural models has become pervasive in both academia and economic policy institutions. In order to answer quantitative questions within a data coherent framework, practitioners have resorted to a variety of full or limited information methods. Nevertheless, while economic theory provides a set of equilibrium conditions, it rarely dictates the complete probability distribution of observables. The latter is necessary to perform full information analysis i.e. counter-factual experiments and probabilistic forecasts, and this forces users to make several auxiliary assumptions. For example, one has to choose which solution concept to use and type (and degree) of approximation to consider.

Although approximations make computation of the solution of the model easier, this can possibly cause a form of misspecification with respect to the exact model. Approximations to non linear models might not necessarily work well, as they can distort the dynamics implied by the model (den Haan and de Wind, 2010). Distorting the dynamics can lead to severely wrong inference about parameters and policy recommendations. Moreover, as shown by Canova and Sala (2009), approximation and model solution can introduce further uncertainties like loss of identification.

With regard to the types of equilibria considered, although some equilibria can be easily discarded due to economic reasoning, it is often the case that this is done with not so strong evidence Pesaran (1987); Blanchard (1979). Different types of equilibria are a priori equally plausible, and selecting one type of equilibrium can have important implications for inference regarding the effectiveness of policy. A classic example is the determination of inflation and the identification of fiscal monetary regimes as discussed in Leeper and Leith (2016).

The most prominent approach to estimating models that are not completely specified is the Generalized Method of Moments (GMM) and its variants (Hansen, 1982). Nevertheless, the performance of GMM is distorted in small samples (Hansen, Heaton, and Yaron, 1996). This paper considers an alternative method for estimating the parameters

of a dynamic structural model which does not require the equilibrium decision rules and produces an estimated probability model for the observables. We propose the use of what we refer to as a "base" conditional probability measure with density $f(X|Z,\varphi)$ where Z is conditioning information. This measure can be generally interpreted as an approximate model for the observables. Utilizing a variation of the method of information projections Kitamura and Stutzer (1997); I.Csiszar (1975) we obtain a probability distribution that satisfies the *conditional* restrictions of the economic model, that is $\mathbb{E}(m(X,\vartheta)|z)=0$, and is as close as possible to the base measure. This is also related to the recent work of Giacomini and Ragusa (2014) in a forecasting context.

We develop the corresponding frequentist inference, while we limit most of our analysis to the case of finite dimensional φ. However, extensions under suitable assumptions are possible[1]. Furthermore, we deal with correctly specified or locally misspecified classes of $f(X|Z,\varphi)$. In case of local misspecification, we show that the proposed method is akin to shrinkage towards the approximate model. More interestingly, an explicit form of the asymptotic variance of the estimator is provided. Under the condition that there exists an admissible parameter of $f(X|Z,\varphi)$ such that the moment conditions are satisfied, the efficiency attained is higher than the semi-parametric lower bound obtained using only sample information (see Chamberlain (1987)). The reason for this result is that since we are using more information on the density, the estimator automatically generates more valid moment restrictions than the purely non-parametric case, and efficiency therefore increases. Moreover, local misspecification of the density in the form of improper finite dimensional restrictions leads to even more efficiency gains and therefore an asymptotic bias - variance trade-off. We provide simulation comparisons of the Mean Squared Error (MSE) of the estimator for the case of local and non local density misspecification which corroborate our theoretical results. We also apply the method to simulated data from the prototypical stochastic growth model, the results of which we report in Appendix C.

[1]Independent work by Shin (2014) proposes Bayesian algorithms to implement the exponential tilting estimation using flexible mixtures of densities. Our contribution is mostly on the frequentist properties of exponential tilting for a general parametric family of densities and our results are therefore complementary

The strand of literature that is closer to the methodology considered in this paper is the literature on Exponential Tilting i.e. Schennah (2007); Kitamura and Stutzer (1997); Imbens, Spady, and Johnson (1998), and Generalized Empirical Likelihood criteria i.e. Newey and Smith (2004) in a conditional moment restrictions framework. Formally, our estimator is not an extension of GEL criteria, in the same way the ETEL estimator (Schennah (2007)) cannot be obtained as a particular version of GEL estimator. We depart from this literature by considering a generalized version of exponential tilting in the "first step", where the form of $f(X|Z,\varphi)$ is parametrically specified.

The paper is organized as follows. In Section 2, we introduce information projections and we provide an asset pricing example. In Section 3 we outline the large sample properties under correct specification of $f(X|Z,\varphi)$. Section 4 provides a formal shrinkage formulation and the asymptotic distribution in case of local misspecification while Section 5 provides simulation evidence. Section 6 concludes. Appendix A provides some analytical details for the example and discusses the computational aspect of the method and the case of non differentiable models. Appendix B contains some of the proofs, while the rest are in the supplemental material. Appendix C contains further Monte Carlo results and a basic application on simulated data.

Finally, a word on notation. Let N_0 denote the length of the data and N_s the length of simulated series. X is an $n_x \times 1$ vector of the variables of interest while Z is an $n_z \times 1$ vector of conditioning variables. Both X and Z induce a probability space $(\Omega, \mathcal{F}, \mathbb{P})$. In the paper three different probability measures are used, the true measure $\mathbb{P}$, the base measure F_φ which is indexed by parameters φ and the $\mathcal{H}_{(\varphi,\vartheta)}$ measure which is obtained after the information projection. Moreover, these measures are considered absolutely continuous with respect to a dominating measure v, where v in most interesting cases is the Lebesgue measure. All these measures possess the corresponding density functions p, f and h. The set of parameters ψ is decomposed in $\vartheta \in \Theta$, the set of structural (economic) parameters and φ the parameters indexing the density $f(X|Z,\varphi)$. In addition, P_s is the conditional distribution where s can be a variable or a parameter. Furthermore,

$m^l(X, Z, \vartheta)$ is a general $X \otimes Z$ measurable moment function and $m(X, Z, \vartheta)$ is an $n_m \times 1$ vector containing these functions. For any matrix function D_i, the subscript i denotes the evaluation at datum (x_i, z_i). The operator $\rightarrow_p$ signifies convergence in probability and $\rightarrow_d$ convergence in distribution; $\mathcal{N}(.,.)$ signifies the Normal distribution with certain mean and variance. In terms of norms, $||.||$ signifies the Euclidean norm unless otherwise stated. In addition $||.||_{TV}$ is the Total Variation distance[2]. $\mathbb{E}_P$ and is the mathematical expectations operator with respect to measure P. Finally, $\mathbb{V}_P(x)$ signifies the variance of variable x under the $P-$ measure while $V_{\tilde{P},s}(x)$ is the second moment of a particular function $\tilde{s}(.)$.

2. Information Projections as Perturbations to the Base Measure

For completeness, we present below the formal problem of an information projection. Given a class of candidate base densities $f(X, Z|\psi)$, a conditional information projection is equivalent to solving for the following program:

$$\min_{h(X|Z,\varphi)\in\mathcal{H}_\theta} \int h(X|Z,\psi) log\left(\frac{h(X|Z,\psi)}{f(X|Z,\varphi)}\right) h(Z) d(X,Z) \tag{2.1}$$

where

$$a) \quad \mathcal{H}_\theta \;:=\; \left\{h \in \mathcal{L}_p : \int h(X|Z,\psi) m(X,Z,\theta) dX = 0, \quad \int h(X|Z,\psi) dX = 1, \quad Z \quad a.e.\right\}$$

In the information projections literature the minimization problem in 2.1 subject to constraint (a) is called exponential tilting as the distance metric *minimized* is the Kullback Leibler distance, whose convex conjugate has an exponential form.

The set $\mathcal{H}_\theta$ is the set of admissible densities i.e. the densities that by construction satisfy the moment conditions. Given this class of densities, we perform extremum estimation

[2] $||.||_{TV} = \sup_{B\in\Omega} \int_B |f - p| dv$

using the log likelihood function as follows:

$$\max_{\psi\in\Psi} \int \log(h^{\star}(X|Z,\psi))d\mathbb{P}(X,Z) \tag{2.2}$$

The above problem can be conveniently rewritten such that the choice of density $h(X|Z,\theta)$ is equivalent to the choice of a perturbation $\mathcal{M}(X,Z,\theta)$ to the prior density, that is $h(X|Z,\vartheta,\varphi) = f(X|Z,\varphi)\mathcal{M}(X,Z,\vartheta)$. The perturbation factor $\mathcal{M}(X,Z,\vartheta)$ will be a function of the sufficient information to estimate θ and is in general not unique. Selecting $h(X|Z,\vartheta,\varphi)$ by minimizing the Kullback-Leibler distance to the prior density is one way of selecting a unique factor $\mathcal{M}$. The program therefore becomes as follows:

$$\min_{\mathcal{M}\in\mathbb{M}} \mathbb{E}_{f(X|Z,\varphi)h(Z)}\mathcal{M}(X,Z,\vartheta)\log\mathcal{M}(X,Z,\vartheta)$$

where

$$\mathbb{M} := \Big\{\mathcal{M}\in\mathcal{L}_p : \mathbb{E}_{f(X|Z,\varphi)}\mathcal{M}(X,Z,\theta)m(X,Z,\theta) = 0$$
$$\mathbb{E}_{f(X|Z,\varphi)}\mathcal{M}(X,Z,\theta) = 1\Big\}$$

The solution to the above problem, that is, the optimal perturbation factor is the following:

$$\mathcal{M}^{\star} \;=\; \exp\left(\lambda(Z) + \mu(Z)'m(X,Z,\vartheta)\right)$$

which implies the choice of the following family of distributions:

$$h(X|Z,\psi) = f(X|Z,\varphi)\exp\left(\lambda(Z) + \mu(Z)'m(Y,\vartheta)\right) \tag{2.3}$$

where μ is the vector of the Lagrange multiplier functions enforcing the conditional moment conditions on $f(X|Z,\varphi)$ and λ is a scaling function.

Had we used an alternative objective function to (2), e.g. another particular case from the general family of divergences in Cressie and Read (1984), this would result to

a different form for $h^\star(X|Z,\psi)$. Under correct specification for $f(X|Z,\varphi)$, this choice does not matter asymptotically, while it matters in finite samples. Exponential tilting ensures a positive density function $h^\star$ while it has been shown that it is robust under misspecification of the moment conditions Schennah (2007).

Moreover, in the case in which $f(X|Z,\varphi)$ belongs to the exponential family and the moment conditions are linear, exponential tilting is the natural choice. We present an illustrative example of projecting on densities that satisfy moment conditions that arise from economic theory. In this simple case, due to linearity, the resulting distribution after the change of measure implied by the projection is conjugate to the prior. Economic theory therefore imposes structure on the moments of the prior density.

2.1. **An Example from Asset Pricing.** Consider the restrictions implied by the consumption - savings decision of the representative household on the joint stochastic process of consumption, C_t, and gross interest rate, R_t. This means that they should satisfy the following Euler equation:

$$\mathbb{E}_{\mathbb{P}}(\beta R_{t+1} U_c(C_{t+1}) - U_c(C_t)|\mathcal{F}_t) = 0$$

where $\mathcal{F}_t$ is the information set of the agent at time t and $U(C_t) = C_t^2$. Under Rational expectations, the agent uses the objective probability measure to formulate expectations.

Suppose that a prior statistical model is a bivariate VAR for consumption and the interest rate which, for analytical tractability, are not correlated. Their joint density conditional on $\mathcal{F}_t$ is therefore:

$$\begin{pmatrix} c_{t+1} \\ R_{t+1} \end{pmatrix} |\mathcal{F}_t \sim N\left(\begin{pmatrix} \rho_c c_t \\ \rho_R R_t \end{pmatrix}, \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix} \right)$$

Given the assumption on the utility function, $\mathbb{E}(R_{t+1}C_{t+1}|\mathcal{F}_t) = \frac{c_t}{\beta}$. This is a covariance restriction as it implies that $Cov(R_{t+1}, C_{t+1}|\mathcal{F}_t) = \frac{C_t}{\beta}(1 - R_t\beta\rho_c\rho_R)$. The new density

$h(C_{t+1}, R_{t+1}|\mathcal{F}_t)$ is therefore:

$$\begin{pmatrix} C_{t+1} \\ R_{t+1} \end{pmatrix} |\mathcal{F}_t \sim N\left(\begin{pmatrix} \rho_c C_t \\ \rho_R R_t \end{pmatrix}, \begin{pmatrix} 1 & \frac{C_t}{\beta}(1 - R_t\beta\rho_c\rho_R) \\ * & 1 \end{pmatrix}\right)$$

Since we know the new density in this case, the perturbation $\mathcal{M}(X, Z; \vartheta)$, can be reverse engineered as follows:

$$\begin{aligned} \mathcal{M} &= \left[N\left(\begin{pmatrix} \rho_c C_t \\ \rho_R R_t \end{pmatrix}, I_2\right)\right]^{-1} N\left(\begin{pmatrix} \rho_c C_t \\ \rho_R R_t \end{pmatrix}, \begin{pmatrix} 1 & \frac{C_t}{\beta}(1 - R_t\beta\rho_c\rho_R) \\ * & 1 \end{pmatrix}\right) \\ &= \exp\left(-\frac{1}{2}\begin{pmatrix} C_{t+1} - \rho_c C_t \\ R_{t+1} - \rho_R R_t \end{pmatrix}' \begin{pmatrix} 1 & \frac{C_t}{\beta}(1 - R_t\beta\rho_c\rho_R) \\ * & 1 \end{pmatrix} \begin{pmatrix} C_{t+1} - \rho_c C_t \\ R_{t+1} - \rho_R R_t \end{pmatrix}\right) \end{aligned}$$

In Appendix A, we illustrate how the same expression for $\mathcal{M}$ can be obtained formally using a conditional density projection[3], that is, solving 2.1 subject to the first constraint (a). Note that in this example, the fact that the Euler equation is a direct restriction on the parameters of the base density is an artifact of the form of the utility function assumed, and is therefore a special case. In more general examples an analytical solution cannot be easily obtained and we therefore resort to simulation. Details of the algorithm are provided in Appendix A.

In the rest of the paper we analyze the frequentist properties of using the tilted density to estimate $\psi \equiv (\vartheta, \varphi)$. The main challenge is the fact that we project on a general possibly misspecified density. Explicitly acknowledging for estimating the parameters of the density yields some useful insight to the behaviour of the estimator.

3. Large Sample Theory

This section illustrates asymptotic results, that is consistency and asymptotic distribution for ψ. The properties of the estimator, as expected, depend crucially on the distance between the prior and the true population conditional density. We provide an explicit

[3]More precisely, what is obtained is the density conditional on $Z = z$.

shrinkage formulation when the distance vanishes at a $N_0^{\frac{1}{2}}$ rate and we comment on the case of employing non-parametric estimators[4].

Before stating the main results, we make certain assumptions that are fairly standard in parametric extremum estimation and are necessary and sufficient for the Propositions to be valid. For a stationary ergodic sequence $\{X_i, Z_i\}_{i=1,n\geq 1}^{N_0}$, we assume the following:

ASSUMPTIONS I.

(1) **(COMP)** $\Theta \subset \mathbb{R}^k, \Phi \subset \mathbb{R}^l$ are compact. Therefore $\Psi \equiv \Theta \times \Phi \subset \mathbb{R}^{k+l}$ is compact.

(2) **(ID)** $\exists! \psi_0 \in int(\Psi) : \psi_0 = \arg\max_{\Psi} \mathbb{E} \log h(x|z, \psi_0)$

(3) **(BD-1**a**)** $\forall l \in 1..M$ and for $d \leq 4, P \in \{F_\varphi, \mathbb{P}\}$:

$\mathbb{E}_{P|z} \sup_\psi ||m_l(x, \vartheta)||^d, \mathbb{E}_{P|z} \sup_\psi ||m_{l,\vartheta}(x, \vartheta)||^d$, and $\mathbb{E}_{P|z} \sup_\psi ||m_{l,\vartheta\vartheta}(x, \vartheta)||^d$ are finite, P_z-a.s.

(4) **(BD-1**b**)** $\sup_\psi \mathbb{E}_{\mathbb{P}(.|z)} ||e^{\mu_i' |m(x,z,\vartheta)|}||^{2+\delta} < \infty$ for $\delta > 0$, $\forall \mu(z) > 0, \mathbb{P}(.|z) - a.s$ [5]

(5) **(BD-2)** $\sup_\psi \mathbb{E}(\log h(x|z, \psi))^{2+\tilde{\delta}} < \infty$ where $\tilde{\delta} > 0$.

(6) **(PD-1)** For any non zero vector ξ and closed $\mathcal{B}_\delta(\psi)$, $\delta > 0$, and $P \in (F_\varphi, \mathbb{P})$, $\inf_{\xi \times \mathcal{B}_\delta(\psi)} \xi' \mathbb{E}_P \mathbf{m}(x, \vartheta) \mathbf{m}(x, \vartheta)' \xi > 0$ and $\sup_{\xi \times \mathcal{B}_\delta(\psi)} \xi' \mathbb{E}_P \mathbf{m}(x, \vartheta) \mathbf{m}(x, \vartheta)' \xi < \infty$

Assumptions (1)-(2) correspond to typical compactness and identification assumptions found in Newey and McFadden (1994) while (3) assumes uniform boundedness of conditional moments, up to a set of measure zero. Assumption (4) assumes existence of exponential absolute $1 + \delta$ moments and (5) boundedness of the population objective function[6]. Finally, (6) assumes away pathological cases of perfect correlation between moment conditions.

Note that the assumptions above correspond to the case of estimation of a density with finite dimensional parameters φ. In case φ is infinite dimensional, the conditions have

[4]Conditional density projections can therefore rationalize regularized versions of "optimal" GMM, see for example Hausman, Lewis, Menzel, and Newey (2011) for the case of the Continuous Updating Estimator (CUE).

[5] Note that **BD-1a** and **BD-1b** imply that $\sup_\psi \mathbb{E}_{\mathbb{P}_{z_i}} ||e^{\mu_i' m(x,z_i,\vartheta) + \lambda(z_i,\vartheta)} m(x, z_i, \vartheta_0)||^{2+\delta} < \infty$ for $d - 2 > \delta > 0$ and $\forall z_i$.

[6]The additional subtlety here is that it has to hold for the base measure and the true measure. Given absolute continuity of $\mathbb{P}(X|Z)$ with respect to $f(X|Z)$, the existence of moments under $\mathbb{P}(X|Z)$ is sufficient for the existence of moments under $f(X|Z)$

to be sufficiently generalized. Such a generalization involves additional conditions that control for parametric or semi-non parametric estimators for $f(x|z)$. In the former class of estimators we would need to define a function $\mathfrak{S}(x,z)$ that essentially replaces the usual score function in the finite dimensional case and corresponding stochastic equicontinuity and mean square differentiability conditions, see again Newey and McFadden (1994). In the semi-non parametric case, since the estimation space becomes a function of the sample size, i.e. $\Phi_n \subseteq \Phi_{n+1}... \subset \Phi$, conditions on the uniform convergence and continuity of the objective function have to be suitably adjusted, see for example Chen (2007).

Although we abstract from the above generalizations, the characterization of the asymptotic distribution using the high level assumption of asymptotically correctly specified $f(X|Z)$ is sufficient to illustrate the main trade-off arising when a practitioner wants to do inference using an estimated probability model without solving for the equilibrium law of motion.

Recall that we maximize the empirical analogue to (2.2), which, abstracting from simulation error, is equivalent to the following:

$$\max_{(\theta,\varphi)\in\Theta\times\Phi} Q_n(\theta,\varphi) \equiv \tfrac{1}{N_0}\sum_{i=1..N_0}\log\left(f(x_i|z_i,\varphi)\exp(\mu_i' m(x_i,z_i,\vartheta)+\lambda_i\right)$$

where

$$\begin{aligned}\forall i=1..n,\quad \mu_i: &\quad \int f(X|z_i,\varphi)\exp(\mu_i' m(X,z_i,\vartheta)m(X,z_i,\vartheta)dX=0\\ \lambda_i: &\quad \int f(X|z_i,\varphi)\exp(\mu_i' m(X,z_i,\vartheta)dX=1\end{aligned}$$

where for notational brevity we substituted $Z=z_i$ for z_i. Comparing our objective function with that of Kitamura, Tripathi, and Ahn (2004), apart from using exponential tilting in the "first step", we also do not smooth using local values for the instrument Z. Accounting for local smoothing would complicate in un unnecessary way the analysis without apparent gain. Most importantly, as the relevant applications are in macroeconomics, instruments will be lagged values of X_t, whose distribution is already pinned down by $f(.)$. In other non-time series applications, Z is treated as fixed.

The corresponding first order conditions of the estimator are going to be useful in order to understand both the asymptotic but also the finite sample results. Denoting by $\mathbf{M}$ the Jacobian of the moment conditions, the first order conditions are the following:

$$\begin{aligned} \vartheta: \quad & \tfrac{1}{N}\sum_i \left(\mu(z_i)'\mathbf{M}(x_i,z_i,\vartheta)+\mu_\theta(z_i)'\mathbf{m}(x_i,z_i,\vartheta)+\lambda_\vartheta(z_i)\right) &= 0 \\ \varphi: \quad & \tfrac{1}{n}\sum_i \left(\mathfrak{s}(x_i,z_i,\varphi)+\mu_\varphi(z_i)'\mathbf{m}(x_i,z_i,\vartheta)+\lambda_\varphi(z_i)\right) &= 0 \end{aligned}$$

where:

$$\begin{aligned} \mu(z_i) &= \arg\min_{\mu\in\mathbb{R}^k} \int f(X|z_i,\varphi)\exp(\mu' m(X,z_i,\vartheta)dX \\ \lambda(z_i) &= 1-\log\left(\int f(X|z_i,\varphi)\exp(\mu(z_i)' m(X,z_i,\vartheta)dX\right) \end{aligned}$$

With regard to the existence of $\mu(Z)$, or equivalently, the existence of the conditional density projection, Komunjer and Ragusa (2016) provide primitive conditions for the case of projecting using a divergence that belongs to the $\phi-$ divergence class and moment restrictions that have unbounded moment functions. Assumptions **BD-1a** and **BD-1b** are sufficient for their primitive conditions (Theorem 3).

In Appendix B we provide expressions for the first and second order derivatives of $(\mu(Z_i),\lambda(M_i))$ which determine the behaviour of $\hat{\psi}$ in the neighborhood of $\psi_0^\star$. More interestingly, these expressions will be useful for the characterization of the properties of our estimator in the case that the total variation distance between the prior density and the true density is not zero. In particular, the shrinkage direction will be towards the approximate model.

We first outline certain Lemmata which are systematically applied in the proofs of all propositions, and they are also useful in understanding the the source of the differences to traditional GEL estimation, apart from using exponential tilting in the "first step". We delegate the proofs to the auxiliary Lemmata to the supplemental material.

Lemma 3.1. *For any $\mathcal{Z}$-measurable function $g(\mu)$, $\mathbb{E}_z g(\hat{\mu}_i) \to \mathbb{E}_z g(\mu_i)$ and consequently, $\mathbb{E}_z \hat{\lambda}_i \to \mathbb{E}_z \lambda$.*

Proof. See Appendix B □

Lemma 3.2. *$\mu_i = O_p(TV(f_{N_0}, p_{N_0}))$. Furthermore,*

$$\forall i \in \{1..n_z\}, \max_i \sup_\vartheta |\mu_i' m(\vartheta, z_i)| = O_p(TV(f_{N_0}, p_{N_0}) N_0^{\frac{1}{d}})$$

Proof. See Appendix B □

A specific case of the above result is that of Newey and Smith (2004), where the total variation distance between the two densities is $TV(f_N, p_N) = O_p(N_0^{-\xi})$ and therefore $\mu_i = o_p(1)$ and if $\frac{1}{d} < \xi < \frac{1}{2}$, $\max_i \sup_\vartheta |\mu_i' m(\vartheta, z_i)| = o_p(1)$.

Corollary 3.2.1. *$\mu_i = O_p(\frac{1}{N_s} \sum_{j=1..N_s} m(x_j, z_i, \vartheta))$.*

Proof. See Appendix B □

Given the above results, we show consistency for both the case of misspecification and correct specification, and the asymptotic distribution under the latter case. We postpone the characterization of the asymptotic distribution under local misspecification for the next section.

3.1. **Consistency, Asymptotic Normality and Efficiency.** Due to the fact that the estimator involves a 'two step' procedure, where the first step involves using only simulated data, we need to make the assumption that the size of simulated data grows at a higher rate than sample size. The uniform consistency of the estimator is then shown by first proving pointwise consistency and then stochastic equicontinuity of the objective function. Details of the proof are in the Appendix.

Proposition 1. *Consistency for $\psi_0^\star$*
Under Assumption I, Lemmata 3.1-3.2:

$$(\hat{\vartheta}, \hat{\varphi}) \underset{p}{\to} (\vartheta_0^\star, \varphi_0^\star)$$

Proof. See the Appendix

□

As expected, under correct specification, consistency is for ϑ_0. This leads to the following corollary:

Corollary 3.2.2. *Consistency for* ϑ_0
If $f(X|Z,\hat{\varphi})$ *is consistent for* $\mathbb{P}(X|Z)$ *or correctly specified, then* $\vartheta_0^\star = \vartheta_0$.

Proof. See Appendix □

We also derive the limiting distribution of the estimator by the usual first order approximation around ψ_0. Below, we present the main result for a general, correctly specified density. Denoting by $G(\psi,.)$ the matrix of first order derivatives with respect to (ϑ,φ), the asymptotic distribution is regular.

Proposition 2. *Asymptotic Normality*
Under asymptotic correct specification, Assumption I, Lemmata 3.1-3.2, and for N_s,$N_0 \to \infty$ *such that* $\frac{N_0^{\bar{\gamma}+1}}{N_s} \to c$ *with* $c > 0$ *and* $\bar{\gamma} > 1 + \frac{2}{d}$*:*

$$N_0^{\frac{1}{2}}(\psi - \psi^*) \underset{d}{\to} N(0, \Omega^{-1})$$

where $\Omega = \mathbb{E}(G(.,z)'\mathbb{V}_g(.,z)^{-1}G(.,z))$.

Proof. See the Appendix □

The condition on $\bar{\gamma}$ states that the higher d is, i.e. the stronger the boundedness requirements on the moment conditions, the closer $\bar{\gamma}$ is to one. Conversely, when moment conditions cannot be guaranteed to be bounded for higher orders, simulation size has to grow at a higher rate.

In the supplemental material we derive the exact form of the variance covariance matrix of the estimator. Given a finite number of conditional moment restrictions and the

specified density, the efficiency attained is higher than the efficiency bound that does not use any additional information, as in Chamberlain (1987). To show this, we analyze the corresponding Jacobian terms and the variance covariance matrix of the first order conditions. For brevity, we denote dependence on data by the subscript i.

With regard to the Jacobian,

$$G(\psi_0) \equiv \begin{pmatrix} \bar{G}_{i,\vartheta\vartheta'}(\tilde{\psi}) & \bar{G}_{i,\vartheta\varphi'}(\tilde{\psi}) \\ \bar{G}_{i,\varphi\vartheta'}(\tilde{\psi}) & \bar{G}_{i,\varphi\varphi'}(\tilde{\psi}) \end{pmatrix}$$

for $M_i(\vartheta) \equiv \mathbb{E}(M(x,\vartheta)|Z)$, $\mathfrak{s}_i \equiv \mathbb{E}(\mathfrak{s}(X,\varphi)|Z)$ and $\mathfrak{B}_i$ the population projection coefficient from projecting the score on the user specified moment conditions, the corresponding components are as follows:

$$\mathbb{E}G_{i,\vartheta_l\vartheta'} = \mathbb{E}M_i(\vartheta)'V_m^{-1}(\vartheta)M_i(\vartheta) \tag{3.1}$$

$$\mathbb{E}G_{i,\vartheta\varphi'} = \mathbb{E}_z M_i(\vartheta)V_m^{-1}\mathbb{E}(m_i(\vartheta)\otimes\mathfrak{s}_i(\varphi)'|Z) \tag{3.2}$$

$$= \mathbb{E}_z M_i'(\vartheta)\mathfrak{B}_i(\psi) \tag{3.3}$$

$$\mathbb{E}G_{i,\varphi\varphi'} = \mathbb{E}_z\mathfrak{s}_i(\varphi)\mathfrak{s}_i(\varphi)' \tag{3.4}$$

Notice that the upper left component is the same as the information matrix corresponding to ϑ when the conventional optimally weighted GMM criterion is employed. The cross derivative involves the coefficient of projection of the score of the density on the economic moment conditions. Moreover, 3.4 is the outer product of the score of the density.

With regard to the covariance matrix, $V_g(\psi,z)$, notice that due to stationarity assumptions, the form of the long run variance will be $V_g(\psi,z) \equiv V_{g,0}(\psi,z) + \sum_i^{N_0-1}(\Gamma_{g,i} + \Gamma_{g,i}')$. More particularly, for $\mathfrak{s}_i^P \equiv \mathbf{m}\mathfrak{B}_i$, the instantaneous variance-covariance matrix,

$$\bar{V}(\psi_0) \equiv \begin{pmatrix} \bar{V}_{11}(\tilde{\psi}) & \bar{V}_{12}(\tilde{\psi}) \\ \bar{V}_{21}(\tilde{\psi}) & \bar{V}_{22}(\tilde{\psi}) \end{pmatrix}$$

has the following components:

$$\begin{aligned}
\bar{V}_{11} &= \mathbb{E}_z M_i(\vartheta)' V_m^{-1} M_i(\vartheta) \\
\bar{V}_{22} &= \mathbb{E}_z(\mathfrak{s}_i(\varphi) + \mathfrak{s}_i^P(\varphi))(\mathfrak{s}_i(\varphi) + \mathfrak{s}_i^P(\varphi))' \\
\bar{V}_{12} &= 2\mathbb{E}_z M_i(\vartheta)' \mathfrak{B}_i(\psi)
\end{aligned}$$

Analogously, the components of the autocovariance terms, $\Gamma_{g,i} = \frac{1}{k}\sum_{k=i+1}^{N_0} \mathbb{E} g_k g_{k-i}$ are :

$$\begin{aligned}
\mathbb{E}(g_k g_{k-i}')_{11} &= \mathbb{E}_z M_k(\vartheta)' \mathbb{E}(m_k(\vartheta) m_{k-i}(\vartheta)') M_{k-i}(\vartheta) \\
\mathbb{E}(g_k g_{k-i}')_{22} &= \mathbb{E}_z(\mathfrak{s}_k(\varphi) + \mathfrak{s}_k^P(\varphi))(\mathfrak{s}_{k-i}(\varphi) + \mathfrak{s}_{k-i}^P(\varphi))' \\
\mathbb{E}(g_k g_{k-i}')_{12} &= 2\mathbb{E}_z M_k(\vartheta)' \mathfrak{B}_{k-i}(\psi)
\end{aligned}$$

Interestingly, the expressions above have an intuitive interpretation. If the moment conditions we use satisfy $m(X, Z, \vartheta) = \mathfrak{s}(X, Z, \varphi) + \mathcal{U}$ and $\mathbb{E}(\mathcal{U}|\mathfrak{s}) = 0$, then the the variance covariance matrix (in the special case of *iid* data) collapses to:

$$\bar{V}_0 = \begin{pmatrix} H'(V_\mathfrak{s} + V_\mathcal{U})^{-1} H & 2(H' + \frac{\partial \mathcal{U}}{\partial \phi}) \\ 2(H + \frac{\partial \mathcal{U}'}{\partial \phi}) & 3(V_\mathfrak{s} + V_\mathcal{U}) + H \end{pmatrix}$$

where $H \equiv \mathbb{E}\frac{\partial^2}{\partial \varphi \varphi'} \log f(X, Z, \varphi)$. Under correct specification of the density, $H = V_\mathfrak{s}$ and therefore

$$\bar{V}_0 = \begin{pmatrix} V_\mathfrak{s}'(V_\mathfrak{s} + V_\mathcal{U})^{-1} V_\mathfrak{s} & V_\mathfrak{s}' + \frac{\partial \mathcal{U}}{\partial \phi} \\ V_\mathfrak{s} + \frac{\partial \mathcal{U}'}{\partial \phi} & 3(V_\mathfrak{s} + V_\mathcal{U}) + V_s \end{pmatrix}$$

In addition, if the moment conditions used span the same space spanned by the scores of the density, and this is the case when the model is solved, then $(G'\bar{V}_0 G)^{-1}$ trivially [7] attains the Cramer - Rao bound as $\mathcal{U} = 0$.

In general, letting $J \equiv M'V_{m0}M$, $\mathcal{W} \equiv ((\mathfrak{s} + \mathfrak{s}^p)(\mathfrak{s} + \mathfrak{s}^p)' - 4\mathfrak{B}'MJ^{-1}M'\mathfrak{B})$ and $\mathcal{Q} \equiv \mathfrak{s}\mathfrak{s}' - 2\mathfrak{B}'MJ^{-1}M'\mathfrak{B}$, the inverse of the variance covariance matrix of the estimator $G'\bar{V}_0 G$ will have the following form:

$$\Omega = \begin{pmatrix} J + M'\mathfrak{B}\mathcal{W}^{-1}\mathfrak{B}'M & M'\mathfrak{B}(I_{n_\vartheta \times n_\varphi} - \mathcal{W}^{-1}\mathcal{Q}) \\ \star & \mathfrak{B}'MJ^{-1}M'\mathfrak{B} + \mathcal{Q}\mathcal{W}^{-1}\mathcal{Q}' \end{pmatrix}$$

As is also known from the properties of GEL estimators, the projection in the simulated first step ensures that the moment conditions are automatically weighted with the variance covariance matrix to achieve maximum efficiency. What is more in our case is that additional moment conditions are generated by optimizing with respect to φ. Since these conditions also have information about ϑ, the optimal weighting makes use of it. By standard arguments, if we just used a trivial inverse $\bar{V}_0$ which was non zero only on the upper left block, i.e. $\bar{V}_{0,11} = (M'V_m^{-1}M)^{-1}$, the variance of the estimator would not be at its minimum level. If no information is used for the density, as in the GEL literature, where a non-parametric estimator for $f(X|Z)$ is employed, then $[\Omega^{-1}]_{11} = J^{-1}$, the semi-parametric lower bound[8].

In the next section, we show that in the case of misspecification of a parametric density, the first order conditions of the estimator can be conveniently rewritten such that they are equivalent to optimal GMM type of first order conditions plus a penalty term, which will be a function of the discrepancy between $f(X|\phi, Z)$ and $p(X|Z)$. Under local misspecification, this penalty has only second order effects. Moreover, misspecification in the form

[7] If we let $\mathcal{U} = 0$ then the covariance matrix becomes singular as both m and $\mathfrak{s}$ give the same information. Moreover, the first order conditions and G collapse to the standard score function and the Hessian (outer score product) respectively.

[8] This finding is also in line with the results of Imbens, Spady, and Johnson (1998) in the context of testing unconditional moment restrictions, who find that exponential tilting utilizes "efficient" estimates of probabilities rather than the inefficient $\frac{1}{N}$ weight used in the empirical likelihood literature. Nevertheless, efficiency gains in our case are of first order importance.

of wrong parametric restrictions can result in a bias - variance trade-off for ϑ. This also provides a shrinkage characterization of the estimator, where shrinkage on the nuisance parameters translates to efficiency gains in the estimates of structural parameters.

4. Shrinkage Towards the Statistical Model

4.1. **Finite dimensional φ.** In this section we investigate the consequences of density misspecification. We treat the unknown structural model as the infeasible case, so any misspecified density will imply certain restrictions on the density of the true structural model. We focus on misspecification of the type $R(\varphi) = 0$, where R is possibly non linear. This is quite general, as it represents not only non-linear restrictions on the space of parameters indexing a single density $f(X|Z, \varphi)$ but also restrictions on the mixture weights in finite mixtures of densities.

We first establish a few facts on the (lack of) first order effects of local misspecification of the density. Recall that the first order conditions of the estimator for ϑ once we substitute for the expressions for $\lambda(Z)$ and $\mu(Z)$ are the following:

$$(M_{\mathbb{P}} - M_H)' V_{m,\kappa,f}^{-1} m_f + M_f' V_{f,m}^{-1} m_{\mathbb{P}} \quad = \quad 0$$

where for notational simplicity we let $m_P \equiv \int m(X, Z) dP(X, Z)$ for any measure P.

Since $M_P - M_H \equiv \int M(x, \vartheta)(dP(x, z) - dH(x, z))$ the latter quantity collapses to zero for almost all (x, z) if and only if the base statistical model is correctly specified for the true data generating process. In this case the population first order conditions become the same as the Continuously Updating GMM estimator (CU) that is:

$$M_{\mathbb{P}}' V_{\mathbb{P},m}^{-1} m_{\mathbb{P}} \quad = \quad 0$$

In case of misspecification, rearranging terms in the above first order condition, the scaled by $N_0^{\frac{1}{2}}$ conditions are as follows:

$$0 = (M_{\mathbb{P}_n} - M_{H_n})' V_{\kappa,f_n}^{-1} N_0^{\frac{1}{2}} (m_{f_n} - m_{\mathbb{P},n}) + (M_{\mathbb{P}_n} - M_{H_n})' V_{\kappa,f_n}^{-1} N_0^{\frac{1}{2}} m_{\mathbb{P}_n} + ... \tag{4.1}$$

$$... + (M'_{f_n} V_{f_n}^{-1} - M'_{\mathbb{P}_n} V_{\mathbb{P}_n}^{-1}) N_0^{\frac{1}{2}} m_{\mathbb{P}_n} + M'_{\mathbb{P}_n} V_{\mathbb{P}_n}^{-1} N_0^{\frac{1}{2}} m_{\mathbb{P}_n} \tag{4.2}$$

The first three terms are functions of the distance between the proposed and the true $f(x|z)$. We utilize the fact that we can derive the rate of convergence of the terms involving functionals of the true and the locally misspecified density. More particularly, we provide below a decomposition that will be useful when thinking about the effects of discrepancies between the conditional density used by the econometrician and the true density. This decomposition will be trivial in the case of smooth parametric models.

Lemma 4.1. *Influence function for plug-in estimator Wasserman (2006)*
For a general function $W(x,z)$, conditional density $Q(x|z)$ and $\mathcal{L}(x,z) \equiv W(x,z) - \int W(x,z) d\mathbb{P}_z(x|z)$

$$\begin{aligned} W_{Q_n} - W_P &\equiv \int W(x,z) d(Q(x|z)\mathbb{P}(z)) - \int W(x,z) d(\mathbb{P}(x|z)\mathbb{P}(z)) \\ &= \int \int \mathcal{L}(x,z) dQ(x|z)\mathbb{P}(z) \end{aligned}$$

We use Lemma 4.1 to characterize the conditions under which local discrepancies between the conditional density used by the econometrician and the true density have an effect on the estimating equations characterizing ϑ. We first present the case that corresponds to the class of densities considered in this paper, that is the parametric class.

Proposition 3. *Parametric Smooth Density.*
For any (x,z) - measurable function $W(.)$ and $P \equiv P(\varphi)$, $\mathbb{P}(\varphi)$ 1-differentiable in ϕ, the

following statement holds:

$$W_{P(\phi_0+hN_0^{-\frac{1}{2}})} - W_P \quad = \quad N_0^{-\frac{1}{2}} h \int \delta_W(z) d\mathbb{P}(z)$$

Proof. See Appendix B □

The distance between any functional will therefore have the same order as that of the distance between the conditional densities. The first three terms in 3.12-3.13 involve functionals of the moment functions and their corresponding Jacobian matrices. Given Proposition 1, we can now determine whether the first order estimating equations for ϑ are affected by the misspecification. What we find is that local misspecification has first order effects on $\hat{\vartheta}$ *only through* $\hat{\phi}$.

Proposition 4. *Indirect first order effects*
Given Proposition 1, the system of equations in (4.1) becomes as follows:

$$0 \quad = \quad O_p(hN_0^{-\frac{1}{2}}) + M'_{P_n} V_{P_n}^{-1} N_0^{\frac{1}{2}} m_{P_n}$$

Proof. See Appendix B □

Note that the misspecification considered is arbitrary as h is arbitrary. Given this result, we can focus on shrinkage properties for ϑ arising solely because of shrinkage in ϕ. We analyze shrinkage by adopting the local asymptotic experiment approach, see for example Hansen (2016). We investigate convergence in distribution along sequences ψ_n where $\psi_n = \psi_0 + hN_0^{-\frac{1}{2}}$ for ψ_n the true value, $\psi_0 \in \Psi_0$ the centering value and h the localizing parameter. The true parameter is therefore "close" to the restricted parameter space up to h.

Proposition 5. *Asymptotic Distribution with Local Restrictions*

For $R(\varphi) \equiv \frac{\partial}{\partial \varphi} r(\varphi)$, $G^{-1} \equiv \begin{pmatrix} G^{11} & G^{12} \\ G^{21} & G^{22} \end{pmatrix}$, $S_1 \equiv [I_{n_1}, 0_{n_1 \times n_2}]$, $S_2 \equiv [0_{n_2 \times n_1}, I_{n_2}]$,

Under assumptions I such that $N_0^{\frac{1}{2}} \hat{G}(\tilde{\psi})^{-1} g(\psi_n) \underset{d}{\to} \mathcal{Z} \sim N(0, \Omega)$*:*

(1) $N_0^{\frac{1}{2}}(\hat{\vartheta} - \vartheta_n) \underset{d}{\to} \mathcal{Z}_r$

$$\text{where } \mathcal{Z}_r \equiv S_1 \mathcal{Z} - G^{12}(\psi_0) R(\varphi_0)(R(\varphi_0)' G^{22}(\psi_0) R(\varphi_0))^{-1} R(\varphi_0)'(S_2(\mathcal{Z} + h))$$

(2) For any non zero vector ξ, $\xi'(\mathbb{V}(S_1\mathcal{Z}) - \mathbb{V}(\mathcal{Z}_r))\xi \geq 0$

Proof. See Appendix B □

There are two main implications of Proposition 4.2 for $\hat{\vartheta}$. First, for $h > 0$, the asymptotic distribution is non regular i.e. the distribution depends on h (see p. 115 in van der Vaart (1998)). Second, the variance of ϑ_n is lower than the conventional semi-parametric lower bound for regular estimators. For ϑ_n arbitrarily close to the restricted subspace of ϑ_0, efficiency increases. More importantly, this increase in efficiency is *not local* as the size of h is left unrestricted. Note that no statement has been made about the implications for MSE. Future work can possibly look at restrictions on the domain of h such that this estimator dominates.

4.2. **A note on the Non Parametric Case.** While in this paper we have not formally dealt with non or semi parametric estimation of the conditional density of the observations, we make a sketch of what can be expected in terms of the behaviour of the estimator. First, it is clear that the conventional Taylor expansion is not valid anymore in the case of infinite dimensional ϕ. We nevertheless can characterize the behaviour of the estimator using the influence function in the non parametric case.

When a non parametric estimator is used, then integrating with respect to $Q(x|Z)$ yields that:

$$W_{Q_n} - W_P = \sum_{i \leq N_0} \omega_i \mathcal{L}(x_i, z_i)$$

where ω_i are local weights that depend on the data and some tuning parameter i.e. bandwidth. Letting $\zeta_i \equiv \omega_i \mathcal{L}(x_i, z_i)$, we make two observations. First, $\mathbb{E}\zeta_i$ is in general not zero as is typical in non parametric estimation i.e. there is a bias which has the same order as the bandwidth. Second, the variance of ζ_i is also typically of order lower than N_0^{-1} and therefore the rate of convergence is typically lower than $N_0^{-\frac{1}{2}}$. From equations 3.14-3.14 we can see that as long as this rate of convergence is not as low as $N_0^{-\frac{1}{4}}$, the first order conditions for ϑ do not have asymptotic first order bias. Moreover, restrictions on the class of densities considered will in general reduce variance and potentially increase bias in the estimate of $f(X|Z)$. In order to investigate the effects on estimates of ϑ we need to compute the influence function for $\hat{f}(X|Z)$ which is beyond the scope of this paper. Intuitively, optimizing the choice of auxiliary parameters like the bandwidth in a way that minimizes mean squared error should also minimize the mean squared error for ϑ, at least in the case of having a rate of convergence faster than $N_0^{-\frac{1}{4}}$. If this is not true, then we should expect slower rates of convergence for ϑ.

Although we have characterized the implications for the estimation of ϑ conditional on the choice of the auxiliary conditional density, we have not yet discussed what would lead to a reasonable choice of density. We provide such a discussion below. Moreover, we provide some simulation evidence on the performance of this method and an application to a small scale equilibrium model with standard agent optimization restrictions.

5. Discussion and Simulation Evidence

5.1. **Discussion on Choice of $F(X|Z)$ and Asymptotic Bias.** An obvious way to avoid distributional misspecification asymptotically is that of non parametrically estimation of $F(X|Z)$, which this paper abstracts from . One of the reasons is that within the

class of General Equilibrium models, once the equilibrium conditions are determined, we know a lot about $F(X|Z)$, even before solving the expectational system.

Recall that what is often specified without economic theory in the background, is the probability distribution of the shocks. Then, the practitioner specifies which moment conditions should be satisfied by the model. For example, a well known specification for the production function is the Cobb Douglas form, that is $\log y_t = \log A_t + (1-\alpha)K_t + \alpha N_t$ where A_t is an efficiency factor. Conditional on K_t and N_t being observable, the law of motion of output is determined by the production function and the process of A_t. Had A_t had been observable too, then we could estimate its law of motion, $\hat{F}(A_t|z_{t-1})$. The next question is whether we should estimate the law of motion for y_t. If $F(A_t|z_{t-1})$ and the Cobb Douglas condition are well specified, then we do not need to estimate $\hat{F}(y_t|z_{t-1})$. Since the Cobb Douglas form of the production function, or any other condition, are derived from economic theory, then they should be correctly specified by assumption. This is in contrast with partial equilibrium models, like in Gallant and Tauchen (1989), where estimating the law of motion is more important as it is left unspecified by the theory posed. In the context of this paper, what is more useful is to look at the extent to which estimates can be biased when the base density is slightly misspecified, when it is in principle observed and estimable, but we have limited sample size. Below, we provide evidence of how severe the effects on MSE can be in a simple setting.

5.2. **Monte Carlo Experiments.** We conducted two Monte Carlo (MC) experiments and an estimation exercise of the stochastic growth model with simulated data. In this section we present the MC experiment for the consumption Euler equation; the rest of the exercises are in Appendix C .

Estimating the Consumption Euler equation. We investigate performance in terms of MSE of our estimator compared to CU-GMM in the case of locally and non-locally misspecified base densities. Similar to the analytical example we used in previous sections, the DGP is a Bivariate log-Normal VAR for the (demeaned) consumption and interest

rate :

$$\begin{pmatrix} \log \tilde{C}_{t+1} \\ \log \tilde{R}_{t+1} \end{pmatrix} \sim N \left(\begin{pmatrix} \rho_C & \rho_{CR} \\ \rho_{RC} & \rho_R, \end{pmatrix} \begin{pmatrix} \log \tilde{C}_t \\ \log \tilde{R}_t \end{pmatrix}, \begin{pmatrix} \sigma_C^2 & \sigma_{CR} \\ \sigma_{RC} & \sigma_R^2 \end{pmatrix} \right)$$

Moreover, assuming a quadratic utility for the representative agent, $U(C_t) := \alpha C_t - \gamma C_t^2$ and that $\beta R_{ss} = 1$ the Euler equation becomes as follows:

$$\mathbb{E}_t \left(\beta \frac{C_{t+1} R_{t+1}}{C_t} - 1 \right) \quad = \quad 0$$

For the DGE we impose that $\rho_C = \rho_{RC} = 0, \rho_R = 0.95, \rho_{CR} = 0.05, \beta = 0.75$ and $\Sigma = [0.05, 0.002; 0.002, 0.05]$. We plot below MSE comparisons for typical sample (and sub-sample) sizes for quarterly macroeconomic data sets i.e. $n = \{20...210\}$ for two experiments. In the first experiment, we compare the performance of the CU-GMM estimator to our estimator, both in the case of knowing the density and estimating σ_{CR}.

As evident, the performance of GMM is much worse than the other two cases, as we use the empirical distribution function rather than the correctly specified density. In Figure 5.2 we present the results of restricting σ_{CR} to zero: the efficiency gain does not overcome the resulting bias. However, as we increase the dimension of the estimated parameters, the MSE gains become noticeable. In fact, in Figure 5.3 we present the case when we estimate $(\rho_{RC}, \rho_R, \sigma_{CR})$ s.t. $\rho_{CR} = 1 - \rho_R$ compared to estimating just σ_{CR}) and imposing local misspecification $T^{-\frac{1}{2}h}$on (ρ_{RC}, ρ_R) for h=0.01 and $\rho_{CR} = 1 - \rho_R$. The bias - variance trade-off holds for a moderately sized samples, indicating that our estimator can be potentially useful for estimating models in small subsamples. Moreover, the dominance over optimal GMM (plotted in Figure 5.1)is clearly visible .

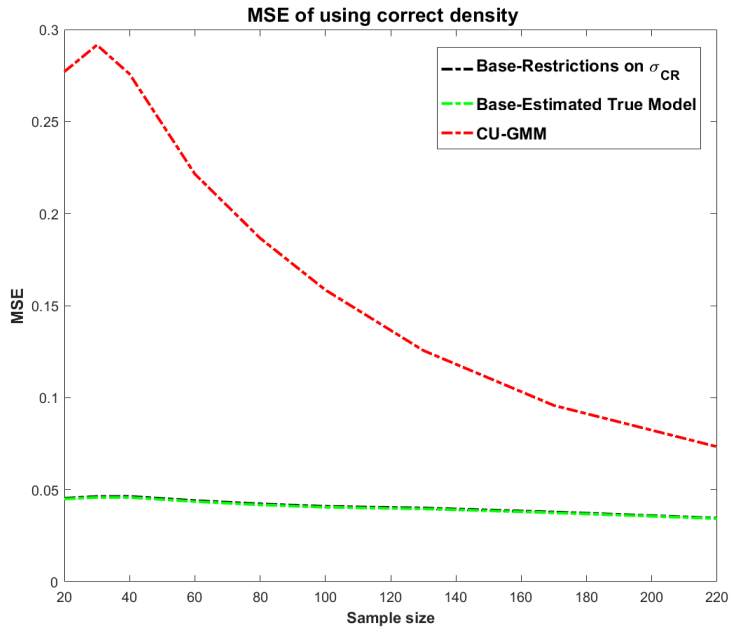

FIGURE 5.1. $\hat{\beta}$ for low dimensional Φ vs optGMM

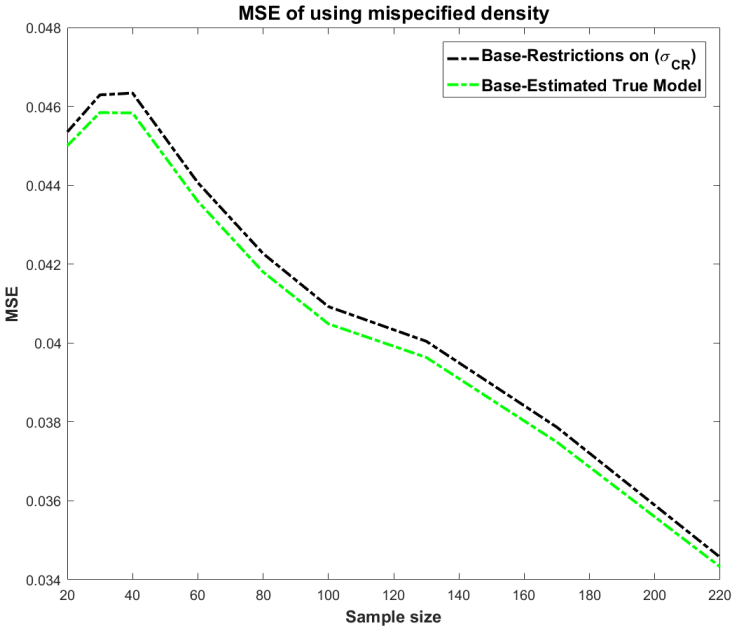

FIGURE 5.2. $\hat{\beta}$ for low dimensional Φ vs True model (1000 MC replications)

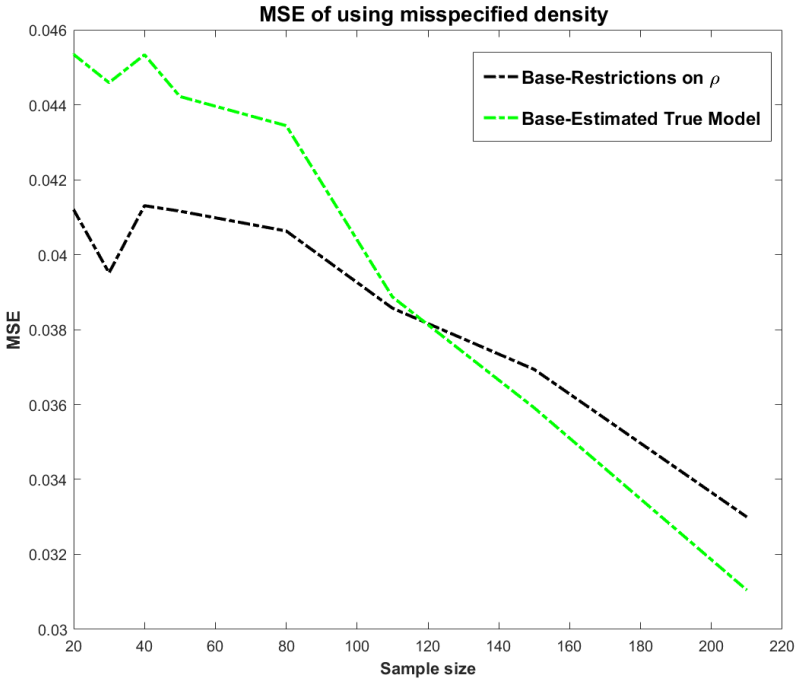

FIGURE 5.3. $\hat{\beta}$ for "high" dimensional Φ vs Restricted model (500 MC replications)

Interpreting GMM as a plug-in estimator using the empirical CDF, where the latter is the most basic infinite dimensional model for the true CDF, it is not surprising that a low dimensional but locally misspecified CDF performs better in terms of MSE in small samples.

6. Conclusion and Future Research

In this paper we have proposed an alternative approach to estimating a probability model that satisfies conditional moment restrictions coming from economic theory. The motivation comes from the fact that solving the equilibrium conditions for the decision rules requires assumptions that may not be valid and more importantly, are not revisable with the sample size. The use of auxiliary information on the predictive density of the observations to obtain a complete model enables one to construct estimated predictive distributions that can be used both for policy and forecasting exercises. We have shown

the asymptotic properties of this method under correct specification and local misspecification of the parametric conditional density of the observations. With regard to the latter, parametric models defined by drifting parameter sequences (that are local to the true parameter) can under some conditions lead to efficiency gains that can justify the use of auxiliary information even if this information is not accurate.

It is worthwhile to note that the results of this paper are general and therefore not confined to the case of equilibrium models, but any model defined by conditional moment restrictions. Using more information on the nuisance parameters leads naturally to efficiency gains. We have also shown simulation evidence for the performance of this method at various sample sizes under local and non-local misspecification and we applied the method to the prototypical stochastic growth model. Since this paper has focused on the econometric analysis and simulation of the method, we leave substantive applications for immediate future research.

Overall our method provides a promising way of estimating the parameters of models that are not probabilistically complete but nevertheless enables practitioners to do exercises that have been only possible with full information methods.

7. Appendix A

7.1. **Analytical derivations for Example 1.** Suppressing λ, the perturbation, $\exp(\mu' m(x,\vartheta)+\lambda)$ is proportional to

$$\exp\left(-\tfrac{1}{2}\left(\begin{pmatrix} c_{t+1}-\rho_c c_t \\ R_{t+1}-\rho_R R_t \end{pmatrix}' \begin{pmatrix} 0 & -\mu_t \\ -\mu_t & 0 \end{pmatrix} \begin{pmatrix} c_{t+1}-\rho_c c_t \\ R_{t+1}-\rho_R R_t \end{pmatrix}\right) - \mu_t \tfrac{c_t}{\beta}(1-R_t\rho_c\rho_R\beta)\right)$$

The trick here is that we can get the representation by rearranging terms, and drooping terms that do not depend on μ, and then do the minimization. Therefore, for $\begin{pmatrix} \epsilon_{1,t+1} \\ \epsilon_{2,t+1} \end{pmatrix} \equiv \begin{pmatrix} c_{t+1}-\rho_c c_t \\ R_{t+1}-\rho_R R_t \end{pmatrix}$ the problem becomes

$$\min_{\mu} \int \exp(-\frac{1}{2}\left(\begin{pmatrix} \epsilon_{1,t+1} \\ \epsilon_{2,t+1} \end{pmatrix}' \begin{pmatrix} 1 & -\mu_t \\ -\mu_t & 1 \end{pmatrix} \begin{pmatrix} \epsilon_{1,t+1} \\ \epsilon_{2,t+1} \end{pmatrix} + 2\mu_t \frac{c_t}{\beta}(1-R_t\rho_c\rho_R\beta)\right) d(R,C)$$

$$= \min_{\mu} \int \exp -\frac{1}{2}\left(\begin{pmatrix} \epsilon_{1,t+1} \\ \epsilon_{2,t+1} \end{pmatrix}' \begin{pmatrix} \frac{1}{(1-\mu_t^2)} & \frac{\mu_t}{(1-\mu_t^2)} \\ \frac{\mu}{(1-\mu_t^2)} & \frac{1}{(1-\mu_t^2)} \end{pmatrix}^{-1} \begin{pmatrix} \epsilon_{1,t+1} \\ \epsilon_{2,t+1} \end{pmatrix} + 2\mu_t \frac{c_t}{\beta}(1-R_t\rho_c\rho_R\beta)\right) d(R,C)$$

We therefore have that the F.O.C

$$\int \exp -\frac{1}{2}\left(\begin{pmatrix} \epsilon_{1,t+1} \\ \epsilon_{2,t+1} \end{pmatrix}' \begin{pmatrix} \frac{1}{(1-\mu_t^2)} & \frac{\mu_t}{(1-\mu_t^2)} \\ \frac{\mu}{(1-\mu_t^2)} & \frac{1}{(1-\mu_t^2)} \end{pmatrix}^{-1} \begin{pmatrix} \epsilon_{1,t+1} \\ \epsilon_{2,t+1} \end{pmatrix} - 2\mu_t \frac{c_t}{\beta}(1-R_t\rho_c\rho_R\beta)\right) \times ...$$

$$... \times (-(\epsilon_{1,t+1}\epsilon_{2,t+1} + \frac{c_t}{\beta}(1-R_t\rho_c\rho_R\beta))d(R,C) = 0$$

Then, for the Normal scaling constant C,

$$C\int N(\begin{pmatrix} \epsilon_{1,t+1} \\ \epsilon_{2,t+1} \end{pmatrix}, \begin{pmatrix} \frac{1}{(1-\mu_t^2)} & \frac{\mu_t}{(1-\mu_t^2)} \\ \frac{\mu}{(1-\mu_t^2)} & \frac{1}{(1-\mu_t^2)} \end{pmatrix})(\epsilon_{1,t+1})(\epsilon_{2,t+1}) - \frac{c_t}{\beta}(1-R_t\rho_c\rho_R\beta))d(R,C) = 0$$

$$\frac{\mu_t}{(1-\mu_t^2)} - \frac{c_t}{\beta}(1-R_t\rho_c\rho_R\beta) = 0$$

μ_t is the solution of the latter equation.

7.2. **Computational Considerations.** This section comments on the computational aspects of using information projections to estimate models defined by moment restrictions. This is important in terms of practice, and is indeed crucial when the number of moments conditions is higher. This makes it more costly to compute the projection with high precision. Moreover, in the case of conditional moment restrictions, the projection involves computing Lagrange multipliers which are both functions defined on $\Theta \times Z$. The dimension of this space can be formidably high.

To begin with, it is instructive to notice that the problem we are solving is a min-max problem, of a particular nature. In traditional empirical likelihood (GEL) computation, it is often advocated that the dual approach (min-max) can be computationally easier in the sense that it is lower dimensional. More particularly, in that case, if there are M constraints, N data points, and K parameters, then the dimension of the constrained optimization is $K+N$ with $M+1$ restrictions, while the min-max problem is of dimension $K+M$. Nevertheless, there is a potential cost to this dimension reduction, and this is the issue that the whole problem is not convex. While the inner loop (the one to obtain the multipliers) has a nice quadratic objective function, and can be handled with a typical Gauss-Newton procedure, the outer loop is often hard to handle.

In this paper, computation of the inner loop is much smoother than the one typically faced in the GEL. This is for the reason that the constraints are imposed on the population density, from which we can sample as much as we can. Furthermore, the issue of dimensionality reduction is more subtle as $\mu(z)$ and $\lambda(z)$ are still functions, and we therefore operate in an infinite dimensional space. The outer loop can nevertheless still be an issue. We use Markov Chain Monte Carlo (MCMC) as in Chernozhukov and Hong (2003) with a partially adaptive variance covariance matrix for the proposal distribution in the Metropolis - Hastings algorithm.

As already mentioned, μ is a vector of functions of the information set and the parameter vector. Therefore, in the estimation algorithm, the projection has to be implemented at all the points of z_i and at every proposal for the vector ϕ. In a high dimensional setting

due to large samples, instead of computing the projection it might be more efficient to estimate the unknown functions $\mu(X, Z)$ and $\lambda(X, Z)$ by simulating at different points of the support of the function and use function approximation methods i.e. splines. In case the model admits a Markov structure, the information set is substantially reduced, making computation much easier.

The general algorithm for the inner loop is therefore as follows:

(1) Given proposal for (φ, ϑ), simulate N_s observations from $F(x; z, \varphi)$

(2) For a finite set $\{z_1, z_2, ..z_k..z_K\}$ compute :

- $\mu(x_j; z_k\vartheta) = \arg\min \frac{1}{n_s} \sum_j \exp(\mu(x_j; z, \vartheta)' m(x_j; z_k, \vartheta))$ and
- $\lambda(x_j; z_k, \vartheta) = 1 - \log(\frac{1}{n_s} \sum_j \exp(\mu(x_j; z, \vartheta)' m(x_j; z_k, \vartheta)))$

(3) Evaluate log-likelihood: $L(x|z, \psi) = \frac{1}{N_0} \sum_i (\log h(x_i, z_i)\vartheta))$

Inner loop. In order to facilitate the quick convergence for the inner minimization and avoid indefinite solutions, we transform the objective function with a one to one mapping, and add a penalizing quadratic function, More particularly, let the objective function be $F(\mu) = \frac{1}{n_s} \sum_{j=..N_s} \exp(\sum_j m_j(x_i, \vartheta))$. The transformed objective function is $\tilde{F}(\mu) = \log(F(\mu) + 1) + \tau||\mu||^2$ where τ is the regularization parameter. We have tried many different examples, and in all the cases, with large enough simulation ($n_s = 5000$), the objective function has a nice quadratic form, something that makes the regularization trivial. Regularization becomes important when the simulation size is smaller, something that makes sense only if we want to reduce computational time. This introduces a bias to the value of μwhich is in the order of τ. The results reported are with $N_s = 5000$, as it has been checked that the objective function converges.

7.3. **Counterfactual Distributions.** An additional advantage of the method used in this paper, is that although the model is not solved for the equilibrium decision rules, we can still perform counterfactual experiments. What is more important is that this method readily gives a counterfactual distribution, while the distribution of the endogenous variables is hardly known in non-linear DSGE models. Knowing the distribution of

outcomes is extremely important for policy analysis, especially when non linear effects take place, and therefore the average effect is not a sufficient statistic to make a decision. Below I present an example which is based on a modification of Example 1, where the only difference is that the utility function is of the Constant Relative Risk Aversion form. The counterfactual experiment consists of increasing the CRRA coefficient ($\sigma = 1$ to $\sigma = 5$). Below I plot the contour maps of the conditional joint density of (R_{t+1}, c_{t+1}) with a change in the risk aversion coefficient. An increase in risk aversion is consistent with higher mean interest rate, and lower mean consumption. Moreover, consumption and interest rates are less negatively correlated. This is also consistent what the log - linearized Euler equation implies, $c_t = -\frac{1}{\sigma} r_t$.:

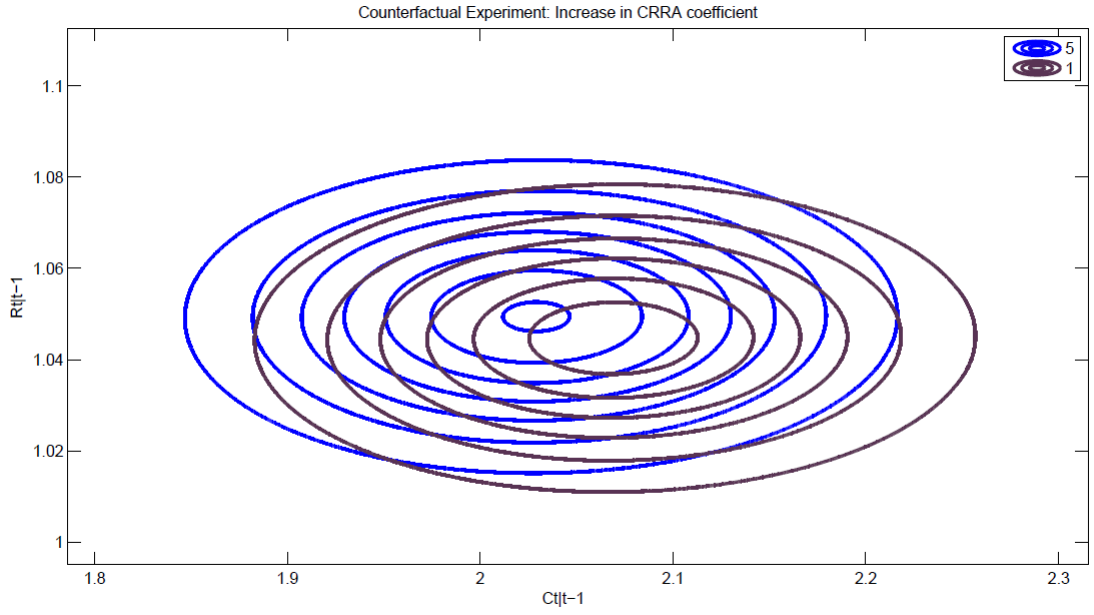

FIGURE 7.1. Increase in Risk Aversion Coefficient

7.4. **A Note On Non Differentiable models.** A not so uncommon case in economic theory in which first order conditions cannot be easily derived due to differentiability issues is the case of discrete choice. Discrete choice might be relevant in a macroeconomic framework in cases when some agents have to decide over finite actions, for example job search, default e.t.c. Discrete choice problems have a special structure, which we can also

make use of. This is the case because we can in principle obtain a *conditional choice probability*, (CCP), originally introduced by Hotz and Miller (1993).

Following Rust (1987), we can define the control variable sequence as $\{d_t, d_{t+1}, d_{t+2}...\}$. Moreover, let (x, ε) be the endogenous and exogenous state variables (with (x', ε') denoting next period), $v(x, d, \theta)$ the instantaneous return function and $p(.)$ the relevant probability density. By a standard Bellman formulation, controls solve the following functional equation,

$$V(x,\varepsilon) \quad = \quad \max_{d}\{v(x,d,\theta) + \beta \int_{(x',\varepsilon')} V((x',\varepsilon')p(x',\varepsilon'|x,\varepsilon,d,\vartheta)dx'd\varepsilon')\}$$

Under a conditional independence assumption, the Markov transition density is factorized as $p(x', \varepsilon'|x, \varepsilon, d, \vartheta) = p_2(\varepsilon'|x', \vartheta_2)p_3(x'|x, d, \vartheta_3)$, and taking expectations, $\mathbb{E}V(x) = \int_\varepsilon V(x, \varepsilon)p_2(\varepsilon|x, \vartheta_2)d\varepsilon$ and $\mathbb{E}V(x, d) = v(x, d, \theta_1) + \beta \int_{x'} \mathbb{E}V(x')p_3(x'|x, d, \vartheta_3)dx')$. Then, the log likelihood of a data point $\{X_i\}$ is as follows:

$$\log l_i(X_{d,i}, \vartheta) \quad = \quad \log P(d_i|x_i; \vartheta)p_3(x_i|x_{i-1}, d_{i-1}, \vartheta_3) \tag{7.1}$$

There are various (often tricky) ways to obtain the CPP $P(d_i|x_i; \vartheta)$ as a function of $v(.)$, ϑ and $\mathbb{E}V(x, d)$ in the microeconometrics literature, which can in principle be applied in the same way here, but I abstract from this and I encourage the interested reader to refer to the papers cited. Using 7.1, we can obtain the likelihood in the following way: Assuming that the data contain both continuous variables (Y_i), discrete variables $(X_{d,i})$, we can include the fixed point requirement $\mathbb{E}V = T(\mathbb{E}V, \vartheta)$ as another restriction i.e. $\mathbb{E}(V - T(\mathbb{E}V)) = Em_d(x_{d,i}, \vartheta_d)$[9]. This restriction will be satisfied for the true parameter vector parameterizing the discrete choice problem, and will also be a function of φ, i.e.

[9]In the case of correct specification of the underlying density, imposing fixed point conditions as restrictions is similar to the MPEC method of Su and Judd (2012).

$\vartheta^d = \tau(\varphi)$. More particularly, the tilted density will be of the form:

$$\begin{aligned} & h(x_{d,i}, y_i | z_i, \vartheta) \\ = \; & f(y_i | x_{d,i}, z_i, \varphi) \exp(\mu' m(y_i, \vartheta) \times p(x_{d,i} | y_i, z_i, \vartheta_d) \delta(x_d = x_{d,i}) \exp(\mu_d m_d(x_{d,i}, \vartheta_d)) \\ = \; & f((y_i, d_i | z_i, \tau(\varphi)) \exp(\mu' m(y_i, d_i, \vartheta)) \end{aligned}$$

where $\delta(.)$ is the Dirac delta function, used to represent the probability density function of a discrete variable. We avoid the non smoothness problem exactly by making the density a smooth function of ϑ and $\mathbb{E}V$. Furthermore, the asymptotic theory presented above, assumes differentiability of the moment function $m(., \vartheta)$ and this appears not to correspond to the class of discrete choice problems. Nevertheless, looking at the fixed point constraint, the "moment function" one can use is $\mathbb{E}_{p_2} V(x) - T(\mathbb{E}_{p_2} V(x), \vartheta)$ where $\mathbb{E}(V(x)) \equiv \int_{\varepsilon'} \max_{x_d} \mathbb{E}V(x, d) p_2(\varepsilon' | x, \vartheta_2) d\varepsilon')$. If T is smooth, we can see that assuming a continuous type of distribution for ε, overcomes the non differentiability of the " max " operator. In our formulation, the moment condition is $\mathbb{E}_x[\mathbb{E}_{p_2} V(x) - T(\mathbb{E}_{p_2} V(x), \vartheta)] = 0$ and therefore consistent with using the "smoothed" moment function.

8. Appendix B

Proof. **of Proposition 1** :

Convergence of μ_i. Consider the sets $\mathcal{V}_{\mu,\delta} = \{\mu \in \mathcal{M} : ||\mu - \mu_0|| < \delta\}$and $\mathcal{V}_{(\vartheta,\phi),\delta} = \{\vartheta \in \Theta : ||\vartheta - \vartheta_0|| < \delta, \phi \in \Phi : ||\phi - \phi_0|| < \delta\}$ and the objective functions they optimize respectively. By assumptions **BD-1** and the definition of $\mu = \arg\inf Q(x, z_i, \vartheta, \phi, \mu)$ $\underset{M\times 1}{\mu}(\phi, \vartheta)$ exists for all ϑ, ϕ and is unique . Fixing $Z = z_i, \forall\delta > 0$, we have that from a Taylor expansion of $Q(\mu, z_i) = \frac{1}{n_s}\sum_{1..n_s} e^{\mu_i' m_i(x_s,\vartheta)}$ with Lagrange Remainder:

$$\begin{aligned} Q(\mu_0, z_i) &\geq Q(\mu, z_i) = Q(\mu_0, z_i) + Q'_\mu(\mu_0, z_i)(\mu - \mu_0) + \frac{1}{2}Q''_\mu(\tilde{\mu}, z_i)(\mu - \mu_0)^2 \\ -\frac{1}{2}Q''_\mu(\tilde{\mu}, z_i)(\mu - \mu_0)^2 &\geq Q'_\mu(\mu_0, z_i)(\mu - \mu_0) \Rightarrow |Q'_\mu(\mu_0, z_i)| > C||\mu - \mu_0|| \end{aligned}$$

By assumption (**BD-1b**), the sequence $\{e^{\mu_i' m(x_s,\vartheta)} m(x_s, \vartheta_0, z_i)\}_{s=1..n_s}$ is uniformly integrable with respect to the $F-$measure , and by the WLLN for U.I sequences, we have that $Q'_\mu(\mu, z_i) = o_p(1)$ as:

$$\frac{1}{n_s}\sum_{1..n_s} e^{\mu_i' m(x_s,z_i,\vartheta)+\lambda} m(x_s, \vartheta_0) \xrightarrow{p} \mathbb{E}_{h|\varphi,z_i} m(x_s, \vartheta_0, z_i) = 0$$

Therefore, $\mu_i - \mu_{i,0} = o_p(1)$. $(a.s)$ We can actually improve on this rate, as by assumption (**BD-1a**)

$$\frac{1}{n_s}\sum_{1..n_s} e^{\mu_i' m(x_s,\vartheta)} m_i^2(x, \vartheta_0) \xrightarrow{u.p} \mathbb{E}_{h|\varphi,z_i} m_i^2(x_t, \vartheta)$$

and by the Central Limit Theorem for Martingale Difference sequences (**CLT-MDS**) Billingsley (1961), we have that

$$\frac{1}{n_s}\sum_{1..n_s} e^{\mu_i' m(x_t,\vartheta)} m(x_s, \vartheta_0) = O_p(n_s^{-\frac{1}{2}}).$$

Correspondingly, for $\psi = \arg\sup G(x, \psi, \mu)$ where

$$G(x, \psi, \mu) = \frac{1}{n}\sum_{i=1..n} \log\left(f(x_i|z_i, \varphi)\exp(\mu_i' m(x_i, z_i, \vartheta)\right)$$

Given the assumption that $\frac{n}{n_s} \to 0$ then $\forall(\phi, \vartheta), n$, $\hat{\mu}_i = \mu_i + o_p(1)$ and $G_n(\psi, \hat{\mu}_\psi) = G_n(\psi, \mu_\psi) + O_p(n_s^{-\frac{1}{2}})$, which follows from the differentiability of $G_n$in μ and the delta method.

8.1. **Uniform Convergence for Q_n.** Despite the fact that the pair $(\hat{\mu}, \hat{\lambda})$ is estimated at one-step, together with (φ, ϑ), the existence of a simulation step necessitates the use of general uniform convergence results. According to Theorem 1 in Andrews D.K 1992, we need to show (i) **BD** (Total Boundedness) of the metric space in which (φ, ϑ) lie together with (ii) **PC** (Pointwise consistency) and (iii) **SE** (Stochastic Equicontinuity).

Regarding (i). since in this section we are dealing with a finite dimensional φ, we rely on assumption 1 **(COMP)** which implies total boundedness. For pointwise convergence (ii),

$$
\begin{aligned}
& Pr(|\tfrac{1}{n}\textstyle\sum_i(\log(h(x_i; z_i, \psi)) - \mathbb{E}\log(h(x_i; z_i, \psi))| > \epsilon) \\
& \leq Pr(\tfrac{1}{n}\textstyle\sum_i |\log(h(x_i; z_i, \psi)) - \mathbb{E}\log(h(x_i; z_i, \psi)|)) > \epsilon) \\
\text{MarkovIn} \leq \quad & \tfrac{1}{n^2\epsilon}\mathbb{V}(\textstyle\sum_i |\log((h(x_i; z_i, \psi) - \mathbb{E}\log(h(x_i; z_i, \psi)|))
\end{aligned}
$$

This probability goes to zero as $\mathbb{E}\log(h(x_i; z_i, \psi)) < \infty$ and the autocovariances are summable by ergodicity.

Regarding (iii), Stochastic equicontinuity for the objective function can be verified by the "weak" Lipschitz condition in Andrews (1992) , as

$$
\begin{aligned}
& \limsup_{n\to\infty} Pr(\sup_\psi \sup_{\psi'} |\tfrac{1}{n}\textstyle\sum_i(\log h(x_i; z_i, \psi) - \log(h(x_i; z_i, \psi'))| > \epsilon) \\
& \leq \limsup_{n\to\infty} Pr(\sup_\psi \sup_{\psi'} |\tfrac{1}{n}\textstyle\sum_i(\log(1 + \tfrac{|h(x_i;z_i,\psi) - h(x_i;z_i,\psi')|}{h(x_i;z_i,\psi'))}) > \epsilon) \\
& \leq \limsup_{n\to\infty} Pr(\sup_\psi \sup_{\psi'} |\log(1 + \tfrac{1}{n}\textstyle\sum_i \tfrac{|h(x_i;z_i,\psi) - h(x_i;z_i,\psi'))|}{h(x_i;z_i,\psi'))}) > \epsilon) \\
\text{by monot.} & \leq \limsup_{n\to\infty} Pr(\sup_\psi \sup_{\psi'} |\tfrac{1}{n}\textstyle\sum_i(h(x_i; z_i, \psi)) - h(x_i; z_i, \psi'))| > \epsilon)
\end{aligned}
$$

Therefore the condition that needs to be shown is that,

$$
|\tilde{Q}_{n_x}(\psi, \hat{\mu}_\psi) - \tilde{Q}_{n_x}(\psi', \hat{\mu}_\psi)| \leq B_n \tilde{g}(d(\psi, \psi')), \forall(\psi, \psi') \in \Psi
$$

where $B_n = O_p(1)$ and $\tilde{g}$:$\lim_{y\to 0} \tilde{g}(y) = 0$. To verify this condition,

$$\begin{aligned}
|\tilde{Q}_n(\psi, \hat{\mu}_\psi) - \tilde{Q}_n(\psi', \hat{\mu}_{\psi'})| &= \frac{1}{n}|\sum_i \left(f_i(\varphi)\exp(\hat{\mu}'_{i,\psi} m_i(\vartheta) + \hat{\lambda}_{i,\psi} - f_i(\varphi')\exp(\hat{\mu}'_i m_i(\vartheta') + \hat{\lambda}_{i,\psi}) \right)| \\
&\leq \frac{1}{n}\sum_i |f_i(\varphi)\exp(\hat{\mu}_i(\psi)' m_i(\vartheta) + \hat{\lambda}_i(\psi) - f_i(\varphi')\exp(\hat{\mu}'_i m_i(\vartheta') + \hat{\lambda}_i(\psi))| \\
\text{sup sum<sum sup} \quad &\leq \frac{1}{n}\sum_i |\exp(\log f_i(\varphi) + \hat{\mu}_i(\psi)' m_i(\vartheta) + \hat{\lambda}_i(\psi)) - \exp(\log f_i(\varphi') + ... \\
&\quad \hat{\mu}'_i m_i(\vartheta') + \hat{\lambda}_i(\psi))|
\end{aligned}$$

Let $q_i(\psi) = \log f_i(\varphi) + \hat{\mu}_i(\psi)' m_i(\vartheta) + \hat{\lambda}_i(\psi)$. Therefore,

$$\begin{aligned}
|\tilde{Q}_n(\psi, \hat{\mu}(\psi)) - \tilde{Q}_n(\psi', \hat{\mu}(\psi))| &= \frac{1}{n}\sum_i |\exp(q_i(\psi)) - \exp(q_i(\psi'))| \\
&= \frac{1}{n}\sum_i |\exp(q_i(\tilde{\psi}))\nabla q_i(\tilde{\psi})(\psi) - \exp(q_i(\tilde{\psi}'))\nabla q_i(\tilde{\psi}')(\psi'))| \\
&\leq \frac{1}{n}\sum_i |\exp(q_i(\bar{\psi}))\nabla q_i(\bar{\psi})||\psi - \psi'| \\
&\leq \frac{1}{n}\sum_i |\exp(q_i(\bar{\psi}))\nabla q_i(\bar{\psi})||\psi - \psi'|
\end{aligned}$$

where $\bar{\psi} = \arg\max_{\{\tilde{\psi}, \tilde{\psi}'\}} |\exp(q_i(\bar{\psi}))\nabla q_i(\bar{\psi})|$

Let $B_{n_x} = \frac{1}{n}\sum_i |\exp(q_i(\bar{\psi}))\nabla q_i(\bar{\psi})|$. Notice that

$$\begin{aligned}
\mathbb{E}\frac{1}{n}\sum_i |\exp(q_i(\bar{\psi}))\nabla q_i(\bar{\psi})| &\leq \mathbb{E}_z\mathbb{E}_{x|z}\frac{1}{n}\sum_i |\exp(q_i(\bar{\psi}))\nabla q_i(\bar{\psi})| \\
&\leq \mathbb{E}|\exp(q_i(\bar{\psi}))||\nabla q_i(\bar{\psi})| \\
\text{C-S} \quad &\leq \left(\mathbb{E}(|\exp(q_i(\bar{\psi}))|)^2\right)^{\frac{1}{2}} \left(\mathbb{E}(|\nabla q_i(\bar{\psi}))|)^2\right)^{\frac{1}{2}} \\
\text{BD-1a,BD-2} \quad &= O_p(1)
\end{aligned}$$

Given the definition of the estimating equation i.e. the estimator of $\hat{\psi}$ is an extremum estimator, established weak uniform convergence, assumptions **ID**, **COMP**, and $\mathbf{BD-2}$(which guarantees continuity of the population objective), we have consistency by standard arguments (i.e. Newey and McFadden (1994) consistency results, Theorem 2.1)

□

Proof. **of Corollary 3.1** Consistency or correct specification of $f(X|Z,\varphi)$ imply that there exists a $\varphi_0 \in \varphi : f(X|Z,\varphi_0) = \mathbb{P}(X|Z)$. By Lemma 3.2, $\lambda(Z_i) = \mu(Z_i) = 0 \forall i$ and therefore $h(X|Z,\psi) = f(X|Z,\varphi)$. By construction, the moment condition holds under the H measure , $\mathbb{E}_H m(X,Z,\vartheta) = \int \mathbb{P}(X,Z) m(X,Z,\vartheta_0^\star) d(X,Z) = 0$. But it is true that $\int \mathbb{P}(X,Z) m(X,Z,\vartheta_0) d(X,Z) = 0$. Since $\theta_0^\star$ is identified, $\theta_0 = \theta_0^\star$. □

Proof. **of Proposition 2 (Asymptotic Normality):** Denoting by subscript i the function evaluation with real data, and by subscript j the evaluation with simulated data, the first order conditions characterizing the estimator are:

$$\begin{aligned} \vartheta: &\quad \tfrac{1}{n}\sum_i \left(\mu_i' M_i(\vartheta) + \mu_{\theta,i}' m_i(\vartheta) + \lambda_{i,\vartheta}\right) = 0 \\ \phi: &\quad \tfrac{1}{n}\sum_i \left(\tfrac{s_i(\varphi)}{f_i(\varphi)} + \mu_{\varphi,i}' m_i(\vartheta) + \lambda_{i,\phi}\right) = 0 \end{aligned}$$

Define :$e_{j,i} = e^{\mu_i' m_{j,i}(\vartheta)}$, $z_{j,i} = e_{j,i}(I_{n_\vartheta} + (\mu_i' M_j \otimes m_j)(M_j' M_j)^{-1} M_j)$,$\tilde{e}_{j,i} = \frac{e_{j,i}}{\frac{1}{n_s}\sum_{j=1..s} e_{j,i}}$, $\kappa_{j.i} = -\frac{(e^{\mu_i' m_{j,i}(\vartheta)} - 1)}{\mu_i m_{j,i}(\vartheta)'}$, $s_j := \frac{\partial}{\partial \phi} \log f(x|\phi,z)$ and $\mathfrak{s}_j := \frac{s_j}{f_j}$, $\tilde{e}_{j,\vartheta} = \tilde{e}_j(m_j' \mu_\vartheta + \mu'(M_j - \sum_j \tilde{e}_j M_j))$ and $\tilde{e}_{j,\phi} = -\tilde{e}_j \sum_j \tilde{e}_j \tilde{\mathfrak{s}}_j$. We have already established that as long as the base density is asymptotically correctly specified, then $\mu_i \underset{p}{\to} 0$ for almost all z_i. Therefore, $e_{j,i} \underset{p}{\to} 1$, $z_{j,i} \underset{p}{\to} 1$ and $\kappa_{j,i} \underset{p}{\to} -1$.

The derivatives of (μ, λ) with respect to ψ are as follows,

$$\begin{aligned} \mu_{\vartheta,i} &= (-\frac{1}{n_s}\sum_j e_j m_{j,i} m_{j,i}')^{-1} (\frac{1}{n_s}\sum_j z_{j,i} M_{j,i}) \\ \lambda_{\vartheta,i} &= -\mu_i' \sum_j \tilde{e}_j M_{j,i} - \sum_j \tilde{e}_j m_{j,i}' \mu_{\vartheta,i} \\ \mu_{\phi,i} &= (\sum_j e_j m_{j,i} m_{j,i}')^{-1} \sum_j e_j m_{j,i} \otimes \mathfrak{s}_{j,i} \\ \lambda_{\phi,i} &= -\sum_j \tilde{e}_j \mathfrak{s}_{j,i} - \sum_j \tilde{e}_j m_{j,i}' \mu_{\phi,i} \end{aligned}$$

Therefore, the estimator has this final implicit form $\underset{1\times(n_\vartheta+n_\varphi)}{G_n} = 0$ where

$$G_n = \left[\begin{array}{c} \frac{1}{n}\sum_i \underbrace{\left(\frac{1}{n_s}\sum_j m_{j,i}(\vartheta)'\right)}_{\hat{C}_{1,i}} \underbrace{\left(\frac{1}{n_s}\sum_j \kappa_{j,i} m_{j,i}(\vartheta) m_{j,i}(\vartheta)'\right)}_{\hat{A}_{1,i}}{}^{-1} \underbrace{\left(M_i(\vartheta) - \frac{1}{n_s}\sum_j \tilde{e}_{j,i} M_{j,i}(\vartheta)\right)}_{\hat{B}_{1,i}} \\ \frac{1}{n}\sum_i \underbrace{m_i(\vartheta)'}_{C_{2,i}} \underbrace{\left(\frac{1}{n_s}\sum_j e_{j,i} m_{j.i}(\vartheta) m_{j,i}(\vartheta)'\right)}_{\hat{A}_{2,i}}{}^{-1} \underbrace{\left(\frac{1}{n_s}\sum_j e_{j,i} m_{j,i}(\vartheta)\otimes \mathfrak{s}_{j,i}(\psi)\right)}_{\hat{B}_{2,i}} \end{array}\right] + ..$$

$$... + \left[\begin{array}{c} \frac{1}{n}\sum_i m_i(\vartheta)' \left(\frac{1}{n_s}\sum_j e_{j,i} m_{j.i}(\vartheta) m_{j,i}(\vartheta)'\right)^{-1} \underbrace{\left(\frac{1}{n_s}\sum_j z_{j,i} M_{j,i}(\vartheta)\right)}_{\hat{B}_{3,i}} - \underbrace{\sum_j \tilde{e}_j m_j' \mu_\vartheta}_{\hat{B}_{4,i}} \\ \frac{1}{n}\sum_i \underbrace{\left(\mathfrak{s}_i - \frac{1}{n_s}\sum_j \tilde{e}_{j,i}\mathfrak{s}_{j,i}\right)}_{\hat{B}_{5,i}} \end{array}\right]$$

We proceed by decomposing $N^{\frac{1}{2}} G_{i,n} = N^{\frac{1}{2}} G_{i,0} + N^{\frac{1}{2}}\Delta_i$ and show that $n^{\frac{1}{2}}\Delta_i = o_p(1)$. Since we have effectively two different samples to handle, which are conditionally independent (conditional on z_i), we have to further decompose in different factors and make use of the pointwise convergence for almost all Z. By *iid* sampling and domination assumptions $\mathbf{BD-1}$ and Lemma 3.1, $\hat{A}_1 \to A_1 \equiv \mathbb{V}_m(z)$, $\hat{A}_2 \to A_2 \equiv \mathbb{V}_m(z)$, $\hat{C}_1 \to C_1 \equiv \mathbb{E}_F(m(\vartheta)|Z) = m_F$, $\hat{B}_2 \to B_2 \equiv M(Z,\vartheta) - \mathbb{E}_H(M(\vartheta)|Z) \equiv M_H$, $\hat{B}_3 \xrightarrow{p} B_3 \equiv \mathbb{E}_F(M(\vartheta)|Z) \equiv M_F$, $\hat{B}_4 \xrightarrow{p} B_4 \equiv m_F'\mu_\vartheta$, $\hat{B}_5 \to B_5 \equiv \mathfrak{s}(\varphi,Z) - \mathbb{E}_H(\mathfrak{s}(\varphi)|Z) \equiv \mathfrak{s}(\varphi,Z) - \mathfrak{s}_H$.

To maximize clarity, we analyze $N^{\frac{1}{2}} G_n$ row-wise. With regard to $G_{1,n}$, we have that

$$\begin{aligned} N^{\frac{1}{2}} G_{1,n} &= N^{-\frac{1}{2}}\sum_i (C_{1,i}A_{1,i}^{-1}B_{1,i} - (C_{1,i}-\hat{C}_{1,i})\hat{A}_{1,i}^{-1}\hat{B}_{1,i} + ... \\ &\quad ... - C_{1,i}(A_{1,i}^{-1} - \hat{A}_{1,i}^{-1})\hat{B}_{1,i} + C_{1,i}A_{1,i}^{-1}(B_{1,i}-\hat{B}_{1,i}) + ... \\ &\quad ... - C_{2,i}(A_{2,i}^{-1} - \hat{A}_{2,i}^{-1})\hat{B}_{3,i} - C_{2,i}A_{2,i}^{-1}(B_{3,i}-\hat{B}_{3,i}) + C_{2,i}A_{2,i}^{-1}B_{3,i} - \hat{B}_{4,i} \\ &= N^{-\frac{1}{2}}\sum_i C_{1,i}A_{1,i}^{-1}B_{1,i} + N^{-\frac{1}{2}}\sum_i C_{2,i}A_{2,i}^{-1}B_{3,i} + O_p(N_s^{-\frac{1}{2}} N^{\frac{1}{d}+\frac{1}{2}}) \end{aligned}$$

$$
\begin{aligned}
N^{\frac{1}{2}} G_{2,n} &= N^{-\frac{1}{2}} \sum_i (C_{2,i} A_{2,i}^{-1} B_{2,i} - C_{2,i}(A_{2,i}^{-1} - \hat{A}_{2,i}^{-1})\hat{B}_{2,i} - C_{2,i} A_{2,i}^{-1}(B_{2,i} - \hat{B}_{2,i}) + \hat{B}_{5,i} \\
&= N^{-\frac{1}{2}} \sum_i C_{2,i} A_{2,i}^{-1} B_{2,i} + N^{-\frac{1}{2}} \sum_i B_{5,i} + O_p(N_s^{-\frac{1}{2}} N^{\frac{1}{d}+\frac{1}{2}})
\end{aligned}
$$

A sufficient condition for the remainder to be negligible is that $\bar{\gamma} > 1 + \frac{2}{d}$. To understand why these rates arise, we illustrate three cases, which are indicative for the treatment of all other terms, which are of lower order.

$$
\begin{aligned}
||\frac{1}{N}\sum_i C_{1,i}\left(\hat{A}_{i,1}^{-1} - A_{i,1}^{-1}\right)\hat{B}_{1,i}|| &\leq \max_i ||C_{1,i}|| \max_i ||\hat{A}_{i,1}^{-1} - A_{i,1}^{-1}|| \frac{1}{N}\sum_i ||\hat{B}_{1,i}|| \\
&= O_p(\kappa_N^{-1}) \times O_p(N_s^{-\frac{1}{2}}) \times O_p(1) = O_p(\kappa_N^{-1} N_s^{-1})
\end{aligned}
$$

$$
\begin{aligned}
||\frac{1}{N}\sum_i C_{2,i}\left(\hat{A}_{i,2}^{-1} - A_{i,2}^{-1}\right)\hat{B}_{3,i}|| &\leq \frac{1}{N}\sum_i ||C_{2,i}||||\hat{A}_{i,2}^{-1} - A_{i,2}^{-1}||||\hat{B}_{3,i}|| \\
&= \max_i \sup_\psi ||\hat{A}_{i,2}^{-1} - A_{i,2}^{-1}|| \frac{1}{N}\sum_i ||\hat{C}_{2,i}||||\hat{B}_{3,i}|| \\
&\leq \max_i \sup_\psi ||\hat{A}_{i,2}^{-1} - A_{i,2}^{-1}|| \max_i \sup_\psi ||C_{2,i}|| \frac{1}{N}\sum_i ||\hat{B}_{3,i}|| \\
&= O_p(N_s^{-\frac{1}{2}}) O_p(N^{\frac{1}{d}}) \times O_p(1) = O_p(N_s^{-\frac{1}{2}} N^{\frac{1}{d}})
\end{aligned}
$$

where κ_n is the rate at which C_1 approaches zero i.e. the rate at which $TV(F_i, P_i)$ converges to zero and d as in Assumption. (See proof of Lemma 3.2)

Denoting by $\hat{\Xi}_2$ the terms that do not vanish faster than $N^{-\frac{1}{2}}$,

$$
\hat{\boldsymbol{\Xi}}_{\mathbf{2}} = \begin{bmatrix} \frac{1}{N}\sum_i C_{1,i} A_{i,1}^{-1} B_{1,i} + \frac{1}{N}\sum_i C_{2,i} A_{i,2}^{-1} B_{3,i} \\ \frac{1}{N}\sum_i C_{2,i} A_{i,2}^{-1} B_{2,i} + \frac{1}{N}\sum_i \hat{B}_{5,i} \end{bmatrix}
$$

Under asymptotic correct specification, the second terms in each row are order $O_p(\kappa_n^{-1})$.

To show asymptotic normality, we make use of the Cramer-Wold device. Let ξ be a vector of real valued numbers, normalized such that $||\xi|| = 1$ then:

$$\begin{aligned} N^{\frac{1}{2}}\xi'_{(n_\vartheta+n_\varphi\times 1)}\Xi_2 &= N^{-\frac{1}{2}}\sum_i \xi_1' C_{1,i}A_{i,1}^{-1}B_{1,i} + +N^{-\frac{1}{2}}\sum_i \xi_2' C_{2,i}A_{i,2}^{-1}B_{2,i} + o_p(1) \\ &= \hat{\Xi}_{21} + \hat{\Xi}_{22} + o(1) \end{aligned}$$

where $\xi'_{p\times 1} = \left(\underset{\dim(\vartheta)}{\xi_1'} , \underset{\dim(\varphi)}{\xi_2'} \right)$.

What needs to be shown is that the variance of the above terms is finite. Then by the **CLT-MDS** we conclude. In the supplemental material we show that the variance of each of the terms is finite, and so is the variance of the sum by appealing to the C-S inequality.

Combining the above results we can see that:

$$\begin{aligned} n^{\frac{1}{2}}\xi'_{p\times 1}(G_n - \mathbb{E}G_{n,2}) &= n^{-\frac{1}{2}}\xi'_{p\times 1}(\Xi_{n,2} - \mathbb{E}\Xi_{n,2}) + o_p(1) \\ &\to N(0, \xi' V_g \xi) \end{aligned}$$

and therefore

$$n^{\frac{1}{2}}(G_n(\psi_0) - \mathbb{E}G_{n,2}(\psi_0)) \to N(0, V_g)$$

8.2. **Efficiency: Form of Ω.** : For brevity, we delegate this derivation to the Supplemental Material.

□

Proof. **of Proposition 3**

In the parametric case within the class of smooth densities, we can rewrite $dQ(x|z) \equiv dP(x|\phi + n^{-\frac{1}{2}}h, z)$. Therefore, using a Taylor expansion of around ϕ_0

$$dP(x|\phi + n^{-\frac{1}{2}}h, z) = dP(x|\phi, z) + s_\phi(x, z)n^{-\frac{1}{2}}h + o(n^{-\frac{1}{2}}h)$$

Evaluating $\int\int \mathcal{L}(x,z)dQ(x|z)\mathbb{P}(z)$ gives the result:

$$\begin{aligned} w_{Q_n} - w_P &\equiv \int w(x,z)(s_\phi(x,z)n^{-\frac{1}{2}}h + o(n^{-\frac{1}{2}}h))d\mathbb{P}(z) \\ &= n^{-\frac{1}{2}}h\int \delta_w(z)d\mathbb{P}(z) \end{aligned}$$

□

Proof. **of Proposition 8**

Substituting the result of Proposition 1 in 4.1 we get that:

$$\begin{aligned} 0 &= N_0^{-\frac{1}{2}}\delta_M' h' V^{-1} h\delta_m + n^{-\frac{1}{2}}h\delta_M' V^{-1} N_0^{\frac{1}{2}} m_{P_n} + ... \\ ... &+ (N_0^{-\frac{1}{2}}\delta_M' h' V^{-1} + M_P' N_0^{-\frac{1}{2}} h\delta_V) N_0^{\frac{1}{2}} m_{P_n} + M_{P_n}' V_{P_n}^{-1} N_0^{\frac{1}{2}} m_{P_n} + o_p\left(hN_0^{-\frac{1}{2}}\right) \\ 0 &= O_p(hN_0^{-\frac{1}{2}}) + M_{P_n}' V_{P_n}^{-1} N_0^{\frac{1}{2}} m_{P_n} \end{aligned}$$

Notice that for the Jacobian terms we also substituted $\mathbb{P}$ for $\mathbb{P}_n$ as the empirical distribution function converges also at the $N_0^{\frac{1}{2}}$ rate [10].

□

Proof. **of Proposition 9**

1) The first order conditions for ϕ under restrictions $r(\phi) = 0$ are as follows:

$$\hat{\phi} - \phi_n = -\hat{G}^{21}(\tilde{\psi})\hat{g}_1(\psi_n) - \hat{G}^{22}(\tilde{\psi})(\hat{g}_2(\psi_n) + \pi R(\phi_n))$$

For notational convenience we drop dependence on ψ. Expanding the constraint around ϕ_0 and substituting for $\hat{\phi} - \phi_n$,

$$\pi = -(R'G^{22}R)^{-1}R'(G^{21}g_1 + G^{22}g_2 + hN_0^{-\frac{1}{2}})$$

[10] This can also be verified by plugging $\mathbb{P}_n$ in Q_n in the decomposition in Lemma 4.1

Substituting for π in $\hat{\phi}-\phi_n$ and plugging in the first order conditions for $\vartheta-\vartheta_n$ the result follows.

2) We show positive definiteness of $\mathbb{V}(S_1\mathcal{Z}S_1')-\mathbb{V}(\mathcal{Z}_r)$ by showing that

$$tr((\mathbb{V}(S_1\mathcal{Z}))^{-1}(\mathbb{V}(\mathcal{Z}_r)))<n_1$$

Let $\tilde{S}_i=S_i\Omega^{\frac{1}{2}}$ for $i=1,2$, $\tilde{R}=[G^{22}]^{\frac{1}{2}}R$ and $J=G^{12}[G^{22}]^{-\frac{1}{2}}\tilde{R}(\tilde{R}'\tilde{R})^{-1}\tilde{R}'$. Recall that $\mathcal{Z}_r\equiv S_1\mathcal{Z}-JS_2(\mathcal{Z}+h)$. Positive definiteness of $\mathbb{V}(S_1\mathcal{Z})-\mathbb{V}(\mathcal{Z}_r)$ is equivalent to:

$$tr(\mathbb{V}(S_1\mathcal{Z})^{-1}\mathbb{V}(\mathcal{Z}_r)) \quad < \quad n_1 \tag{8.1}$$

where n_1 is the dimension of g_1. Absence of restrictions implies that $R=0$ and therefore $\mathcal{Z}_r=S_1\mathcal{Z}$. This implies that $tr(\mathbb{V}(S_1\mathcal{Z})^{-1}\mathbb{V}(\mathcal{Z}_r))=n_1$. What needs to be shown therefore is that the inequality in 8.1 holds for any $R\neq 0$. Towards this, we first rewrite the left hand side of 8.1 as follows:

$$\begin{aligned} tr(\mathbb{V}(S_1\mathcal{Z})^{-1}\mathbb{V}(\mathcal{Z}_r)) &= tr((S_1\Omega S_1')^{-1}(S_1-JS_2)\Omega(S_1-JS_2)') \\ &= tr((\tilde{S}_1\tilde{S}_1')^{-1}(\tilde{S}_1-J\tilde{S}_2)(\tilde{S}_1-J\tilde{S}_2)') \\ &= tr((\tilde{S}_1-J\tilde{S}_2)'(\tilde{S}_1\tilde{S}_1')^{-1}(\tilde{S}_1-J\tilde{S}_2)) \end{aligned}$$

For $V'\equiv\underset{n_1\times n}{\tilde{S}_1}\ \underset{n_1\times n_2}{J}$, $B\equiv\begin{pmatrix} \underset{n\times n}{I} & \underset{n\times n_2}{0} \\ \underset{n_2\times n}{0} & \underset{n_2\times n_2}{0}\end{pmatrix}$, $C\equiv\begin{pmatrix} \underset{n\times n}{I} & \underset{n\times n_2}{-\tilde{S}_2'}\end{pmatrix}'$

$A\equiv CC'=\begin{pmatrix} \underset{n\times n}{I} & \underset{n\times n_2}{-\tilde{S}_2'} \\ \underset{n_2\times n}{-\tilde{S}_2} & \underset{n_2\times n_2}{\Omega_{22}}\end{pmatrix}$ where $n=n_1+n_2$ and $\tilde{S}_2=\begin{pmatrix}[\Omega]_{12}^{\frac{1}{2}} & [\Omega]_{22}^{\frac{1}{2}}\end{pmatrix}$,

$$tr(\mathbb{V}(S_1\mathcal{Z})^{-1}\mathbb{V}(\mathcal{Z}_r)) = tr((V'(J)BV(J))^{-1}V(J)'AV(J))$$

We therefore need to show the following:

$$\max_{V} \quad tr((V'(J)BV(J))^{-1}V(J)'AV(J)) \quad = n_1 \tag{8.2}$$

The problem defined by the LHS is of 8.2 is a well defined problem in discriminant analysis for a *general matrix* V, and is equivalent to:

$$\begin{aligned} \max_{V} \quad & tr(V(J)'AV(J)) \\ \text{s.t} \quad & V'(J)BV(J) < K \end{aligned}$$

Using that A is symmetric, the first order conditions are:

$$AV(J) \quad = \quad BV(J)\Lambda \tag{8.3}$$

where Λ is the $n_1 \times n_1$ matrix that contains the lagrange multipliers for the second set of constraints. Noticing that:

$$\begin{aligned} tr((V'(J)BV(J))^{-1}V(J)'AV(J)) \quad &= \quad tr((V'(J)BV(J))^{-1}V(J)'AA^{-1}BV(J)\Lambda) \\ &= \quad tr(\Lambda) \end{aligned}$$

$$\max_{V} tr(V(J)'AV(J)) \quad = \quad \sum_{i \leq n_1} \lambda_i$$

Since the system of equations in 8.3 is a generalized eigenvalue problem, then in order for the maximum to be achieved, $\sum_{i \leq n_1} \lambda_i$ must be the sum of the $n_1 - th$ largest admissible eigenvalues of $B^{-1}V$ and V the matrix containing the corresponding eigenvectors. A complication arises here because B is non invertible, and we therefore cannot compute the eigenvalues of $B^{-1}A$ directly. We proceed as follows: We compute the eigenvalues μ_i of $A^{-1}B$ and use the fact that $\lambda_i = \mu_i^{-1}$.

$$
\begin{aligned}
A^{-1}B &= \begin{pmatrix} \underset{n\times n}{\Xi} & \underset{n\times n_2}{0} \\ \underset{n_2\times n}{0} & \underset{n_2\times n_2}{0} \end{pmatrix} \\
\Xi &\equiv \begin{pmatrix} \underset{n_1\times n_1}{I} - [\Omega]_{12}^{\frac{1}{2}}[\Omega]_{22}^{-\frac{1}{2}}[\Omega]_{21}^{\frac{1}{2}} & -[\Omega]_{12}^{\frac{1}{2}}[\Omega]_{22}^{-\frac{1}{2}} \\ -[\Omega]_{22}^{-\frac{1}{2}}[\Omega]_{21}^{\frac{1}{2}} & \underset{n_2\times n_2}{0} \end{pmatrix}
\end{aligned}
$$

Therefore,

$$
\begin{aligned}
det\left(A^{-1}B - \mu \underset{(n+n_2)\times(n+n_2)}{I}\right) &= det\begin{pmatrix} \underset{n\times n}{\Xi} - \lambda \underset{n\times n}{I} & \underset{n\times n_2}{0} \\ \underset{n_2\times n}{0} & -\lambda \underset{n_2\times n_2}{I} \end{pmatrix} \\
&= det(\underset{n\times n}{\Xi} - \lambda \underset{n\times n}{I})det\left(-\lambda \underset{n_2\times n_2}{I}\right) \\
&= det(\underset{n\times n}{\Xi} - \lambda \underset{n\times n}{I})(-\lambda)^{n_2}
\end{aligned}
$$

Therefore, we establish that there exist n_2 zero eigenvalues.

With regard to $det(\underset{n\times n}{\Xi} - \lambda \underset{n\times n}{I})$:

$$
det\begin{pmatrix} \underset{n_1\times n_1}{I} - [\Omega]_{12}^{\frac{1}{2}}[\Omega]_{22}^{-\frac{1}{2}}[\Omega]_{21}^{\frac{1}{2}} - \lambda \underset{n_1\times n_1}{I} & -[\Omega]_{12}^{\frac{1}{2}}[\Omega]_{22}^{\frac{1}{2}} \\ -[\Omega]_{22}^{\frac{1}{2}}[\Omega]_{21}^{\frac{1}{2}} & -\lambda \underset{n_2\times n_2}{I} \end{pmatrix} = 0
$$

and therefore:

$$
\begin{aligned}
det(I - [\Omega]_{12}^{\frac{1}{2}}[\Omega]_{22}^{-\frac{1}{2}}[\Omega]_{21}^{\frac{1}{2}} - \lambda I + [\Omega]_{12}^{\frac{1}{2}}[\Omega]_{22}^{-1}\lambda^{-1}[\Omega]_{21}^{\frac{1}{2}})\lambda^{n_2} &= 0 \\
det(\lambda I + (1-\lambda)[\Omega]_{12}^{\frac{1}{2}}[\Omega]_{22}^{-\frac{1}{2}}[\Omega]_{21}^{\frac{1}{2}} - \lambda^2 I)\lambda^{n_2} &= 0 \\
det(\lambda(1-\lambda)I + (1-\lambda)[\Omega]_{12}^{\frac{1}{2}}[\Omega]_{22}^{-\frac{1}{2}}[\Omega]_{21}^{\frac{1}{2}}\lambda^{n_2} &= 0 \\
det(\lambda I + [\Omega]_{12}^{\frac{1}{2}}[\Omega]_{22}^{-\frac{1}{2}}[\Omega]_{21}^{\frac{1}{2}})(1-\lambda)^{n_1}\lambda^{n_2} &= 0
\end{aligned}
$$

Since $[\Omega]_{12}^{\frac{1}{2}}[\Omega]_{22}^{-\frac{1}{2}}[\Omega]_{21}^{\frac{1}{2}}$ is positive definite, $det(\lambda I + [\Omega]_{12}^{\frac{1}{2}}[\Omega]_{22}^{-\frac{1}{2}}[\Omega]_{21}^{\frac{1}{2}})$ is not zero for any value of λ. We therefore have that the eigenvalues of $A^{-1}B$ are 1, with multiplicity n_1 and 0

with multiplicity $2n_2$. Therefore, the eigenvalues that solve equation 8.3 are $\lambda_i = 1$ for $i \leq n_1$ and $\lambda_i = \infty$ for $i = n_1...n + n_2$.

Notice that in the analysis above we have not constrained the set of eigenvectors we considered beyond the bound on $V'BV$. Since the vectors V we specified have a certain structure, the maximum value attained should be less than or equal to the value implied by the set of solutions that correspond to λ_i.

Since the set of potential maximum values are either $\sum_{i \leq n_1} \lambda_i = n_1$ or ∞ it is easier to search for the admissible vectors $V*$(in terms of R) that could possibly achieve this maximum. The system that determines the eigenvector is the following:

$$\begin{pmatrix} \underset{n\times n}{I} & \underset{n\times n_2}{-\tilde{S}_2'} \\ \underset{n_2\times n}{-\tilde{S}_2} & \underset{n_2\times n_2}{\Omega_{22}} \end{pmatrix} \begin{pmatrix} \underset{n\times n_1}{\tilde{S}_1'} \\ \underset{n_2\times n_1}{J'} \end{pmatrix} = \lambda \begin{pmatrix} \underset{n\times n_1}{\tilde{S}_1'} \\ \underset{n_2\times n_1}{J'} \end{pmatrix}$$

From the first set of equations, we have that:

$$\begin{aligned} \tilde{G}^{12}\tilde{R}(\tilde{R}'\tilde{R})^{-1}\tilde{R}'\tilde{S}_2 &= \tilde{S}_1(1-\lambda) \\ &\therefore \\ \tilde{G}^{12}\tilde{R} &= \tilde{S}_1(1-\lambda)\tilde{S}_2'(\tilde{S}_2\tilde{S}_2')^{-1}\tilde{R} \end{aligned}$$

Solving for the second set of equations,

$$\tilde{G}^{12}\tilde{R} = ([\Omega]_{11}^{\frac{1}{2}}[\Omega]_{12}^{\frac{1}{2}} + [\Omega]_{12}^{\frac{1}{2}}[\Omega]_{22}^{\frac{1}{2}})(\lambda I - \Omega_{22})^{-1}\tilde{S}_2'(\tilde{S}_2\tilde{S}_2')^{-1}\tilde{R}$$

First, note that any value of $\tilde{R}$ satisfies both equations for $\lambda \notin \{1, \infty\}$. Moreover, we discard the possibility that corresponds to $\lambda_i = \infty$ as for $\tilde{R}$ to satisfy the first set, a non differentiable $r(\vartheta)$ is required. We then turn to the only possibility left, that of $\lambda_i = 1$. For $\lambda = 1$, the only admissible solution of the first set is $R = 0$, while the second set is

also satisfied. $\tilde{R} = 0$ is then the only admissible solution. The constrained maximum is therefore equal to $\sum_{i \leq n_1} 1 = n_1$. Thus, for $R \neq 0$, $tr(\mathbb{V}(S_1 \mathcal{Z})^{-1} \mathbb{V}(\mathcal{Z}_r)) < n_1$. □

9. Appendix C

9.1. **Additional Experiments.**

Second MC Experiment. We first present the true data generating process (DGP) for the vector of observables (X, Y), which is partially unknown to the econometrician, up to a single non linear unconditional moment condition.

Let $\{y_i, x_i\}_{i=1, n \geq 1}^{n}$ an iid sequence generated by the following DGP:

$$\begin{aligned} y_i &= \delta_1 + u_i \\ u_i &= \varepsilon_i + \delta_2 x_i + \delta_3 x_i^2 \\ \varepsilon_i &\sim iidD_1(\alpha_1, \alpha_2) \\ x_i &\sim iidD_2(\gamma_1, \gamma_2) \end{aligned}$$

In the following simulation experiments D is a generic distribution. Different assumptions on D will be made to investigate different cases of misspecification i.e. in the location, scale, skewness and kurtosis. As already noted, the above model satisfies the following (arbitrary) moment restriction:

$$\mathbb{E}(y^{-\beta_0} - 2\beta_0 yx) = 0$$

To perform the experiments, we adopt the following base model:

$$\begin{aligned} y_i &= \delta_1 + u_i \\ u_i &\sim iidD_3(\alpha_{1b}, \alpha_{2b}) \\ x_i &\sim iidD_4(\gamma_{1b}, \gamma_{2b}) \end{aligned}$$

Clearly the probability model used in this exercise goes wrong in many dimensions i.e. has omitted variables and has different distributional assumptions. We plot below the MSE (left panel) when using the true and the misspecified density and the implied true and misspecified densities of u_t (right panel) for the following four cases:

TABLE 1. Distributions Used

Case	$D_1(\alpha_1, \alpha_2), D_2(\gamma_1, \gamma_2)$	$D_3(\alpha_{1b}, \alpha_{2b}), D_4(\gamma_{1b}, \gamma_{2b})$
1	$t(7)$,$\Gamma(2,5)$	$N(0,4)$,$\Gamma(2,5)$
2	$N(0,4)$,$\Gamma(2,5)$	$N(0,4)$,$\Gamma(2,5)$
3	$t(7)$,$U(0,1)$	$N(0,4)$,$U(0,1)$
4	$N(0,4)$,$U(0,1)$	$N(0,4)$,$U(0,1)$

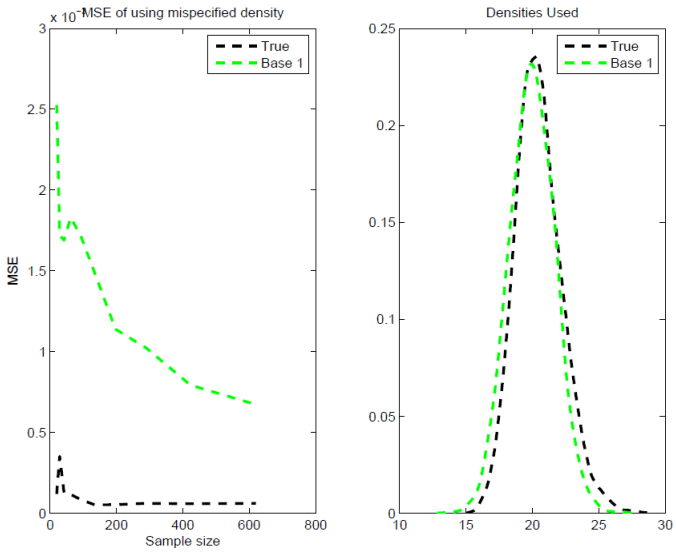

FIGURE 9.1. Monte Carlo Case 1

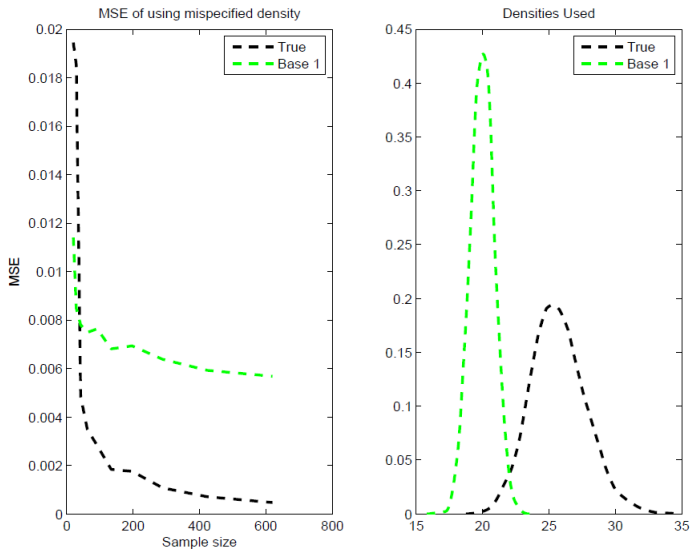

FIGURE 9.2. Monte Carlo Case 2

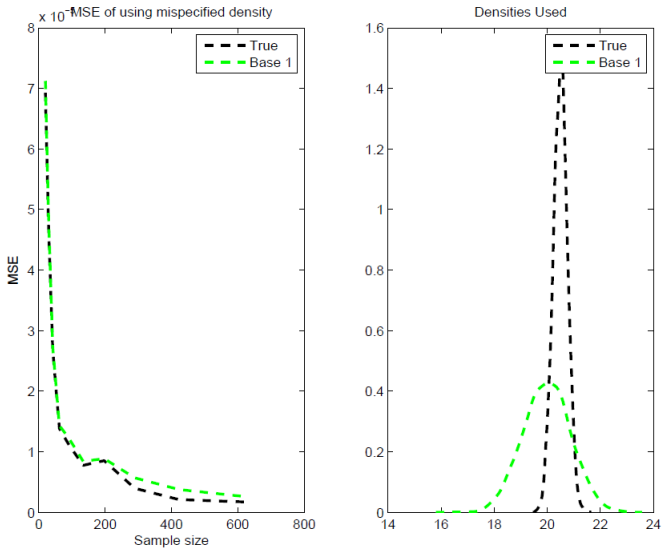

FIGURE 9.3. Monte Carlo Cases 3

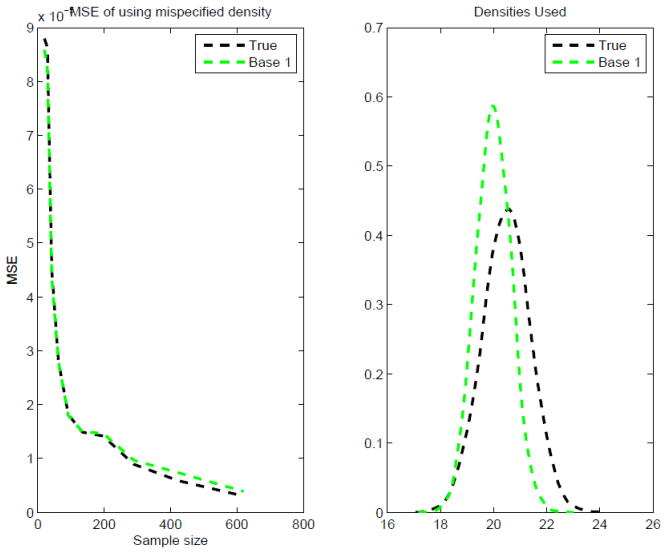

FIGURE 9.4. Monte Carlo Case 4

Evidently, the biggest differences arise when density misspecification is severe i.e. in case 2. In this case the auxiliary density assigns very little mass on the support of u_t. In the rest of cases differences are very small, especially at sample sizes comparable to the conventional size of macroeconomic datasets. Also, note that we have not estimated any of the parameters of the base densities[11].

Estimating a prototypical DSGE. The prototypical DSGE model estimated is the standard stochastic growth model with full depreciation, see for example Ireland (2004). Let $x_t \equiv (y_t, c_t, h_t, k_t)$ be output, consumption, hours, capital. The first order equilibrium conditions of the model are the following:

[11] Given that estimation involves also the finite dimensional nuisance parameter φ_0, it is instructive to notice that since $\varphi_0 \xrightarrow{p} \arg\min_\Phi \int p(x|z) \log\left(\frac{p(x|z)}{h(x|z,\varphi_0,\vartheta)}\right) dx \equiv KL(P,H)$ for any $\vartheta \in \Theta$, and by Pinsker inequality, we know that $TV(P,H) \leq KL(P,H)$. Therefore, minimizing $KL(P,H)$ implies minimizing also $\int |M(x,\vartheta)||p(x) - h(x,\varphi,\vartheta)|dx$.

$$Y_t = A_t K_t^{\theta} H_t^{1-\theta} \tag{9.1}$$

$$K_{t+1} = Y_t - C_t \tag{9.2}$$

$$\gamma C_t H_t = (1-\theta) Y_t I_t \tag{9.3}$$

$$\frac{1}{C_t} = \beta \mathbb{E}_t \left\{ \frac{1}{C_{t+1}} \left(\theta \left(\frac{Y_{t+1}}{K_{t+1}} \right) \right) \right\} \tag{9.4}$$

$$log(I_{t+1}) = \rho_I log(I_t) + logN(0, \sigma_I^2) \tag{9.5}$$

$$log(A_{t+1}) = \rho_I log(A_t) + logN(0, \sigma_A^2) \tag{9.6}$$

where 3.11 is the typical Cobb Douglas production function, 3.12 is capital accumulation equation, 3.13 the distorted (by a marginal efficiency shock I_t) intra-temporal efficiency condition and 3.14 the inter-temporal efficiency condition (consumption Euler equation).

In this case, we know much more information about the conditional predictive density, $h(x_{t+1}|x_t;\varphi)$, since the only equation that is not immediately solved is the Euler equation . The rest of the equations of the system can be readily reduced to a single equation, and then plugged in the Euler equation. This leads to great efficiency gains as the mapping of a subset of ϕ to ϑ is now known. The only mapping that is still unknown is that of the reduced form of consumption, since we do not solve for consumption. Moreover, uncertainty about the consumption function translates to uncertainty about the exact solution for hours H_t and output Y_t.

For simplicity we assume that we in principle observe all the variables of the system. Different sets of observables would lead to a different form for 3.18 that would be used for estimation. Future work could look at the accommodation of unobserved variables. Our conjecture is that exogenous unobserved components can be easily accommodated while endogenous unobserved variables are much more challenging [12].

[12]For recent advances towards this direction, see Gallant, Giacomini, and Ragusa (2016)

With regard to the solution of the model, the true solution vector for $C_{t+1}, H_{t+1}, K_{t+1}$ is the following:

$$\begin{aligned} C_{t+1} &= (1-\beta\theta)Y_{t+1} \\ H_{t+1} &= \frac{1-\theta}{\gamma(1-\beta\theta)} I_{t+1} \\ K_{t+1} &= \theta\beta Y_t \\ log(I_{t+1}) &= \rho_I log(I_t) + logN(0,\sigma_I^2) \\ log(A_{t+1}) &= \rho_I log(A_t) + logN(0,\sigma_A^2) \end{aligned}$$

which is essentially log-linear. In the following experiment, we will simulate 200 observations for $X_t \equiv (A_t, I_t, C_t, H_t, K_t)'$ and $(\beta, \theta, \gamma, \rho_A, \rho_I, \sigma_A, \sigma_I) := (0.96, 0.3, 0.5, 0.9, 0.9, 1, 1)$ and then use this as a pseudo-dataset. As a base conditional density , $h(X_{t+1}|X_t)$ we use the log-Normal distribution, $logN(B(\psi)X_t, C(\psi)\Sigma C'(\psi))$ where ψ includes both (β, γ, θ), $(\rho_I, \rho_A, \sigma_A, \sigma_I)$ and nuisance parameters $\underset{n_\psi \times 1}{\phi}$. At the end of this section we show the explicit form of B and C when solution is partially unknown and observations on X_t are used. The corresponding moment condition used as a constraint in the projection is the following:

$$\frac{1}{C_t} = \beta\theta\mathbb{E}_t\left\{\frac{1}{C_{t+1}} A_{t+1}\left(\frac{H_{t+1}}{K_{t+1}}\right)^{1-\theta}\right\} \tag{9.7}$$

Due to identification issues, we set $\beta = 0.96$ and $\gamma = 0.5$, and we therefore estimate θ together with the rest of the nuisance parameters. We also set $\sigma_A = 1, \sigma_I = 1$. Due to the fact that we use 5 observables and we only have two independent sources of variation, we add measurement error to (C_t, H_t, K_t), with variance $\sigma_{me} = 0.25$. We report below the point estimates and confidence bands from a chain of 30000 draws :

TABLE 2. Parameter Estimates

Parameter	$q_{2.5\%}$	Point	$q_{97.5\%}$
θ	**0.19**	**0.35**	**0.49**
b_{31}	0.75	1.11	1.54
b_{32}	-0.21	0.32	0.71
b_{34}	0.12	0.45	0.72
b_{35}	0.19	0.45	0.81
c_{31}	-0.05	0.74	1.67
c_{32}	0.19	0.84	1.63
c_{41}	-1.05	-0.20	0.46
c_{42}	0.03	0.77	1.55

We also performed the estimation in the case of knowing the full likelihood function of the model. The corresponding point estimate for θ is 0.3173 and the two sided 95% confidence interval is $(0.10, 0.49)$. The results are therefore similar.

Reduced form coefficients.

$$B \equiv \begin{pmatrix} \rho_A & 0 & 0 & 0 & 0 \\ 0 & \rho_I & 0 & 0 & 0 \\ b_{31} & b_{32} & 0 & b_{34} & b_{35} \\ \frac{1+\theta}{\theta}\rho_A - \frac{1}{\theta}b_{31} & \frac{1}{\theta}(\rho_I - b_{32}) & 0 & 1-\theta-\frac{1}{\theta}b_{34} & 1-\frac{1}{\theta}b_{35} \\ \rho_A & 0 & 0 & 1-\theta & \theta \end{pmatrix}$$

$$C \equiv \begin{pmatrix} 1 & 0 \\ 0 & 1 \\ c_{31} & c_{32} \\ c_{41} & c_{42} \\ 1 & 0 \end{pmatrix}$$

where $\phi \equiv (vec(b), vec(c)')$

REFERENCES

BILLINGSLEY, P. (1961): "The Lindeberg-LÃ©vy Theorem for Martingales," Proceedings of the American Mathematical Society, 12(5), pp. 788–792.

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

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

Chamberlain, G. (1987): "Asymptotic efficiency in estimation with conditional moment restrictions," Journal of Econometrics, 34(3), 305 – 334.

Chen, X. (2007): "Chapter 76 Large Sample Sieve Estimation of Semi-Nonparametric Models," vol. 6, Part B of Handbook of Econometrics, pp. 5549 – 5632. Elsevier.

Chernozhukov, V., and H. Hong (2003): "An {MCMC} approach to classical estimation," Journal of Econometrics, 115(2), 293 – 346.

Cressie, N., and T. R. C. Read (1984): "Multinomial Goodness-of-Fit Tests," Journal of the Royal Statistical Society. Series B (Methodological), 46(3), pp. 440–464.

den Haan, W. J., and J. de Wind (2010): "How well-behaved are higher-order perturbation solutions?," DNB Working Papers 240, Netherlands Central Bank, Research Department.

Gallant, A. R., R. Giacomini, and G. Ragusa (2016): "Bayesian Estimation of State Space Models Using Moment Conditions," Working paper.

Gallant, A. R., and G. Tauchen (1989): "Seminonparametric Estimation of Conditionally Constrained Heterogeneous Processes: Asset Pricing Applications," Econometrica, 57(5), pp. 1091–1120.

Giacomini, R., and G. Ragusa (2014): "Theory-coherent forecasting," Journal of Econometrics, 182(1), 145 – 155, Causality, Prediction, and Specification Analysis: Recent Advances and Future Directions.

Hansen, B. E. (2016): "Efficient shrinkage in parametric models," Journal of Econometrics, 190(1), 115–132.

Hansen, L. P. (1982): "Large Sample Properties of Generalized Method of Moments Estimators," Econometrica, 50(4), 1029–1054.

Hansen, L. P., J. Heaton, and A. Yaron (1996): "Finite-Sample Properties of Some Alternative GMM Estimators," Journal of Business and Economic Statistics, 14(3), pp. 262–280.

Hausman, J., R. Lewis, K. Menzel, and W. Newey (2011): "Properties of the {CUE} estimator and a modification with moments," Journal of Econometrics, 165(1), 45 – 57, Moment Restriction-Based Econometric Methods.

Hotz, V. J., and R. A. Miller (1993): "Conditional Choice Probabilities and the Estimation of Dynamic Models," The Review of Economic Studies, 60(3), pp. 497–529.

I.Csiszar (1975): "I-Divergence Geometry of Probability Distributions and Minimization Problems," Annals of Probability, 3(1), 146–158.

Imbens, G. W., R. H. Spady, and P. Johnson (1998): "Information Theoretic Approaches to Inference in Moment Condition Models," Econometrica, 66(2), 333–357.

Ireland, P. N. (2004): "A method for taking models to the data," Journal of Economic Dynamics and Control, 28(6), 1205 – 1226.

Kitamura, Y., and M. Stutzer (1997): "An Information-Theoretic Alternative to Generalized Method of Moments Estimation," Econometrica, 65(4), 861–874.

Kitamura, Y., G. Tripathi, and H. Ahn (2004): "Empirical Likelihood-Based Inference in Conditional Moment Restriction Models," Econometrica, 72(6), pp. 1667–1714.

Komunjer, I., and G. Ragusa (2016): "Existence and Characterization of Conditional Density Projections," Econometric Theory, FirstView, 1–41.

Leeper, E., and C. Leith (2016): "Chapter 30 - Understanding Inflation as a Joint Monetary -Fiscal Phenomenon," vol. 2 of Handbook of Macroeconomics, pp. 2305 – 2415. Elsevier.

Newey, W. K., and D. McFadden (1994): "Chapter 36 Large sample estimation and hypothesis testing," vol. 4 of Handbook of Econometrics, pp. 2111 – 2245. Elsevier.

Newey, W. K., and R. J. Smith (2004): "Higher Order Properties of Gmm and Generalized Empirical Likelihood Estimators," Econometrica, 72(1), 219–255.

Pesaran, M. H. (1987): The Limits to Rational Expectations. Basil Blackwell.

Rust, J. (1987): "Optimal Replacement of GMC Bus Engines: An Empirical Model of Harold Zurcher," Econometrica, 55(5), pp. 999–1033.

Schennah, S. (2007): "Point Estimation with Exponentially Tilted Likelihood," Annals of Statistics, 35(2), 634–672.

Shin, M. (2014): "Bayesian GMM," Manuscript.

Su, C.-L., and K. L. Judd (2012): "Constrained Optimization Approaches to Estimation of Structural Models," Econometrica, 80(5), 2213–2230.

van der Vaart, A. (1998): Asymptotic statistics, Cambridge Series in Statistical and Probabilistic Mathematics. Cambridge University Press.

Wasserman, L. (2006): All of Nonparametric Statistics (Springer Texts in Statistics). Springer-Verlag New York, Secaucus, NJ, USA.

SFB 649 Discussion Paper Series 2017

For a complete list of Discussion Papers published by the SFB 649, please visit http://sfb649.wiwi.hu-berlin.de.

001 "Fake Alpha" by Marcel Müller, Tobias Rosenberger and Marliese Uhrig-Homburg, January 2017.
002 "Estimating location values of agricultural land" by Georg Helbing, Zhiwei Shen, Martin Odening and Matthias Ritter, January 2017.
003 "FRM: a Financial Risk Meter based on penalizing tail events occurrence" by Lining Yu, Wolfgang Karl Härdle, Lukas Borke and Thijs Benschop, January 2017.
004 "Tail event driven networks of SIFIs" by Cathy Yi-Hsuan Chen, Wolfgang Karl Härdle and Yarema Okhrin, January 2017.
005 "Dynamic Valuation of Weather Derivatives under Default Risk" by Wolfgang Karl Härdle and Maria Osipenko, February 2017.
006 "RiskAnalytics: an R package for real time processing of Nasdaq and Yahoo finance data and parallelized quantile lasso regression methods" by Lukas Borke, February 2017.
007 "Testing Missing at Random using Instrumental Variables" by Christoph Breunig, February 2017.
008 "GitHub API based QuantNet Mining infrastructure in R" by Lukas Borke and Wolfgang K. Härdle, February 2017.
009 "The Economics of German Unification after Twenty-five Years: Lessons for Korea" by Michael C. Burda and Mark Weder, April 2017.
010 "DATA SCIENCE & DIGITAL SOCIETY" by Cathy Yi-Hsuan Chen and Wolfgang Karl Härdle, May 2017.
011 "The impact of news on US household inflation expectations" by Shih-Kang Chao, Wolfgang Karl Härdle, Jeffrey Sheen, Stefan Trück and Ben Zhe Wang, May 2017.
012 "Industry Interdependency Dynamics in a Network Context" by Ya Qian, Wolfgang Karl Härdle and Cathy Yi-Hsuan Chen, May 2017.
013 "Adaptive weights clustering of research papers" by Larisa Adamyan, Kirill Efimov, Cathy Yi-Hsuan Chen, Wolfgang K. Härdle, July 2017.
014 "Investing with cryptocurrencies - A liquidity constrained investment approach" by Simon Trimborn, Mingyang Li and Wolfgang Karl Härdle, July 2017.
015 "(Un)expected Monetary Policy Shocks and Term Premia" by Martin Kliem and Alexander Meyer-Gohde, July 2017.
016 " Conditional moment restrictions and the role of density information in estimated structural models" by Andreas Tryphonides, July 2017.

SFB 649, Spandauer Straße 1, D-10178 Berlin
http://sfb649.wiwi.hu-berlin.de

This research was supported by the Deutsche Forschungsgemeinschaft through the SFB 649 "Economic Risk".